TRENDS IN INTERNATIONAL MATHEMATICS AND SCIENCE STUDY

TIMSS

Methods and Procedures: TIMSS 2019 Technical Report

Michael O. Martin,
Matthias von Davier,
Ina V.S. Mullis, Editors

IEA
TIMSS & PIRLS
International Study Center
Lynch School of Education
BOSTON COLLEGE

Methods and Procedures: TIMSS 2019 Technical Report
Michael O. Martin, Matthias von Davier, and Ina V.S. Mullis, Editors

Publishers: TIMSS & PIRLS International Study Center,
Lynch School of Education and Human Development, Boston College and
International Association for the Evaluation of Educational Achievement (IEA)

Library of Congress Catalog Card Number: 2020901029
ISBN-978-1-889938-53-0

For more information about TIMSS contact:
TIMSS & PIRLS International Study Center
Lynch School of Education and Human Development
Boston College
Chestnut Hill, MA 02467
United States

tel: +1-617-552-1600
e-mail: timssandpirls.bc.edu
timssandpirls.bc.edu

Assessment Development

Chapter 1
Developing the TIMSS 2019 Mathematics and Science Achievement Instruments **1.1**
Kerry E. Cotter, Victoria A.S. Centurino, and Ina V.S. Mullis

Chapter 2
Updating the TIMSS 2019 Instruments for Describing the Contexts for Student Learning. **2.1**
Ina V.S. Mullis and Bethany Fishbein

Chapter 3
Sample Design in TIMSS 2019 . . . . **3.1**
Sylvie LaRoche, Marc Joncas, and Pierre Foy

Data Collection Procedures

Chapter 4
eAssessment System for TIMSS 2019 . . . . **4.1**
Mark Cockle and Heiko Sibberns

Chapter 5
Instrument Translation and Layout Verification for TIMSS 2019 . . . . **5.1**
David Ebbs, Erin Wry, Jan-Philipp Wagner, and Andrea Netten

Chapter 6
Survey Operations Procedures for TIMSS 2019 . . . . **6.1**
Ieva Johansone

Chapter 7
International Quality Assurance Program for TIMSS 2019 . . . . **7.1**
Ieva Johansone and Sue Flicop

Chapter 8
Creating the TIMSS 2019 International Database **8.1**
Kamil Kowolik, Mark Cockle, and Milena Tavena

Analysis and Reporting

Chapter 9
Sample Implementation in TIMSS 2019 . . . . **9.1**
Sylvie LaRoche and Pierre Foy

Chapter 10
Reviewing the TIMSS 2019 Achievement Item Statistics **10.1**
Bethany Fishbein, Pierre Foy, and Lillian Tyack

Chapter 11
TIMSS 2019 Scaling Methodology: Item Response Theory, Population Models, and Linking Across Modes **11.1**
Matthias von Davier

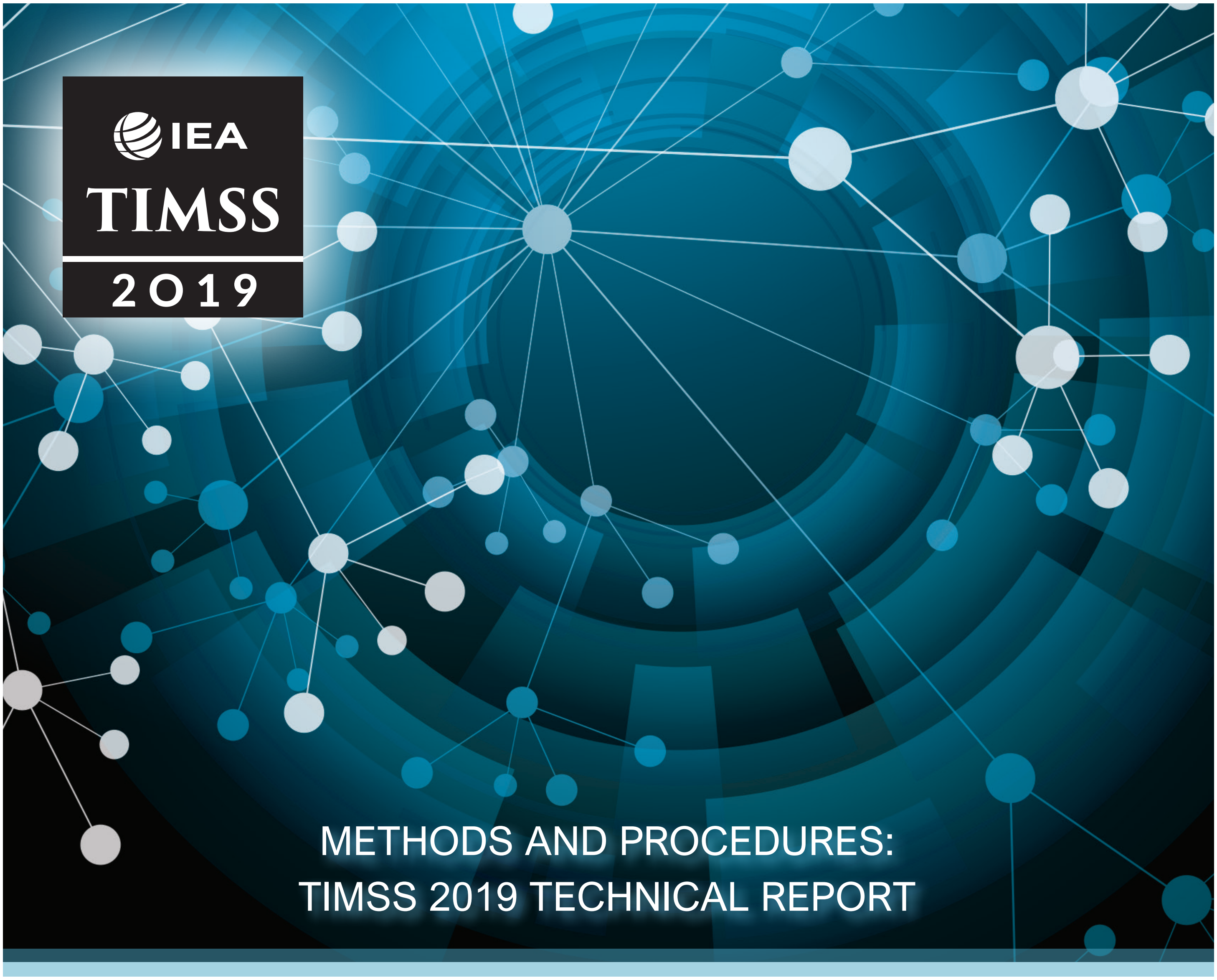
Chapter 12
Implementing the TIMSS 2019 Scaling Methodology. **12.1**
Pierre Foy, Bethany Fishbein, Matthias von Davier, and Liqun Yin

Chapter 13
Examining eTIMSS Country Differences Between eTIMSS Data and Bridge Data: A Look at Country-Level Mode of Administration Effects. . . **13.1**
Matthias von Davier, Pierre Foy, Michael O. Martin, and Ina V.S. Mullis

Chapter 14
Estimating Standard Errors in the TIMSS 2019 Results **14.1**
Pierre Foy and Sylvie LaRoche

Chapter 15
Using Scale Anchoring to Interpret the TIMSS 2019 Achievement Scales. **15.1**
Ina V.S. Mullis and Bethany Fishbein

Chapter 16
Creating and Interpreting the TIMSS 2019 Context Questionnaire Scales **16.1**
Liqun Yin and Bethany Fishbein



METHODS AND PROCEDURES:
TIMSS 2019 TECHNICAL REPORT

ASSESSMENT DEVELOPMENT

IEA
TIMSS & PIRLS
International Study Center
Lynch School of Education
BOSTON COLLEGE

CHAPTER 1

Developing the TIMSS 2019 Mathematics and Science Achievement Instruments

Kerry E. Cotter
Victoria A.S. Centurino
Ina V.S. Mullis

Unique Characteristics of TIMSS 2019

The TIMSS assessments are designed to provide valid measurement of the mathematics and science content and skills that are valued by the international education community and included in the curricula of participating countries. The general approach to developing the TIMSS mathematics and science achievement items to meet this goal is similar from one assessment cycle to the next, but each cycle has some unique characteristics that influence instrument development. Besides providing measures on another cycle for the TIMSS trend lines monitoring changes in educational achievement since 1995, TIMSS 2019 also was remarkable for several reasons.

- TIMSS 2019 marked the beginning of the transition to eTIMSS—a digital version of TIMSS designed for computer- and tablet-based administration. eTIMSS offered an engaging, interactive, and visually attractive assessment that enabled TIMSS 2019 to better assess complex areas of the mathematics and science frameworks and increase operational efficiency in translation, assessment delivery, data entry, and scoring.
- As a part of the transition to digital assessment, eTIMSS 2019 included a series of extended Problem Solving and Inquiry (PSI) tasks in mathematics and science at both the fourth and the eighth grades. The eTIMSS PSIs were designed to simulate real world or laboratory situations in which students could integrate and apply process skills and content knowledge to solve mathematics problems or conduct virtual scientific experiments and investigations.
- Building on the success of TIMSS Numeracy 2015, the TIMSS 2019 assessment design was expanded to support a less difficult version of the fourth grade mathematics assessment that had some blocks of items in common with the regular fourth grade mathematics assessment. The less difficult version enabled TIMSS 2019 to provide improved measurement for participating countries where fourth grade students were still developing fundamental mathematics skills.

Results for the two versions of the fourth grade mathematics assessment were linked through the common items and reported on the same TIMSS achievement scale.

Transitioning TIMSS to eTIMSS

Transitioning to digital assessment is important to "keep up with the times" and to increase both construct representation and data utility. Because not all TIMSS countries were prepared to conduct digital assessments, IEA decided to implement the transition over two assessment cycles—TIMSS 2019 and TIMSS 2023. More than half of the 64 countries participating in TIMSS 2019 elected to administer the "e" version of the assessments, while the rest of the countries administered TIMSS in paper-and-pencil format, as in previous assessment cycles (paperTIMSS).

The eTIMSS 2019 assessments included a variety of technology-enhanced item formats, with colorful graphics and interactive features. These features extended coverage of the mathematics and science frameworks and promoted student engagement. The digital mode of administration also allowed for a substantial proportion of the eTIMSS mathematics items to be machine scored.

The eTIMSS 2019 assessments were created and administered using IEA's eAssessment System, which houses a collection of online tools used for instrument creation, translation and adaptation, verification, delivery to students, scoring, and data entry. The digital mode of administration allowed eTIMSS to collect information about how students work through the items, such as screen-by-screen timing data and additional process variables that can be analyzed to study students' interactions with the achievement items.

The eTIMSS 2019 PSIs, designed exclusively for eTIMSS, were a new and pioneering effort to improve measurement of higher-order mathematics and science skills by capitalizing on the digital mode of administration. Each PSI consisted of a sequence of 4 to 16 items that were set in a cohesive context and addressed a range of topics from the *TIMSS 2019 Assessment Frameworks* (Mullis & Martin, 2017), such as solving a series of mathematics problems to plan a school event or conducting a virtual scientific experiment to study plant growth. The items within these situational tasks included a broader array of innovative digital features than the regular eTIMSS achievement items and provided scaffolding for complex mathematics problems and science investigations.

Less Difficult Mathematics at the Fourth Grade

For a variety of reasons, there are some countries where most children in the fourth grade are still developing fundamental mathematics skills. To offer countries the most effective assessment of fourth grade mathematics, IEA offers options for matching the TIMSS fourth grade mathematics assessment to the country's educational development and students' mathematics proficiency. For some countries, the less difficult version of the TIMSS 2019 fourth grade mathematics assessment was a better match with students' learning.

The TIMSS 2019 fourth grade assessment with less difficult mathematics was developed together with the regular fourth grade mathematics assessment and reflected the mathematics described in the TIMSS 2019 Mathematics Framework (Lindquist, Philpot, Mullis, & Cotter, 2017). The regular and less difficult versions of the assessment were equivalent in scope, and about one-third of the items were the same between the two versions. The other two-thirds of the items in each version of the assessment addressed the same areas of the mathematics framework, but the items in the less difficult version involved less complex numbers and situations. The items in common between the two versions of the fourth grade mathematics assessment enabled the two assessments to be linked so that the results could be reported together and directly compared. Expert committees reviewed both the regular and less difficult mathematics items together at each phase of development.

The design of the TIMSS 2019 less difficult mathematics assessment improved upon the design of TIMSS Numeracy 2015, which was a stand-alone mathematics assessment that did not include science. For TIMSS 2019 at the fourth grade, countries could opt for either regular or less difficult mathematics, together with science. A substantial portion of the items in the less difficult version of the TIMSS 2019 mathematics assessment was carried forward from TIMSS Numeracy 2015, which enabled trend measurement for countries that participated in TIMSS Numeracy 2015.

The TIMSS Approach to Measuring Trends

Because TIMSS is designed to measure trends, the assessments of mathematics and science cannot change dramatically from cycle to cycle. That is, TIMSS is based on a well-known premise for designing trend assessments (ascribed to John Tukey and Albert Beaton):

"If you want to measure change, do not change the measure."

However, the achievement items also need to be updated with each cycle to prevent the assessments from becoming dated and no longer relevant to current learning goals and policy issues. It is important that TIMSS reflects the most recent discoveries in the field and is presented in ways consistent with students' instructional and everyday experiences.

To maintain continuity with past assessments while keeping up with current topics and technology, the TIMSS assessments evolve with each cycle. TIMSS has a specific design for rotating items out of the assessment after each cycle and replacing them with newly developed items for the following cycle. The remaining achievement items, referred to as "trend items," are kept secure to be re-administered in subsequent cycles. With this design for item replacement, each TIMSS assessment includes items from three cycles—newly developed items, items from the previous cycle, and items from two cycles before.

Overview of the TIMSS 2019 Achievement Items

Although the majority of the TIMSS achievement items are carried forward from the previous assessment cycle to measure trends, the task of updating the instruments for each new cycle—every four years since 1995—is a substantial undertaking. Because TIMSS assesses two subjects at two grades, it actually encompasses five different assessments of achievement—regular and less difficult mathematics at the fourth grade, mathematics at the eighth grade, and science at the fourth and eighth grades.

The TIMSS 2019 fourth grade assessments required developing and field testing 261 new mathematics and science items in both digital and paper formats as well as 66 new paper-based items for the less difficult version of the mathematics assessment. The TIMSS 2019 eighth grade assessments required developing and field testing 325 new mathematics and science items in both digital and paper formats. For eTIMSS 2019, the field test also included eight mathematics and science PSI tasks at the fourth grade and seven mathematics and science PSI tasks at the eighth grade.

Since the beginning in 1995, the TIMSS assessments have included two general item formats: selected response (i.e., questions to which students choose their answer(s) from a set of options) and constructed response (i.e., questions to which students construct their own responses). For each constructed response item, a unique scoring guide is developed along with the item with clear distinctions among correct, incorrect, and, if applicable, partially correct answers. The format of each item is chosen based on the mathematics or science content and cognitive domain being assessed.

The Item Development Process

The TIMSS & PIRLS International Study Center at Boston College employs a collaborative process inspired by the principles of the evidence-centered design framework (ECD; Mislevy, Almond, & Lukas, 2003) to develop the new achievement items needed for each TIMSS cycle. With this approach, validity is supported by adhering to best practices in assessment design throughout the development process—namely, clearly defining the target construct to be measured, specifying the items needed to measure it, establishing standards for items and test forms, and ensuring that the assessments meet the test specifications. A broad overview of this process to support coherence between the assessment goals and data includes:

- Updating the assessment frameworks to identify and prioritize the mathematics and science content and skills that the assessment will measure
- Developing achievement items as well as scoring guides for constructed response items to meet the assessment specifications delineated in the frameworks
- Conducting a full-scale field test to evaluate the measurement properties of the item pool and practice the data collection and scoring procedures

- Selecting the new items to meet the assessment specifications based on the field test results and trend items from previous cycles
- Conducting training in how to reliably score students' responses to constructed response items to ensure the quality of the data

The development process is directed and managed by the staff of the TIMSS & PIRLS International Study Center, who collectively have considerable experience in the measurement and assessment of mathematics and science achievement. For TIMSS 2019, Executive Director, Ina Mullis, and Assistant Director of Mathematics, Kerry Cotter, managed the mathematics assessment development. Executive Director, Michael Martin, and Assistant Director of Science, Victoria Centurino, managed the science assessment development.

Also playing a key role in achievement item development were the TIMSS 2019 National Research Coordinators (NRCs) designated by their countries to be responsible for the complex tasks involved in implementing TIMSS in their countries. The TIMSS & PIRLS International Study Center worked with the NRCs and experts from the participating countries to update the assessment frameworks and develop the new achievement items, including the scoring guides for constructed response items. The NRCs reviewed the items prior to the field test and helped select the items for the assessment after the field test.

The TIMSS & PIRLS International Study Center prepared an international version of all the TIMSS achievement items in English. Subsequently, the items were translated by participating countries into their languages of instruction with the goal of creating high quality translations that were appropriately adapted for the national context and at the same time remained internationally comparable. Therefore, a significant portion of the NRCs' development and review effort was dedicated to ensuring that the achievement items could be translated accurately.

Additional advice and guidance was provided through periodic reviews by the TIMSS 2019 Science and Mathematics Item Review Committee (SMIRC). SMIRC members for each TIMSS cycle are nominated by countries participating in TIMSS and provide guidance in developing the TIMSS assessments. The TIMSS 2019 SMIRC consisted of 13 members: 7 experts in mathematics and mathematics education and 6 experts in science and science education.

SMIRC members met four times for TIMSS 2019. At the 1st TIMSS 2019 SMIRC meeting in Amsterdam, The Netherlands (April 2017), the committee reviewed the mathematics and science content frameworks and initial drafts of the mathematics and science PSIs. At the 2nd meeting in Windsor, England (September 2017), SMIRC reviewed draft field test items, together with the scoring guides for constructed response items. At the 3rd meeting in Tromsø, Norway (July 2018), SMIRC reviewed field test results and made recommendations regarding the items to include in the TIMSS 2019 mathematics and science assessments. At the final meeting in Singapore (May 2020), SMIRC conducted the TIMSS 2019 scale anchoring process (see Using Scale Anchoring to Interpret the TIMSS 2019 Achievement Scales). Exhibit 1.1 lists the TIMSS 2019 SMIRC members.

Exhibit 1.1: TIMSS 2019 Science and Mathematics Item Review Committee (SMIRC)

Mathematics

Ray Philpot
Australian Council for Educational Research (ACER)
Australia

Kiril Bankov
Faculty of Mathematics and Informatics, University of Sofia
Bulgaria

Khattab Mohammad Ahmad Abulibdeh
National Center for Human Resources Development
Jordan

Arne Hole
Department of Teacher Education and School Research, University of Oslo
Norway

Cheow Kian Soh
Ministry of Education, Curriculum Planning, and Development Division, Mathematics Branch
Singapore

Mary Lindquist
Professor Emeritus Mathematics Education, Columbus State University
United States

Linda Hall
Mathematics Consultant
United States

Science

Svatava Janoušková
Science Faculty Department of Teaching and Didactics of Chemistry, Charles University Prague
Czech Republic

Emily Jones
National Foundation of Educational Research (NFER)
England

Jouni Viiri
Department of Teacher Education, University of Jyväskylä
Finland

Berenice Michels
Faculty of Science, Freudenthal Institute for Science and Mathematics Education
The Netherlands

Galina Kovaleva
Federal Institute for Strategy of Education Development of the Russian Academy of Education Center for Evaluating the Quality of Education
Russian Federation

Christopher Lazzaro
The College Board
United States

Developing the PSIs and technology-enhanced achievement items to meet the ambitious development goals for eTIMSS 2019 necessitated even more expert review and collaboration than previous TIMSS cycles. Several SMIRC members worked closely with staff at the TIMSS & PIRLS International Study Center throughout the development process to achieve these goals. For mathematics, Mary Lindquist and Ray Philpot provided additional subject-matter expertise and support. For science, Emily Jones, Christopher Lazzaro, and Berenice Michels served in this capacity.

The TIMSS 2019 Development Schedule

In preparation for the transition to eTIMSS, development work for TIMSS 2019 began over three years before the TIMSS 2019 Field Test and included a series of novel activities to develop the eTIMSS user interface, eAssessment System, and PSIs. Essentially, the first two years were devoted to updating the assessment frameworks and pilot testing the mathematics and science PSIs and trend items in digital format. The third year was dedicated to writing new achievement items in both digital and paper format, continuing to refine the PSIs, and testing components of the eAssessment System to ensure successful delivery of eTIMSS across a variety of digital devices and testing conditions.

The TIMSS 2019 Field Test was conducted from March through May 2018. After a thorough review of the results, the materials for data collection were finalized in August 2018. TIMSS 2019 Data Collection began in the Southern Hemisphere in September 2018 and continued in the Northern Hemisphere through May 2019.

Exhibit 1.2 shows the TIMSS 2019 development schedule for the achievement items beginning with initial work on the eAssessment System through TIMSS 2019 Data Collection.

Exhibit 1.2: TIMSS 2019 Development Schedule for Achievement Items

Date(s)		Group and Activity
January	2015	TIMSS & PIRLS International Study Center and IEA Hamburg began designing the eTIMSS assessment system, user interface, and digital item types, including the PSIs, in preparation for the transition to eTIMSS
March	2015	TIMSS & PIRLS International Study Center began work with members of the Science and Mathematics Item Review Committee (SMIRC), other external expert consultants, and IEA Hamburg to design and operationalize prototype PSIs
August	2015	Consultants and staff at the TIMSS & PIRLS International Study Center began drafting additional PSIs for both the fourth and eighth grade assessments (Boston, USA)
August	2015	American Institutes for Research (AIR) conducted cognitive laboratories for two prototype PSIs (one fourth grade mathematics and one eighth grade science) and a sample of TIMSS trend items converted to digital format
October	2015	Consultants and staff at the TIMSS & PIRLS International Study Center reviewed the results of the cognitive laboratories, continued revising the draft PSIs, and drafted new PSIs (Boston, USA)
June	2016	TIMSS & PIRLS International Study Center presented an informational video introducing the features of the eTIMSS assessments and debuting the PSIs to National Research Coordinators (NRCs) (8^{th} NRC meeting—Quebec, Canada)
June–September	2016	TIMSS & PIRLS International Study Center conducted content analysis of the curricular topics described in the *TIMSS 2015 Encyclopedia* and proposed updates to the mathematics and science frameworks for TIMSS 2019
September	2016	SMIRC reviewed the draft *TIMSS 2019 Assessment Frameworks* and provided feedback to staff at the TIMSS & PIRLS International Study Center. The staff then met with SMRIC consultants to incorporate SMIRC's comments (Boston, USA)

Exhibit 1.2: TIMSS 2019 Development Schedule for Achievement Items (continued)

Date(s)		Group and Activity
October	2016	Australia, Canada, and Singapore administered the eTIMSS prePilot, which included a sample of trend items converted to digital format and draft PSIs
November	2016	Consultants and staff at the TIMSS & PIRLS International Study Center reviewed the results of the eTIMSS prePilot and revised the PSIs and user interface specifications based on these results. The group also drafted one additional PSI for each grade, fulfilling the development requirements for the eTIMSS 2019 Field Test (Boston, USA)
February	2017	NRCs reviewed the draft *TIMSS 2019 Assessment Frameworks* (1st NRC meeting—Hamburg, Germany). Following the meeting, NRCs completed an online survey to provide feedback as to whether each topic area should be kept as is, modified, or deleted
March–April	2017	TIMSS & PIRLS International Study Center prepared draft *TIMSS 2019 Item Writing Guidelines*, including specific guidelines for the enhanced item formats available for eTIMSS. Staff also revised the draft *TIMSS 2019 Assessment Frameworks* based on feedback from NRCs
April	2017	SMIRC reviewed the draft *TIMSS 2019 Assessment Frameworks*, *TIMSS 2019 Item Writing Guidelines*, and PSIs (1st TIMSS 2019 SMIRC meeting—Amsterdam, The Netherlands)
May	2017	NRCs reviewed the *TIMSS 2019 Assessment Frameworks* and developed draft field test items and scoring guides using the *TIMSS 2019 Item Writing Guidelines* (2nd NRC meeting—Hamburg, Germany)
May	2017	The eTIMSS Pilot/Item Equivalence Study, designed to investigate mode effects for the TIMSS trend items, was conducted to provide information about the robustness of the eAssessment System and countries' readiness to conduct a digital assessment
July	2017	Consultants and staff at the TIMSS & PIRLS International Study Center reviewed and revised draft field test items and scoring guides, including PSIs (Boston, USA)
September	2017	SMIRC reviewed the draft field test items and scoring guides, including PSIs (2nd SMIRC meeting—Windsor, England)
September	2017	Consultants and staff at the TIMSS & PIRLS International Study Center reviewed the updated field test items and PSIs and refined the scoring guides with special attention to machine scoring (Boston, USA)
November	2017	NRCs reviewed and approved the TIMSS 2019 Field Test instruments (3rd NRC meeting—Melbourne, Australia)
December	2017	TIMSS & PIRLS International Study Center and IEA Hamburg assembled all TIMSS 2019 Field Test instruments and released the international instruments to countries for translation
January–March	2018	TIMSS & PIRLS International Study Center and IEA Hamburg collaborated to establish specifications for eTIMSS data capture and machine-scored constructed response items
January	2018	TIMSS & PIRLS International Study Center collected student responses to constructed response items from English-speaking countries to develop scoring training materials for the field test
January	2018	Consultants and staff at the TIMSS & PIRLS International Study Center reviewed the field test scoring guides and prepared scorer training materials (Boston, USA)
March–May	2018	Countries conducted the TIMSS 2019 Field Test

Exhibit 1.2: **TIMSS 2019 Development Schedule for Achievement Items (continued)**

Date(s)		Group and Activity
March	2018	NRCs received scoring training for constructed response field test items (4th NRC meeting—Madrid, Spain)
May	2018	Countries submitted TIMSS 2019 Field Test achievement data for analysis and review
May	2018	NRCs provided feedback to the TIMSS & PIRLS International Study Center about the field-tested PSIs. Based on the NRC's evaluations, the TIMSS & PIRLS International Study Center selected the PSIs to move forward to eTIMSS 2019 Data Collection and began editing the tasks based on NRC feedback
June	2018	IEA Hamburg completed data processing and TIMSS & PIRLS International Study Center completed scoring of machine-scored items
June	2018	TIMSS & PIRLS International Study Center reviewed the field test item statistics and assembled sets of proposed items for data collection
July	2018	SMIRC reviewed the proposed items for data collection in conjunction with the field test results (3rd SMIRC meeting—Tromsø, Norway)
August	2018	NRCs reviewed and approved the proposed item blocks for TIMSS 2019 Data Collection (5th NRC meeting—Stockholm, Sweden)
September	2018	TIMSS & PIRLS International Study Center and IEA Hamburg finalized all TIMSS 2019 Data Collection instruments and released the international instruments to countries for translation
September–December	2018	Southern Hemisphere countries conducted TIMSS 2019 data collection
September	2018	Consultants and staff at the TIMSS & PIRLS International Study Center reviewed and updated scoring guides and scorer training materials (Boston, USA)
November	2018	NRCs from Southern Hemisphere countries received scoring training for constructed response items (Cape Town, South Africa)
November	2018	TIMSS & PIRLS International Study Center finalized scoring guides and training materials for constructed response items and distributed them to NRCs from Southern Hemisphere countries
March	2019	NRCs from Northern Hemisphere countries received scoring training for constructed response items (Limassol, Cyprus)
March–June	2019	Northern Hemisphere countries conducted TIMSS 2019 data collection

Updating the Assessment Frameworks for TIMSS 2019

The first step in developing the TIMSS achievement instruments is to define and prioritize the mathematics and science content and skills that the assessment will measure. The assessment frameworks cannot drastically change from cycle to cycle, but are routinely updated to keep up with fresh ideas and current information about curricula, standards, and instruction in mathematics and science education around the world. The first two chapters of the *TIMSS 2019 Assessment Frameworks* (Mullis & Martin, 2017), respectively, describe the mathematics and science frameworks in detail.

Writing and Reviewing the TIMSS 2019 Field Test Items and Scoring Guides

The TIMSS 2019 Field Test included approximately one and a half times the number of achievement items needed for data collection, to ensure a sufficient number of high quality items for the TIMSS 2019 assessments. In all, about 800 items were field tested. With the exception of the PSIs (eTIMSS only) and less difficult mathematics items (paper only), all items were prepared and administered in both digital and paper format. These items were designed to be identical in content across eTIMSS and paperTIMSS, with the only difference being the response mode (e.g., a drag and drop item in eTIMSS may be a matching item in paperTIMSS).

The TIMSS & PIRLS International Study Center uses a collaborative process involving the participating countries to develop the substantial number of new items and scoring guides needed for the field test. Most of the 2nd TIMSS 2019 NRC meeting in Hamburg, Germany was devoted to a workshop for developing the field test items. The NRCs, together with experienced item writers from participating countries and staff from the TIMSS & PIRLS International Study Center, drafted the majority of the new items for the mathematics and science field tests during this workshop.

In preparation for the item writing workshop, staff at the TIMSS & PIRLS International Study Center identified the scope of the item writing task for the field test. Considerations included the total items needed based on the weight assigned to a particular topic in the *TIMSS 2019 Assessment Frameworks* (Mullis & Martin, 2017), as well as how many items existed from previous assessments. The TIMSS & PIRLS International Study Center also updated the item writing manual specifically developed for TIMSS assessments. The manual contains general information about procedures for obtaining good measurement of mathematics and science achievement (e.g., items must be independent and not provide clues to the correct responses of other items), as well as specific information on how to deal with translation and comparability issues (e.g., using TIMSS' fictitious unit of currency, the "zed," for items involving money). The manual also includes the necessary steps for developing scoring guides for constructed response items, as well as checklists for reviewing TIMSS items.

Updated for the transition to eTIMSS, the *TIMSS 2019 Item Writing Guidelines* provided additional instructions for taking advantage of the technology-enhanced item formats—drag and drop, sorting, selection, drop-down menus, and a line drawing tool. These guidelines included examples of how each enhanced item format might be used (e.g., using drag and drop for adding labels to graphs or diagrams) and some details about the functionality of the formats (e.g., the maximum number of "draggable" parts available in a drag and drop item).

At the TIMSS 2019 Item Writing Workshop, country representatives were divided into teams and given specific item writing assignments based on their areas of expertise to ensure that enough field test items were developed for each of the content areas and cognitive processes specified in the frameworks.

Staff from the TIMSS & PIRLS International Study Center used the *TIMSS 2019 Item Writing Guidelines* to provide training to the teams on item writing procedures. The teams were asked to provide a complete draft of each item they developed, including the content topic and cognitive area from the framework that the item addressed and the information needed to score the item (i.e., an answer key for selected response items or scoring guide for constructed response items). Once teams had completed their own item writing assignments, they reviewed the items drafted by other teams. In addition, some teams continued to send items to the TIMSS & PIRLS International Study Center for several weeks after the Item Writing Workshop.

Exhibit 1.3 shows the number of participants in the TIMSS 2019 Item Writing Workshop and the approximate number of items written.

Exhibit 1.3: TIMSS 2019 Item Writing Workshop to Develop Field Test Items

Participants	
Number of Countries and Benchmarking Entities	53
Number of Country Representatives	118
Approximate Number of Field Test Items Written at Item Writing Workshop	
Fourth Grade Mathematics	300
Fourth Grade Science	200
Eighth Grade Mathematics	300
Eighth Grade Science	200

Following the item writing workshop, staff at the TIMSS & PIRLS International Study Center reviewed each item in light of the framework specifications and selected an optimal group of items for further review and revision. Consultants from the Australian Council for Educational Research (ACER) and the National Foundation of Educational Research (NFER) drafted additional mathematics and science items, respectively, to improve coverage of areas of the frameworks that are especially challenging to measure.

In July 2017, several SMIRC members with particular item writing skills met with staff from the TIMSS & PIRLS International Study Center to continue revising the draft field test items. SMIRC then reviewed all of the proposed draft field test items at the 2nd TIMSS 2019 SMIRC meeting. After SMIRC's review, the items were revised again, and the NRCs reviewed the complete set of draft field test items at the 3rd TIMSS 2019 NRC meeting in Melbourne, Australia. Following this meeting, staff at the TIMSS & PIRLS International Study Center implemented the final suggested revisions and provided the international versions of the field test instruments in digital or paper format to the NRCs so that they could begin translating the field test materials into their languages of instruction.

Preparing eTIMSS Field Test Items for Digital Delivery

Preparing the eTIMSS field test items for digital delivery required the additional step of entering each item into IEA's Item Builder, a web-based application for creating digital achievement items and instruments for delivery to students via computers and tablets. For eTIMSS 2019, the Item Builder included templates for both traditional (e.g., standard multiple-choice) and enhanced (e.g., drag and drop) item formats as well as a variety of tools for designing the items, such as features for uploading and adding text to images, creating tables, and previewing items as they would appear to students during the field test. After drafting and reviewing the field test items on paper, staff at the TIMSS & PIRLS International Study Center entered all eTIMSS items into the Item Builder and collaborated with IEA Hamburg to conduct extensive quality control tests to ensure each item would appear and function as intended for students.

Developing Problem Solving and Inquiry (PSIs) Tasks for eTIMSS

In many ways, PSI development work followed the standard TIMSS procedures for ensuring that the items provide valid measurement of the *TIMSS 2019 Assessment Frameworks* (Mullis & Martin, 2017). However, because the PSIs involved a new and more innovative approach to assessing mathematics and science achievement in a digital environment, PSI development required additional efforts.

Developing engaging problem contexts with cohesive sets of achievement items necessitated even more rounds of expert review than is typical for TIMSS items. Staff at the TIMSS & PIRLS International Study Center began collaborating with SMIRC members to develop the PSIs in March 2015, nearly two years before item writing for the rest of the TIMSS 2019 items began. Several SMIRC members worked closely with TIMSS staff to develop the PSIs, which included providing initial ideas for the tasks and participating in a series of meetings to develop and refine the problem contexts, items, and scoring guides.

Leading up to the field test, several SMIRC members and staff at the TIMSS & PIRLS International Study Center met a total of five times at Boston College and conducted many online reviews to refine the PSIs. SMIRC as a whole conducted its first in-depth review of the PSIs at the 1st TIMSS 2019 SMIRC meeting, which focused on the alignment between the tasks and the frameworks, the extent to which the technology in the tasks supported the intended response processes, and the cross-cultural appropriateness of the problem scenarios. The NRCs reviewed the PSIs prior to the field test at the 3rd TIMSS 2019 NRC meeting.

In addition to extensive expert review, cognitive laboratories and a pilot test were conducted in several eTIMSS countries in advance of the field test to gain insight into students' interactions with the PSIs and to test the functionality of the eAssessment System. This strand of development work provided critical information about the usability of innovative item types and the eTIMSS interface, the amount of time it took students to complete each task, and the approximate difficulty of the tasks. Following each outing and review, improvements were made to both the PSIs and their software with the aim of eliciting the intended types of responses from students.

eTIMSS Cognitive Laboratories

Staff at the TIMSS & PIRLS International Study Center partnered with the American Institutes for Research (AIR) to conduct cognitive laboratories in the very early stages of the transition to eTIMSS (August 2015). The goal of this study was to investigate two aspects of digital assessment that would inform next steps in eTIMSS development: students' interactions with drafts of the first PSIs, and students' experiences with the eTIMSS interface.

The TIMSS & PIRLS International Study Center prepared two prototype PSIs and a set of TIMSS trend items in digital format at each grade, along with a list of research questions, from which AIR developed interview protocols. During the interviews, students explained their thoughts while engaging with the items on tablets, providing insight into how the PSI format and eTIMSS interface could be improved.

AIR conducted the interviews with a purposive sample of 32 fourth and eighth grade students from the greater Washington, D.C. area. Following the interviews, AIR prepared a report to address each of the TIMSS & PIRLS International Center's research questions. The reports from the cognitive laboratories prompted substantial revisions to the PSI item format and the eTIMSS interface. In particular, the students reported difficulties in using a stylus to write or draw, so the device keyboards were enabled for items requiring a written response and a new tool for drawing lines was developed.

eTIMSS prePilot

The eTIMSS prePilot was conducted in September 2016 to collect more information on students' interactions with the draft PSIs and eTIMSS interface in a standard testing situation. The prePilot instruments included a total of 12 PSI tasks across both subjects and grades and incorporated a broader variety of interactive features and item types than the first prototypes. The instruments also were designed to be administered on both computers and tablets to accommodate a wider range of devices and support more countries' participation in eTIMSS.

The eTIMSS prePilot was conducted in three English-speaking countries with experience in conducting digital assessments: Australia, Canada, and Singapore. Each country selected two to four classes at each grade to participate and made efforts to include students with a range of mathematics and science ability. This sample yielded approximately 100 responses per item at both the fourth and the eighth grade.

Students' responses to the draft PSIs and participating countries' reports on their experiences carrying out the study prompted additional changes to both the PSIs and their software before the field test.

The TIMSS 2019 Field Test

In preparation for data collection, TIMSS routinely conducts a full-scale field test for the purposes of evaluating the measurement properties of the item pool and practicing the data collection and scoring procedures. For TIMSS 2019, the field test was a particularly critical "dress rehearsal" because it was the first large-scale administration of eTIMSS on computers and tablets. In addition to providing important information about how well each prospective item and PSI functioned, the field test results prompted a number of improvements to the components in the eAssessment System as well as to the directions for test administrators and students.

All eTIMSS and paperTIMSS materials and operational procedures were field tested with samples of students selected according to rigorous sampling procedures. The field test in each country was designed to be conducted in approximately 30 schools and yield at least 200 student responses to each mathematics and science item. The school samples for the TIMSS 2019 Field Test and Data Collection were drawn simultaneously, using the same random sampling procedures. This ensured that the field test samples closely approximated the data collection samples, and that a school was selected for either the field test or data collection, but not both. For example, if a country needed 150 schools for data collection and another 30 for the field test, then a larger sample of 180 schools was selected and a systematic sample of 30 schools was selected for the field test from the 180 schools. See Chapter 3 for details about the school and classroom sampling techniques used in TIMSS 2019.

Preparing for the eTIMSS 2019 Field Test was quite complicated and involved several additional steps beyond those included in paperTIMSS. After translating and adapting the international instruments in IEA's online translation system, countries checked the functionality of their national instruments, loaded the eTIMSS Player software onto each computer or tablet to be used in the field test, and checked the compatibility of the software with the devices. Following each testing session, test administrators uploaded students' responses to IEA's servers.

Exhibit 1.4 shows the total number of items in each fourth and eighth grade field test, as well as the number of students, teachers, and schools that participated. Exhibits 1.5 through 1.8 provide a detailed summary of the number of field test items in the eTIMSS and paperTIMSS field tests by format, content domain, and cognitive domain.

Exhibit 1.4: Overview of the TIMSS 2019 Field Test

	Fourth Grade			Eighth Grade	
	eTIMSS	paperTIMSS	Less Difficult Mathematics	eTIMSS	paperTIMSS
Items in Field Test					
Mathematics	174	127	130	201	158
Science	164	134	134	212	167
Total	**338**	**261**	**264**	**413**	**325**
Responses per Item per Country (approx.)	200	200	200	200	200
Participants					
Countries	31	18	7	22	14
Benchmarking Entities	6	–	–	5	–
Students	50,158	19,656	8,128	37,512	16,225
Teachers	3,337	1,176	471	5,009	1,826
Schools	1,340	526	203	852	342

Counts for eTIMSS include the items from the PSI tasks.
Five item blocks (64 items) were common to both the regular and less difficult fourth grade mathematics assessment.

Exhibit 1.5: TIMSS 2019 Number of Field Test Items by Content Domain and Item Format – Fourth Grade

Content Domain	Number of Selected Response Items	Number of Constructed Response Items	Total Number of Items	Percentage of Total Items
Mathematics – eTIMSS and paperTIMSS				
Number	27	25	52	40%
Measurement and Geometry	22	20	42	32%
Data	15	22	37	28%
Total	**64**	**67**	**131**	
Mathematics – Less Difficult				
Number	31	24	55	42%
Measurement and Geometry	21	17	38	29%
Data	17	20	37	28%
Total	**69**	**61**	**130**	
Science – eTIMSS and paperTIMSS				
Life Science	40	22	62	46%
Physical Science	28	13	41	31%
Earth Science	18	13	31	23%
Total	**86**	**48**	**134**	

Four mathematics items were only field tested in eTIMSS and four items were only field tested in paperTIMSS. Counts include all eight of these items.
Five item blocks (64 items) were common to both the regular and less difficult fourth grade mathematics assessments.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Exhibit 1.6: TIMSS 2019 Number of Field Test Items by Cognitive Domain and Item Format – Fourth Grade

Cognitive Domain	Number of Selected Response Items	Number of Constructed Response Items	Total Number of Items	Percentage of Total Items
Mathematics – eTIMSS and paperTIMSS				
Knowing	29	14	43	33%
Applying	27	38	65	50%
Reasoning	8	15	23	18%
Total	**64**	**67**	**131**	
Mathematics – Less Difficult				
Knowing	36	11	47	36%
Applying	25	30	55	42%
Reasoning	8	20	28	22%
Total	**69**	**61**	**130**	
Science – eTIMSS and paperTIMSS				
Knowing	42	16	58	43%
Applying	28	17	45	34%
Reasoning	16	15	31	23%
Total	**86**	**48**	**134**	

Four mathematics items were only field tested in eTIMSS and four items were only field tested in paperTIMSS. Counts include all eight of these items.

Five item blocks (64 items) were common to both the regular and less difficult fourth grade mathematics assessments.

Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Exhibit 1.7: TIMSS 2019 Number of Field Test Items by Content Domain and Item Format – Eighth Grade

Content Domain	Number of Selected Response Items	Number of Constructed Response Items	Total Number of Items	Percentage of Total Items
Mathematics – eTIMSS and paperTIMSS				
Number	18	28	46	29%
Algebra	25	28	53	34%
Geometry	9	22	31	20%
Data and Probability	14	14	28	18%
Total	**66**	**92**	**158**	
Science – eTIMSS and paperTIMSS				
Biology	42	19	61	36%
Chemistry	22	16	38	23%
Physics	24	13	37	22%
Earth Science	24	7	31	19%
Total	**112**	**55**	**167**	

Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Exhibit 1.8: TIMSS 2019 Number of Field Test Items by Cognitive Domain and Item Format – Eighth Grade

Cognitive Domain	Number of Selected Response Items	Number of Constructed Response Items	Total Number of Items	Percentage of Total Items
Mathematics – eTIMSS and paperTIMSS				
Knowing	28	19	47	30%
Applying	32	46	78	49%
Reasoning	6	27	33	21%
Total	**66**	**92**	**158**	
Science – eTIMSS and paperTIMSS				
Knowing	46	12	58	35%
Applying	39	23	62	37%
Reasoning	27	20	47	28%
Total	**112**	**55**	**167**	

Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

The eTIMSS 2019 Field Test also included eight mathematics and science PSI tasks at the fourth grade, comprising 72 items, and seven mathematics and science PSI tasks at the eighth grade, comprising 83 items. Because the PSIs were designed with the distinct goals of increasing coverage of traditionally difficult to measure areas of the mathematics and science frameworks in the applying and reasoning cognitive domains by capitalizing on technology, choices about the content topics to assess with each task were largely guided by the problem contexts and potential uses of technology to enhance measurement. Following the field test, two-thirds of the PSI tasks were selected for data collection (see Exhibit 1.18 for a description of the selected tasks).

Developing the Materials for TIMSS 2019 Field Test Scoring Training

To ensure the quality of the TIMSS assessment results, it is critical that students' responses to the achievement items demonstrate the knowledge, application, or reasoning in mathematics or science required by the item to receive credit. It also is critical that students' responses are evaluated consistently to enable comparisons of students' mathematics and science achievement across countries and over time. For these reasons, TIMSS expends considerable effort to ensure the validity and reliability of the scores assigned to students' responses to the TIMSS achievement items.

In addition to developing a unique scoring guide for each constructed response item, the TIMSS & PIRLS International Study Center provided training for the NRCs and their scoring supervisors to ensure that the scoring guides for all human-scored constructed response items were applied consistently within and across countries. The TIMSS 2019 training materials consisted of sets of student responses for a selected group of items with the most complicated scoring guides. For each item, the training set consisted of 8 to 12 student responses illustrating the codes in the scoring guide (example responses) followed by 8 to 12 student responses without pre-assigned score codes (practice responses).

To allow for field test scoring to begin immediately upon completion of data collection, it was necessary to prepare scoring training materials for the newly developed constructed response items in advance of the field test. To provide "grist" for these materials, Australia, England, and Ireland pilot tested a selection of the newly developed constructed response field test items in several classrooms with English-speaking students in January 2018. Because students may express their answers in different ways when typing versus writing by hand, both typed and handwritten responses were collected for the all items in both the eTIMSS and paperTIMSS assessments.

Exhibit 1.9 provides the number of items included in the pilot test and the number of student responses collected. Only a small number of mathematics items required scoring training, so the majority of the items in the pilot were in science.

Exhibit 1.9: **Pilot Test Student Responses for Field Test Scoring Training Materials Development**

	Number of Items	Approximate Number of Responses	
		eTIMSS	paperTIMSS
Fourth Grade			
Mathematics	5	93	96
Science	21	93	96
Countries		England	Australia and Ireland
Eighth Grade			
Mathematics	6	80	43
Science	19	80	43
Countries		England	Ireland

Consultants and staff at the TIMSS & PIRLS International Study Center met in January 2018 to review responses collected in the pilot test and create the training materials. For the TIMSS 2019 Field Test, training sets of example and practice responses were created for a total of 23 fourth grade items and 30 eighth grade items. These sets included both typed and handwritten responses to prepare scorers to score student responses in both modes of administration.

The TIMSS 2019 NRCs and their scoring supervisors received scoring training for the field test in March 2018 in Madrid, Spain, as part of the 4th TIMSS 2019 NRC meeting. At the training sessions, the trainers explained the purpose of each item and read it aloud. The trainer then described the scoring guide, explaining each category and the rationale for the score given to each example paper. The country representatives were then given time to score the practice papers so they could apply the scoring guides and learn how to make distinctions among categories. The correct codes for each practice paper were then reviewed, any inconsistencies in scoring were discussed, and, as necessary, the scoring guides were clarified and sometimes categories were revised.

Finalizing the TIMSS 2019 Achievement Instruments

Subsequent to the field test, the TIMSS & PIRLS International Study Center analyzed the TIMSS field test data and selected the new items to be combined with the trend items for data collection. When selecting the items, both the measurement properties (item statistics) of the individual items and the overall content and cognitive domain coverage of the group of items were considered to ensure that the final achievement instruments met the assessment specifications in the frameworks.

To review the measurement properties of the field test items, staff at the TIMSS & PIRLS International Study Center prepared almanacs containing summary item statistics for each field test item. The achievement data almanacs displayed for each item, row by row for each country: the number of students to whom the item was administered, the item difficulty and discrimination, the percentage of students answering each option (selected response) or in each score category (constructed response), the point-biserial correlation for each selected response option or constructed response category, and the degree of scoring agreement for human-scored constructed response items. The field test data were used by the TIMSS & PIRLS International Study Center, expert committees, and NRCs to assess the quality of the field test items.

First, staff at the TIMSS & PIRLS International Study Center reviewed the field test data to make an initial judgment about the quality of each item based on its measurement properties. Items were eliminated from further consideration if they had poor measurement properties, such as being too difficult or too easy or having low discrimination. Particular attention was paid to unusual item statistics in individual countries because these could indicate errors in translation.

After the item-by-item review, staff at the TIMSS & PIRLS International Study Center collaborated with a subset of SMIRC members to choose a set of recommended achievement items. The group reviewed the viable field test items for each content domain topic in relation to the trend items to select a coherent group of items for each topic, then verified that the set of items were appropriately distributed across the cognitive domains and item formats. SMIRC scrutinized the recommendations for the newly developed achievement items at the 3rd TIMSS 2019 SMIRC meeting, reviewing the items and scoring guides for content accuracy, clarity, and adherence to the frameworks.

To allow for any major revisions to the PSIs to be completed in time for data collection, the NRCs were asked to provide feedback on the PSIs when they submitted their field test data. Staff at the TIMSS & PIRLS International Study Center reviewed all NRC comments in conjunction with the data, selected the PSIs for the eTIMSS 2019 assessments based on the NRCs' recommendations, and began editing the selected tasks in June 2018. SMIRC also reviewed the PSIs at their 3rd meeting.

Next, staff at the TIMSS & PIRLS International Study Center implemented SMIRC's recommendations and assembled the items into assessment blocks for the NRCs' penultimate review. The NRCs had the opportunity to review the recommended assessment blocks in light of the field test results and within the security of their own countries. Each country also could check any unusual national results that might be indicative of translation errors and correct the translation as necessary or recommend revisions to better accommodate translation. Finally, the 5th TIMSS 2019 NRC meeting held in Stockholm, Sweden, in August 2018 was devoted to reviewing all the newly developed items.

Following the final review, the newly developed item blocks and existing trend item blocks were arranged into digital block combinations for eTIMSS and booklets for paperTIMSS according to the TIMSS 2019 Assessment Design (Martin, Mullis & Foy, 2017). For eTIMSS, the trend item blocks were converted from paper to digital format to be administered via the eAssessment System along with the new item blocks. The results of the TIMSS 2019 Item Equivalence Study (Fishbein, Martin, Mullis, &

Foy, 2018), a pilot conducted in 25 eTIMSS countries to investigate potential differences in student achievement on the trend items between the paper and digital modes of administration, provided evidence that the mathematics and science constructs assessed by the trend items were mostly unaffected in the transition to eTIMSS at both grades. Still, to ensure that the eTIMSS and paperTIMSS results could be reported on the same achievement scale, eTIMSS 2019 countries that had participated in TIMSS 2015 also re-administered the trend items in paper booklets to a nationally representative sample of students during data collection to provide a "bridge" between paperTIMSS and eTIMSS (see Chapter 12 for additional details).

Distribution of the TIMSS 2019 Achievement Items

It is critical to document the coherence between the assessment frameworks and achievement instruments to ensure that an assessment measures what it is intended to measure and provide evidence for the validity of the assessment results. Because the TIMSS assessments encompass two domains (content and cognitive) and include both trend and newly developed items in a variety of formats, it is necessary to demonstrate the alignment between the items and assessment specifications from multiple perspectives.

Achievement Items by Content and Cognitive Domain

The TIMSS 2019 assessments consisted of approximately 40 percent new items and 60 percent trend items, which were used to continue trend measurement from the previous assessment cycles. Therefore, it is important to confirm that the distribution of both the trend and new items across the content and cognitive domains reflects the specifications described in the assessment frameworks. The distribution of the trend items typically varies from the target specifications because the assessment frameworks are updated with each cycle and items are "retired" from the assessment, so the new items are selected to ensure the final assessments are aligned with the frameworks.

Exhibits 1.10 and 1.11 present the number of trend and newly developed items as well as the number of score points in the TIMSS 2019 fourth grade assessments by content domain and cognitive domain, respectively. The number of items represents the number of distinct questions in the assessment, while the number of score points represents the complexity and weight given to each item. Exhibits 1.12 and 1.13 present the TIMSS 2019 eighth grade assessments by content and cognitive domain.

Exhibit 1.10: **TIMSS 2019 Achievement Items by Content Domain – Fourth Grade**

Content Domain	Trend		New		Total		Target Percentage of Score Points
	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points	
Mathematics – eTIMSS and paperTIMSS							
Number	55 (59)	61%	29 (30)	32%	84 (89)	47%	50%
Measurement and Geometry	26 (27)	28%	27 (31)	33%	53 (58)	31%	30%
Data	11 (11)	11%	27 (32)	34%	38 (43)	23%	20%
Total	**92 (97)**		**83 (93)**		**175 (190)**		
Mathematics – Less Difficult							
Number	67 (68)	59%	29 (32)	42%	96 (100)	52%	50%
Measurement and Geometry	31 (34)	29%	20 (21)	28%	51 (55)	29%	30%
Data	13 (14)	12%	19 (23)	30%	32 (37)	19%	20%
Total	**111 (116)**		**68 (76)**		**179 (192)**		
Science – eTIMSS and paperTIMSS							
Life Science	44 (47)	46%	34 (36)	46%	78 (83)	46%	45%
Physical Science	36 (37)	36%	26 (26)	33%	62 (63)	35%	35%
Earth Science	18 (18)	18%	17 (17)	22%	35 (35)	19%	20%
Total	**98 (102)**		**77 (79)**		**175 (181)**		

Score points are shown in parentheses.
Two mathematics items involving an on-screen ruler tool were only included in eTIMSS assessment.
Four item blocks (48 items) were common to both the regular and less difficult fourth grade mathematics assessments.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Exhibit 1.11: TIMSS 2019 Achievement Items by Cognitive Domain – Fourth Grade

Cognitive Domain	Trend		New		Total		Target Percentage of Score Points
	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points	
Mathematics – eTIMSS and paperTIMSS							
Knowing	34 (34)	35%	29 (29)	31%	63 (63)	33%	40%
Applying	40 (42)	43%	34 (39)	42%	74 (81)	43%	40%
Reasoning	18 (21)	22%	20 (25)	27%	38 (46)	24%	20%
Total	**92 (97)**		**83 (93)**		**175 (190)**		
Mathematics – Less Difficult							
Knowing	56 (56)	48%	25 (26)	34%	81 (82)	43%	40%
Applying	39 (40)	34%	27 (32)	42%	66 (72)	38%	40%
Reasoning	16 (20)	17%	16 (18)	24%	32 (38)	20%	20%
Total	**111 (116)**		**68 (76)**		**179 (192)**		
Science – eTIMSS and paperTIMSS							
Knowing	42 (45)	44%	31 (32)	41%	73 (77)	43%	40%
Applying	35 (36)	35%	30 (30)	38%	65 (66)	36%	40%
Reasoning	21 (21)	21%	16 (17)	22%	37 (38)	21%	20%
Total	**98 (102)**		**77 (79)**		**175 (181)**		

Score points are shown in parentheses.
Two mathematics items involving an on-screen ruler tool were only included in eTIMSS assessment.
Four item blocks (48 items) were common to both the regular and less difficult fourth grade mathematics assessments.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Exhibit 1.12: TIMSS 2019 Achievement Items by Content Domain – Eighth Grade

Content Domain	Trend		New		Total		Target Percentage of Score Points
	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points	
Mathematics – eTIMSS and paperTIMSS							
Number	36 (37)	30%	28 (30)	30%	64 (67)	30%	30%
Algebra	31 (32)	26%	31 (32)	32%	62 (64)	29%	30%
Geometry	25 (28)	22%	18 (20)	20%	43 (48)	21%	20%
Data and Probability	25 (28)	22%	17 (17)	17%	42 (45)	20%	20%
Total	**117 (125)**		**94 (99)**		**211 (224)**		
Science – eTIMSS and paperTIMSS							
Biology	39 (48)	32%	38 (43)	39%	77 (91)	35%	35%
Chemistry	22 (23)	18%	22 (25)	22%	44 (48)	20%	20%
Physics	33 (33)	27%	22 (25)	22%	55 (58)	25%	25%
Earth Science	28 (29)	23%	16 (17)	16%	44 (46)	20%	20%
Total	**122 (133)**		**98 (110)**		**220 (243)**		

Score points are shown in parentheses.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Exhibit 1.13: TIMSS 2019 Achievement Items by Cognitive Domain – Eighth Grade

Cognitive Domain	Trend		New		Total		Target Percentage of Score Points
	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points	
Mathematics – eTIMSS and paperTIMSS							
Knowing	35 (35)	28%	30 (32)	32%	65 (67)	30%	35%
Applying	58 (61)	49%	39 (40)	40%	97 (101)	45%	40%
Reasoning	24 (29)	23%	25 (27)	27%	49 (56)	25%	25%
Total	**117 (125)**		**94 (99)**		**211 (224)**		
Science – eTIMSS and paperTIMSS							
Knowing	45 (50)	37%	35 (36)	36%	80 (86)	36%	35%
Applying	46 (50)	38%	36 (44)	37%	82 (94)	37%	35%
Reasoning	31 (33)	25%	27 (30)	28%	58 (63)	26%	30%
Total	**122 (133)**		**98 (110)**		**220 (243)**		

Score points are shown in parentheses.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Achievement Items by Item Formats within Content and Cognitive Domains

To assess the broad range of mathematics and science topics and skills described in the assessment frameworks, the TIMSS 2019 fourth and eighth grade assessments included a wide variety of selected response and constructed response items. Both the digital and paper versions of the TIMSS 2019 assessments included two general types of selected response items—single selection, in which students choose one of four response options, and multiple selection, in which students chose more than one option from a number of response options or made a series of selections to respond to a question. In eTIMSS, the answer options for some selected response items were presented in drop-down menus or as clickable pictures or words. Most TIMSS 2019 selected response items were worth one score point, although some multiple selection items were worth two score points. The 2-point multiple selection items were scored as fully correct (all parts answered correctly; 2 score points), partially correct (most parts answered correctly; 1 score point), or incorrect (few or no parts answered correctly; 0 score points).

Constructed response items, which involve writing or typing words or numbers, drawing, or dragging and dropping for eTIMSS, were worth one or two score points depending on the degree of complexity involved. The 1-point constructed response items were scored as correct (1 score point) or incorrect (0 score points), whereas 2-point constructed response items were scored as fully correct (2 score points), partially correct (1 score point), or incorrect (0 score points). Fully correct responses show a complete or deeper understanding of a task while partially correct responses demonstrate only a partial understanding of the concepts or procedures embodied in the task.

To ensure sufficient coverage of the assessment frameworks, it is important to verify that an assortment of selected and constructed response items are used to assess each domain. Exhibits 1.14 through 1.17 display the number of items (and score points) by item format for each content and cognitive domain in the fourth and eighth grade assessments.

Exhibit 1.14: TIMSS 2019 Achievement Items by Content Domain and Item Format – Fourth Grade

Content Domain	Selected Response Items		Constructed Response Items		Total Items	Percentage of Score Points
	Single Selection	Multiple Selection	1 Point	2 Points		
Mathematics – eTIMSS and paperTIMSS						
Number	39 (39)	7 (7)	33 (33)	5 (10)	84 (89)	47%
Measurement and Geometry	25 (25)	6 (6)	17 (17)	5 (10)	53 (58)	31%
Data	8 (8)	3 (3)	22 (22)	5 (10)	38 (43)	23%
Total	**72 (72)**	**16 (16)**	**72 (72)**	**15 (30)**	**175 (190)**	
Achieved Percentage of Score Points	**46%**		**54%**			
Mathematics – Less Difficult						
Number	46 (46)	1 (1)	45 (45)	4 (8)	96 (100)	52%
Measurement and Geometry	26 (26)	2 (2)	19 (19)	4 (8)	51 (55)	29%
Data	10 (10)	2 (3)	16 (16)	4 (8)	32 (37)	19%
Total	**82 (82)**	**5 (6)**	**80 (80)**	**12 (24)**	**179 (192)**	
Achieved Percentage of Score Points	**46%**		**54%**			
Science – eTIMSS and paperTIMSS						
Life Science	35 (35)	6 (7)	33 (33)	4 (8)	78 (83)	46%
Physical Science	35 (35)	5 (5)	21 (21)	1 (2)	62 (63)	35%
Earth Science	24 (24)	4 (4)	7 (7)	– –	35 (35)	19%
Total	**94 (94)**	**15 (16)**	**61 (61)**	**5 (10)**	**175 (181)**	
Achieved Percentage of Score Points	**61%**		**39%**			

Score points are shown in parentheses.
Two fourth grade mathematics items involving an on-screen ruler tool were only included in eTIMSS assessment.
Four item blocks (48 items) were common to both the regular and less difficult fourth grade mathematics assessments.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Exhibit 1.15: **TIMSS 2019 Achievement Items by Cognitive Domain and Item Format – Fourth Grade**

Cognitive Domain	Selected Response Items		Constructed Response Items		Total Items	Percentage of Score Points
	Single Selection	Multiple Selection	1 Point	2 Points		
Mathematics – eTIMSS and paperTIMSS						
Knowing	33 (33)	12 (12)	18 (18)	– –	63 (63)	33%
Applying	25 (25)	2 (2)	40 (40)	7 (14)	74 (81)	43%
Reasoning	14 (14)	2 (2)	14 (14)	8 (16)	38 (46)	24%
Total	**72 (72)**	**16 (16)**	**72 (72)**	**15 (30)**	**175 (190)**	
Achieved Percentage of Score Points	**46%**		**54%**			
Mathematics – Less Difficult						
Knowing	46 (46)	2 (2)	32 (32)	1 (2)	81 (82)	43%
Applying	25 (25)	2 (3)	34 (34)	5 (10)	66 (72)	38%
Reasoning	11 (11)	1 (1)	14 (14)	6 (12)	32 (38)	20%
Total	**82 (82)**	**5 (6)**	**80 (80)**	**12 (24)**	**179 (192)**	
Achieved Percentage of Score Points	**46%**		**54%**			
Science – eTIMSS and paperTIMSS						
Knowing	42 (42)	10 (11)	18 (18)	3 (6)	73 (77)	43%
Applying	33 (33)	3 (3)	28 (28)	1 (2)	65 (66)	36%
Reasoning	19 (19)	2 (2)	15 (15)	1 (2)	37 (38)	21%
Total	**94 (94)**	**15 (16)**	**61 (61)**	**5 (10)**	**175 (181)**	
Achieved Percentage of Score Points	**61%**		**39%**			

Score points are shown in parentheses.
Two fourth grade mathematics items involving an on-screen ruler tool were only included in eTIMSS assessment.
Four item blocks (48 items) were common to both the regular and less difficult fourth grade mathematics assessments.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Exhibit 1.16: TIMSS 2019 Achievement Items by Content Domain and Item Format – Eighth Grade

Content Domain	Selected Response Items		Constructed Response Items		Total Items	Percentage of Score Points
	Single Selection	Multiple Selection	1 Point	2 Points		
Mathematics – eTIMSS and paperTIMSS						
Number	27 (27)	4 (6)	32 (32)	1 (2)	64 (67)	30%
Algebra	32 (32)	1 (1)	27 (27)	2 (4)	62 (64)	29%
Geometry	15 (15)	2 (2)	21 (21)	5 (10)	43 (48)	21%
Data and Probability	18 (18)	5 (7)	18 (18)	1 (2)	42 (45)	20%
Total	**92 (92)**	**12 (16)**	**98 (98)**	**9 (18)**	**211 (224)**	
Achieved Percentage of Score Points	**48%**		**52%**			
Science – eTIMSS and paperTIMSS						
Biology	37 (37)	9 (12)	20 (20)	11 (22)	77 (91)	37%
Chemistry	19 (19)	4 (5)	18 (18)	3 (6)	44 (48)	20%
Physics	29 (29)	7 (7)	16 (16)	3 (6)	55 (58)	24%
Earth Science	30 (30)	4 (6)	10 (10)	– –	44 (46)	19%
Total	**115 (115)**	**24 (30)**	**64 (64)**	**17 (34)**	**220 (243)**	
Achieved Percentage of Score Points	**60%**		**40%**			

Score points are shown in parentheses.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

Exhibit 1.17: TIMSS 2019 Achievement Items by Cognitive Domain and Item Format – Eighth Grade

Cognitive Domain	Selected Response Items		Constructed Response Items		Total Items	Percentage of Score Points
	Single Selection	Multiple Selection	1 Point	2 Points		
Mathematics – eTIMSS and paperTIMSS						
Knowing	41 (41)	5 (7)	19 (19)	– –	65 (67)	30%
Applying	40 (40)	4 (4)	49 (49)	4 (8)	97 (101)	45%
Reasoning	11 (11)	3 (5)	30 (30)	5 (10)	49 (56)	25%
Total	**92 (92)**	**12 (16)**	**98 (98)**	**9 (18)**	**211 (224)**	
Achieved Percentage of Score Points	**48%**		**52%**			
Science – eTIMSS and paperTIMSS						
Knowing	56 (56)	11 (14)	10 (10)	3 (6)	80 (86)	35%
Applying	38 (38)	8 (10)	26 (26)	10 (20)	82 (94)	39%
Reasoning	21 (21)	5 (6)	28 (28)	4 (8)	58 (63)	26%
Total	**115 (115)**	**24 (30)**	**64 (64)**	**17 (34)**	**220 (243)**	
Achieved Percentage of Score Points	**60%**		**40%**			

Score points are shown in parentheses.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.

eTIMSS Problem Solving and Inquiry Tasks (PSIs) by Content and Cognitive Domain

Exhibit 1.18 provides a brief description of the eTIMSS 2019 PSI problem scenarios and the total number of items (and score points) in each task. The tasks covered a range of mathematics and science content domain topics and, consistent with the goal of the PSIs to assess higher-order skills, the majority of the items in the PSIs involved applying and reasoning.

Exhibit 1.18: eTIMSS 2019 Mathematics and Science Problem Solving and Inquiry Tasks (PSIs)

Fourth Grade PSIs	Total Items
Mathematics	
School Party – Students plan a party for a school by determining the price for tickets and the amount of food, drinks, and decorations to purchase for the party	11 (14)
Robots – Students use a robot that can follow input-output rules to solve mathematics problems and determine the robot's rules	6 (7)
Little Penguins – Students add information to a website about Little Penguins by solving a series of mathematics problems involving facts about penguins	12 (14)
Science	
Farm Investigation – Students carry out a virtual investigation to identify the farm animal responsible for eating garden plants	10 (16)
Sugar Experiment – Students design and carry out a virtual experiment to test which of three types of sugar dissolves fastest in water	9 (13)
Eighth Grade PSIs	**Total Items**
Mathematics	
Dinosaur Speed – Students use the relationships between foot length, leg height, and stride length to estimate how fast a dinosaur could run	12 (13)
Building – Students determine the dimensions of a shed to store equipment, including a barrel to collect rainwater	9 (11)
Robots – Students determine functions using a robot that applies a function to determine *y* for any given value of *x*	4 (4)
Science	
Sunken Ship – Students carry out a virtual investigation into the circumstances that resulted in the sinking of a ship	16 (17)
Pepper Plants – Students design and carry out a virtual experiment to test the effects of two fertilizers on the growth and development of pepper plants	13 (18)

Score points are shown in parentheses.

The addition of the PSIs for eTIMSS resulted in a slight increase in coverage of the applying and reasoning cognitive domains at both the fourth and the eighth grade. However, comprising only a small part of the whole assessment (approximately 12 percent), the PSIs did not substantially alter the framework coverage provided by the eTIMSS assessments. The pie charts in Exhibits 1.19 and 1.20 show the percentage of assessment score points in each content and cognitive domain in the eTIMSS 2019 assessments, both with and without the PSIs included, compared to the target percentage of testing time allocated to each domain.

Exhibit 1.19: Comparison of Target and Achieved Percentages of Domain Coverage in the eTIMSS 2019 Mathematics and Science Assessments – Fourth Grade

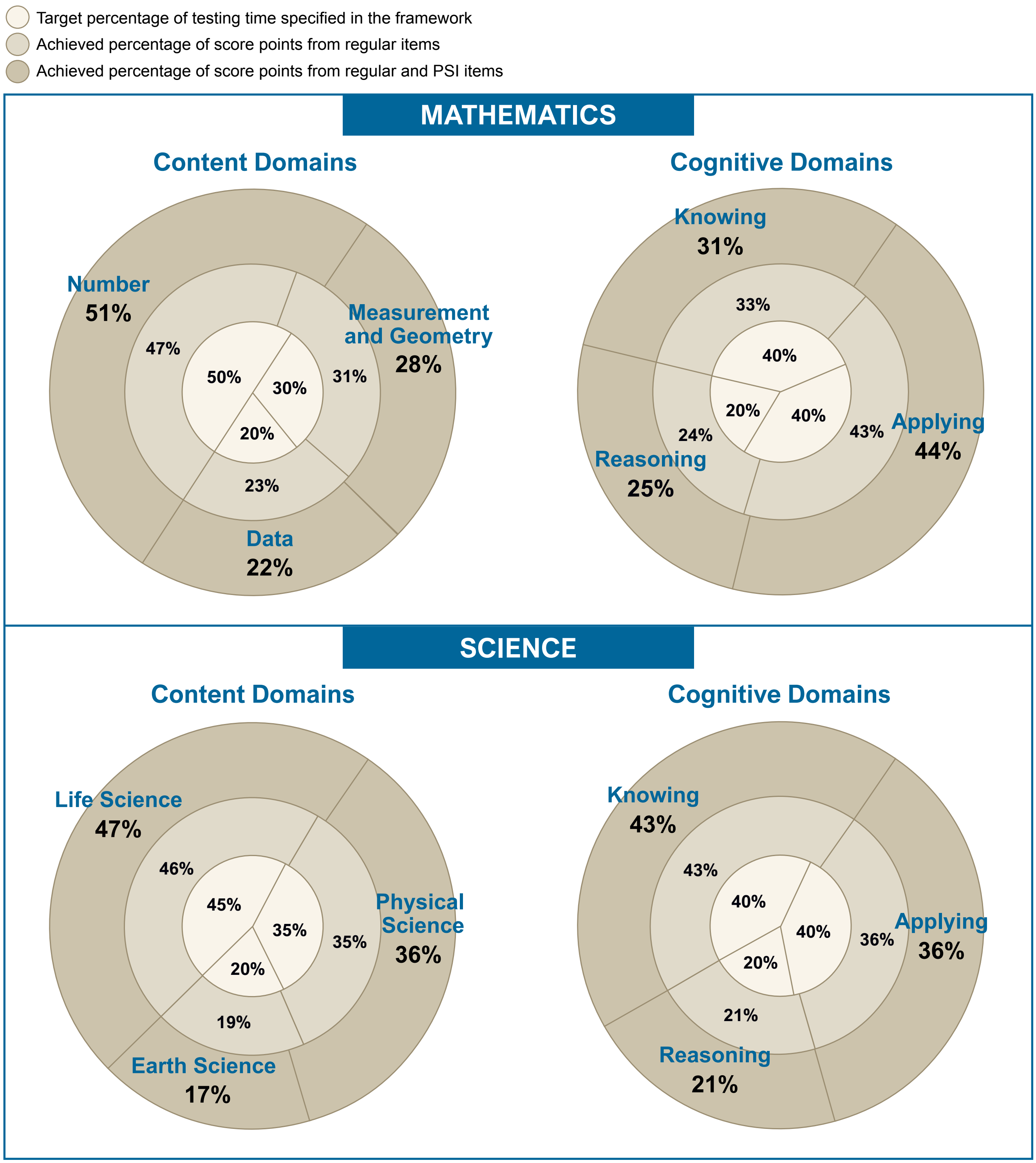
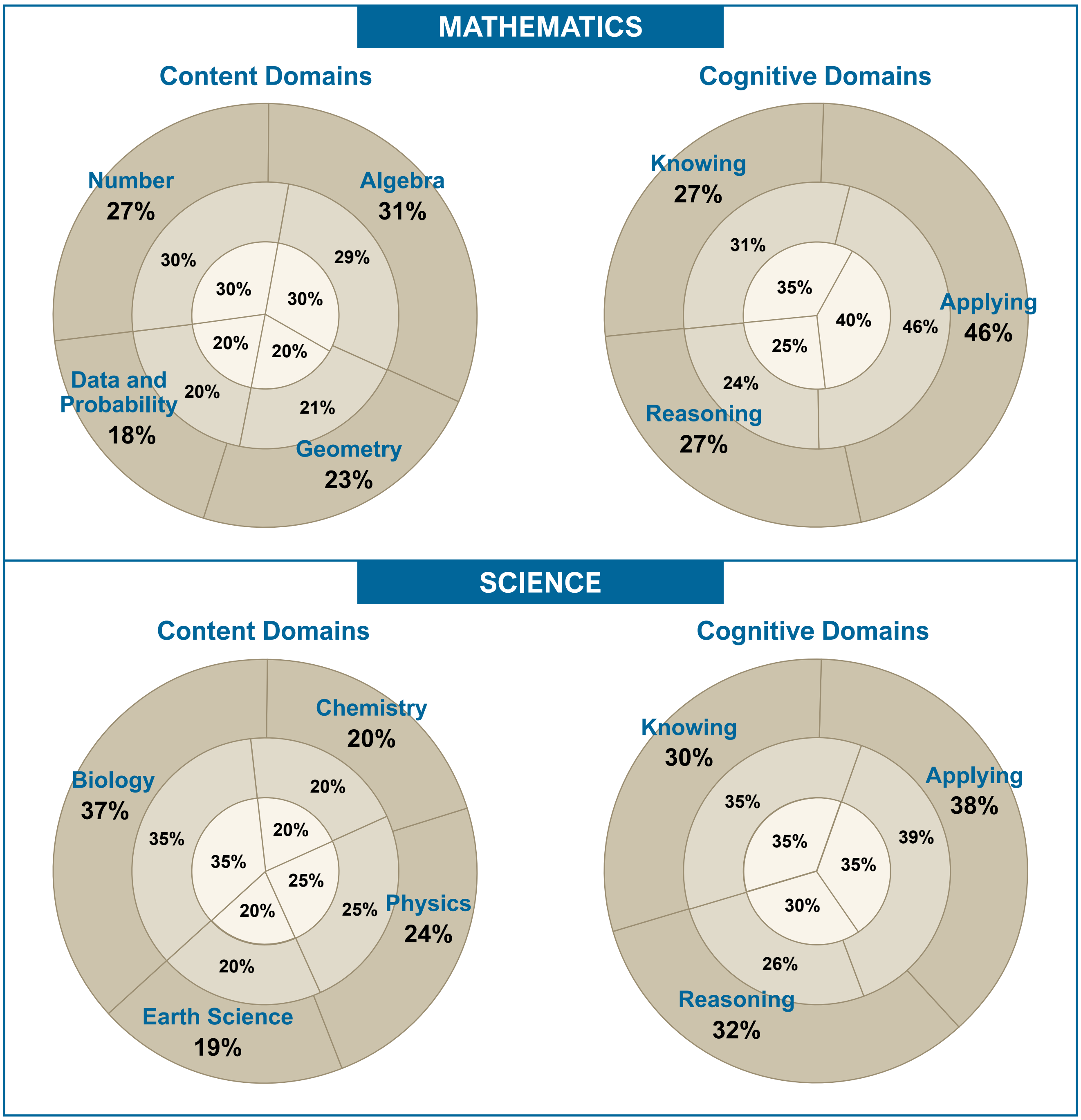


Exhibit 1.20: **Comparison of Target and Achieved Percentages of Domain Coverage in the eTIMSS 2019 Mathematics and Science Assessments – Eighth Grade**

- Target percentage of testing time specified in the framework
- Achieved percentage of score points from regular items
- Achieved percentage of score points from regular and PSI items



TIMSS 2019 Constructed Response Scoring Training

In preparation for the main data collection scoring training, some TIMSS 2019 scoring guides were further refined or clarified based on the results of the field test. This also included a thorough review of the field test scoring training materials to ensure that the student responses were still suitable for the updated scoring guides. In some cases, example and practice sets used in the field test were expanded to further illustrate particular aspects of a scoring guide.

The TIMSS 2019 scoring training materials also included the training sets for the trend items used in TIMSS 2015. These training materials were updated for TIMSS 2019 to include both typed and handwritten responses. In all, the TIMSS 2019 scoring training materials included sets of example and practice responses for a total of 26 fourth grade items and 27 eighth grade items.

To provide scoring training for all the countries participating in TIMSS 2019, the TIMSS & PIRLS International Study Center conducted two training sessions. First, the NRCs for Southern Hemisphere countries and their scoring supervisors received scoring training in November 2018 in Cape Town, South Africa. The NRCs for Northern Hemisphere countries and their scoring supervisors received scoring training in March 2019 in Limassol, Cyprus as part of the 6th TIMSS 2019 NRC meeting. Exhibit 1.21 shows the number of participants in the two scoring training sessions.

Exhibit 1.21: TIMSS 2019 Scoring Training Participation

Participants	Southern Hemisphere	Northern Hemisphere
Number of Countries	7	52
Number of Benchmarking Entities	–	5
Number of Country Representatives	24	150

After participating in scoring training, the NRCs and their scoring supervisors organized and carried out scoring activities in their respective countries. In addition to scoring the student responses, all countries participated in several supplementary scoring activities to document the scoring reliability of the human-scored items. The procedures used to establish scoring reliability within each country, across countries, and across assessment cycles are described in Survey Operations Procedures for TIMSS 2019.

The Process Following Instrument Development

After the participating countries received the international version of the achievement instruments, they began the process of translation and cultural adaptation (some adaptation to local usage typically is necessary even in English-speaking countries) and production of the materials needed to administer the assessment. The tasks involved in producing the materials differed depending on whether eTIMSS or paperTIMSS was being administered. At the same time, countries made final arrangements for data collection, including the host of activities necessary to obtain school participation and implement test administration.

References

Fishbein, B., Martin, M. O., Mullis, I. V. S., & Foy, P. (2018). The TIMSS 2019 item equivalence study: Examining mode effects for computer-based assessment and implications for measuring trends. *Large-scale Assessments in Education, 6*(1), 11.

Lindquist, M., Philpot, R., Mullis, I. V. S., & Cotter, K. E. (2017). TIMSS 2019 mathematics framework. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2019 assessment frameworks* (pp. 11-25). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/

Martin, M. O., Mullis, I. V. S., & Foy, P. (2017). TIMSS 2019 assessment design. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2019 assessment frameworks* (pp. 79-91). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design* (Research Report No. RR-03-16). Princeton, NJ: Educational Testing Service.

Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 assessment frameworks*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/

CHAPTER 2

Updating the TIMSS 2019 Instruments for Describing the Contexts for Student Learning

Ina V.S. Mullis
Bethany Fishbein

Introduction

Beginning with the first TIMSS assessments in 1995, each TIMSS assessment cycle has collected an array of information from each participating country about the national, home, school, and classroom contexts in which students learn mathematics and science. The purpose for collecting these data is to learn more about the educational factors that are related to mathematics and science achievement by examining these factors internationally across and within countries. The fundamental idea is for countries to learn from each other about possible ways to improve their own education systems.

Considering countries' mathematics and science achievement together with the factors that can facilitate that achievement is at the core of TIMSS. Collecting comparable data across countries about students' opportunities for learning mathematics and science is as central to TIMSS as collecting comparable data about students' mathematics and science achievement.

The areas of the student learning contexts addressed in the TIMSS 2019 context questionnaire instruments were described in the TIMSS 2019 Context Questionnaire Framework. Because TIMSS has been administered every four years since 1995, making TIMSS 2019 the seventh TIMSS administration, many aspects of collecting the contextual data have become relatively stable across cycles. Similar to previous TIMSS assessments, in TIMSS 2019:

- The home, school, teacher, and student context questionnaires were administered together with the mathematics and science assessments
- Substantial portions of the *TIMSS 2019 International Results in Mathematics and Science* were devoted to reporting the data collected via the home, school, teacher, and student

questionnaires in relation to countries' achievement on the mathematics and science assessments

- The *TIMSS 2019 Encyclopedia* was based on countries' responses to the TIMSS 2019 Curriculum Questionnaire and each country wrote a chapter for the *Encyclopedia* describing its mathematics and science curricula and general education policies
- Many of the topics covered in the context questionnaires and the information provided by countries for their chapters in the *TIMSS 2019 Encyclopedia* were similar to those in previous assessments, although updated for TIMSS 2019.

Description of the TIMSS 2019 Context Questionnaires

This section describes the TIMSS 2019 Home, School, Teacher, Student, and Curriculum Questionnaires, including who was responsible for completing each questionnaire, the content covered, and the method for administering the questionnaire. The TIMSS 2019 context questionnaires can be viewed in their entirety on the *TIMSS 2019 Context Questionnaires webpage.*

Home Questionnaire

The Home Questionnaire (also known as the "Early Learning Survey") was administered at the fourth grade to the students' parents or guardians. It asked about home resources for fostering literacy and numeracy skills, the parents' highest level of education, employment situations, opinions about their child's school, their child's attendance in preprimary education programs, the emphasis on literacy and numeracy activities in the home before the child attended school (such as reading books, singing songs, writing words and numbers, and counting), and the level of their child's literacy and numeracy skills when beginning school. Countries asked students' parents or guardians to complete the questionnaire online or sent it to the students' homes in paper-and-pencil format.

School Questionnaire

The School Questionnaire was administered at the fourth and eighth grades to the principals of the students' schools. It asked about the level of students' literacy and numeracy skills when they first enter the school, the availability of instructional resources, the socioeconomic background of the students attending the school, the school's emphasis on academic success, the need for discipline, and the principals' education. Countries administered the questionnaire either online or via paper-and-pencil.

Teacher Questionnaires

A single version of the Teacher Questionnaire was administered at the fourth grade to students' teachers, given that generally the same teachers taught the students both mathematics and science. At the eighth grade, there were separate versions of the questionnaire for the students' mathematics teachers and the students' science teachers. The questionnaires asked about the teachers' education, professional development, and career satisfaction as well as about students' readiness for instruction, the frequency they do various instructional activities, difficulties in providing instruction, curriculum topics covered, assessment practices, and availability of computers for instruction. Countries administered the Teacher Questionnaire either online or via paper-and-pencil.

Student Questionnaire

Administered to all students at the fourth and eighth grades, the Student Questionnaire asked students about their educational experiences at home and school related to learning mathematics and science. It also included several scales about their attitudes toward learning mathematics and science. At the eighth grade, there were two versions of the questions about science—one for countries that taught science as an integrated subject and one for countries where science was taught as separate subjects (e.g., biology, earth science, chemistry, and physics). The separate science questionnaire asked some of the questions for each content area individually. Regardless of whether they were participating in eTIMSS or paperTIMSS, students were administered a paper-and-pencil questionnaire at the end of their testing session.

For countries that participated in eTIMSS, students also answered several questions on their digital devices at the end of the assessment about their experience taking the eTIMSS assessment and their familiarity with digital devices.

Curriculum Questionnaire

The Curriculum Questionnaire was administered at the fourth and eighth grades to the National Research Coordinators (NRCs) of the participating countries. This questionnaire collected information about national curriculum policies and practices related to the countries' educational systems and the organization and content of the mathematics and science curricula in their country. The Curriculum Questionnaire was administered online.

Maintaining Continuity with Previous Assessments

Much of the information in the TIMSS 2019 context questionnaires was collected in the form of context questionnaire scales (typically including 8–12 items) that measure particular factors or constructs that have been found to be related to mathematics and science achievement as assessed by TIMSS. Many of the scales included in the TIMSS 2019 questionnaires were brought forward from 2015 because they addressed home and school factors that have been of interest for several assessment cycles. These scales either were brought forward in their entirety or modified for the 2019 cycle.

The following existing scales were included in the Home Questionnaire:

- Home Resources for Learning
- Home Early Literacy and Numeracy Activities Before Primary School
- Could Do Early Literacy and Numeracy Tasks When Beginning Primary School
- Parents' Perceptions of Their Child's School

Existing scales included in the School Questionnaire covered:

- Instruction Affected by Resource Shortages
- School Emphasis on Academic Success (also included in the Teacher Questionnaire)
- School Discipline
- Schools Where Students Entered Primary Grades with Literacy and Numeracy Skills

Existing scales included in the Teacher Questionnaire covered:

- School Emphasis on Academic Success (also in the School Questionnaire)
- Safe and Orderly School
- Teachers' Job Satisfaction
- Classroom Teaching Limited by Students Not Ready for Instruction

The following existing scales were included in the Student Questionnaire:

- Student Bullying
- Sense of School Belonging

TIMSS 2019 also continued the long-standing practice of asking students about their attitudes toward mathematics and science, primarily via the following scales:

- Students Like Learning Mathematics
- Students Like Learning Science
- Students Confident in Mathematics
- Students Confident in Science
- Students Value Mathematics
- Students Value Science

Other topics also were brought forward to TIMSS 2019 from previous assessments. Collecting information about the curriculum has been central to TIMSS from the beginning, and TIMSS 2019 continued this by asking countries to describe their mathematics and science curricula in the curriculum questionnaire and in their chapters for the *TIMSS 2019 Encyclopedia*. The chapters detailed each country's nationally specified (or formal) curricula in mathematics and science (sometimes called the intended curriculum by TIMSS). To collect information about students' opportunity to learn the country's curriculum, the teachers of the TIMSS students were asked which TIMSS topics had been covered during the current or previous school years.

Teacher education policies and practices also have been of continued interest across assessment cycles. TIMSS 2019 asked countries to describe the education and credentialing procedures for becoming a teacher in the Curriculum Questionnaire and professional development requirements and programs were described in the *Encyclopedia* chapters. The Teacher Questionnaire asked teachers themselves about their education degrees, areas of concentrated study while earning their degrees, and participation in professional development after becoming a teacher.

Updates to the TIMSS 2019 Context Questionnaires and Encyclopedia Chapters

Although a number of scales and questions were brought forward from TIMSS 2015, the TIMSS 2019 Home, School, Teacher, and Student Questionnaires as well as the Curriculum Questionnaire and the outline for the countries' *Encyclopedia* chapters were updated to address important areas of current research, such using digital devices in mathematics and science instruction. Considering recommendations from the participating countries about the most useful information to collect, TIMSS 2019 had three main goals for improving the context questionnaires: 1) enhancing the measures of teacher instructional quality, 2) addressing areas relevant to using technology in instruction and assessment, and 3) reducing the response burden for teachers.

TIMSS 2019 focused on scales of teacher instructional quality based on students' reports. New items were written for the Student Questionnaire to enhance existing measures aligned with research about "instructional clarity." To address the topic of classroom management, a new scale was developed asking students about the extent disorderly behavior occurs in their mathematics lessons. New items in the Student Questionnaire also asked about how often teachers use instructional activities related to problem solving and inquiry, such as conducting science experiments.

With more than half the participating countries transitioning to eTIMSS, TIMSS 2019 renewed efforts to collect data about technology use for instruction and assessment. The School Questionnaire asked principals about the availability of technology resources, such as digital learning resources (e.g., digital books). Teachers were asked about using computers to support students in mathematics and science lessons, and whether students take mathematics and science tests on digital devices. To cover additional questionnaire topics relevant to digital assessment, students who took eTIMSS answered questions about their familiarity with using digital devices for schoolwork.

New items for the Home Questionnaire improved coverage for early numeracy activities and skills, such as drawing shapes and measuring quantities. The Student Questionnaire scale at the eighth grade about students' bullying experiences also was revamped to better reflect the current trends related to social media and cyberbullying. The new scale included a greater emphasis on bullying experienced through digital devices.

Based on feedback from NRCs and in response to high rates of teacher nonresponse in several TIMSS 2015 countries, several items and scales that were given a lower priority were retired from the Teacher Questionnaire, including those asking about school working conditions, collaborating with other teachers, and confidence in teaching the curriculum.

TIMSS 2019 also retired some content to reduce the burden for NRCs. Several topics were moved from the *Encyclopedia* chapters to the Curriculum Questionnaire, including the countries' language(s) of instruction, additional education requirements for mathematics and science teachers, and the first grade of schooling taught by subject specialist teachers. Several topics deemed to be outdated were deleted from the Curriculum Questionnaire, such as policies for student tracking and the process for approving instructional materials.

Overview of the Updating Process

With each new assessment cycle, the TIMSS & PIRLS International Study Center at Boston College follows a collaborative and iterative process to update the TIMSS data collection instruments for the contexts for learning mathematics and science. For TIMSS 2019, Executive Directors Ina Mullis and Michael Martin and TIMSS Questionnaire Coordinator Martin Hooper (through 2018) led the development process, which involved updating the questionnaires from 2015, conducting several iterations of review, and a full-scale field test. Based on the field test results, minor revisions were made to the questionnaires and final reviews were conducted prior to data collection.

The National Research Coordinators (NRCs) who were designated by the participating countries to be responsible for implementing TIMSS 2019 played a key role in reviewing the TIMSS 2019 context questionnaires. They provided feedback and proposed new topics at NRC meetings throughout the development process, including at the first TIMSS 2019 NRC meeting, as well as the NRC meetings before the field test and prior to TIMSS 2019 Data Collection.

The TIMSS 2019 Questionnaire Item Review Committee (QIRC) consisted of NRCs with experience and expertise in education policy analysis and survey development. Members of QIRC made major contributions in updating the TIMSS 2019 Context Questionnaire Framework and in modifying and developing the context questionnaires. This included conducting an online review and attending two committee meetings—a first meeting prior to the field test and a second meeting prior to data collection. The members of the TIMSS 2019 QIRC are listed in Exhibit 2.1.

Exhibit 2.1: TIMSS 2019 Questionnaire Item Review Committee (QIRC)

Sue Thomson
Australian Council for Educational Research (ACER)
Australia

Josef Basl
Czech School Inspectorate
Czech Republic

Heike Wendt
Institute for School Development Research (IFS)
TU Dortmund University
Germany

Laura Palmerio
Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione (INVALSI)
Italy

Kyongah Sang
Center for Global Education
Korea Institute for Curriculum & Evaluation
Korea, Republic of

Martina Meelissen
Department of Research Methodology, Measurement, and Data Analysis
University of Twente
The Netherlands

Trude Nilsen
Department of Teacher Education and School Research
ILS, University of Oslo
Norway

Vijay Reddy
Human Sciences Research Council (HSRC)
South Africa

Sean P. "Jack" Buckley
American Institutes for Research
United States

Reviewing the Field Test Results for the TIMSS 2019 Context Questionnaires

The field test is an important step for assessing the quality of the home, school, teacher, and student questionnaire instruments and measurement scales before data collection. Particularly for newly developed items, this step in the updating process also gives countries' an opportunity to ensure the items are appropriately translated and adapted to their national contexts so that their data are internationally comparable (see Chapter 5: Instrument Translation and Layout Verification for TIMSS 2019).

Subsequent to conducting the TIMSS 2019 Field Test, the TIMSS & PIRLS International Study Center analyzed the field test data, consisting of responses from: 1) 66,626 parents or caregivers to the Home Questionnaire, 2) 2,682 principals to the School Questionnaire, 3) 10,993 teachers to the Teacher

Questionnaire, and 4) 121,454 students to the Student Questionnaire. The staff at the TIMSS & PIRLS International Study Center produced data almanacs containing item statistics for each questionnaire item, including the percentage of students responding to each response option, with the corresponding average student achievement in mathematics or science, respectively. The staff also prepared context questionnaire scale summaries to evaluate the suitability of the items for scaling with one parameter item response theory (Rasch) model. The scales were evaluated for unidimensionality, reliability, and their relationship with achievement. More information about the TIMSS 2019 context questionnaire scales and their measurement properties can be found in Chapter 16: Creating and Interpreting the TIMSS 2019 Context Questionnaire Scales.

The TIMSS & PIRLS International Study Center reviewed the field test results and updated the questionnaires as necessary for the final round of reviews by the TIMSS 2019 QIRC and NRCs. The next section contains the complete schedule of activities included in the updating process.

Schedule of Activities for Updating the TIMSS 2019 Instruments for Describing Contexts for Student Learning

Exhibit 2.2 presents the schedule for updating the TIMSS 2019 instruments used to collect information about students' home, school, and classroom contexts for learning mathematics and science. The iterative review process formally began in February 2017 at the 1st TIMSS 2019 NRC meeting and ended with finalizing the Curriculum Questionnaire in April 2019.

Exhibit 2.2: TIMSS 2019 Schedule of Activities for Updating Context Questionnaires

Date(s)		Group and Activity
February	2017	NRCs reviewed the TIMSS 2015 context questionnaires, providing ideas for new topics that should be addressed in TIMSS 2019 (1st NRC meeting—Hamburg, Germany)
February–June	2017	TIMSS & PIRLS International Study Center drafted the TIMSS 2019 Context Questionnaire Framework incorporating NRC feedback
June–July	2017	The TIMSS 2019 Questionnaire Item Review Committee (QIRC) conducted an online review of the draft TIMSS 2019 Context Questionnaire Framework
July–August	2017	TIMSS & PIRLS International Study Center finalized the TIMSS 2019 Context Questionnaire Framework incorporating QIRC feedback and drafted the updated TIMSS 2019 Field Test Home, School, Teacher, and Student Questionnaires
August	2017	TIMSS & PIRLS International Study Center published *TIMSS 2019 Assessment Frameworks*, including the TIMSS 2019 Context Questionnaire Framework
September	2017	QIRC reviewed the draft TIMSS 2019 Field Test Home, School, Teacher, and Student Questionnaires (1st QIRC meeting—Hengelo, the Netherlands)
September–November	2017	TIMSS & PIRLS International Study Center incorporated the QIRC suggestions into the draft TIMSS 2019 Field Test Home, School, Teacher, and Student Questionnaires

Exhibit 2.2: TIMSS 2019 Schedule of Activities for Updating Context Questionnaires (continued)

Date(s)		Group and Activity
November	2017	NRCs reviewed the draft field test home, school, teacher, and student questionnaires (3rd NRC meeting—Melbourne, Australia)
November–December	2017	TIMSS & PIRLS International Study Center finalized the field test home, school, teacher, and student questionnaires, incorporating suggestions from the NRCs
December	2017	TIMSS & PIRLS International Study Center provided the TIMSS 2019 field test questionnaires to the NRCs for translation
March–May	2018	Countries conducted TIMSS 2019 Field Test
April–May	2018	Countries submitted field test data to IEA Hamburg for review
May–June	2018	TIMSS & PIRLS International Study Center analyzed the field test data and reviewed the results
July	2018	QIRC reviewed questionnaires together with the field test results and proposed revisions to the home, school, teacher, and student questionnaires. QIRC also reviewed the draft TIMSS 2019 Curriculum Questionnaire and *Encyclopedia* chapter outline (2nd QIRC meeting—Oslo, Norway)
July–August	2018	TIMSS & PIRLS International Study Center incorporated the QIRC suggestions into the questionnaires
August	2018	NRCs reviewed the proposed TIMSS 2019 Home, School, Teacher, and Student Questionnaires (5th NRC meeting—Stockholm, Sweden)
August	2018	TIMSS & PIRLS International Study Center distributed the TIMSS 2019 Home, School, Teacher, and Student Questionnaires to NRCs for translation and verification
October–December	2018	Southern Hemisphere countries conducted TIMSS 2019 Data Collection
January–March	2019	TIMSS & PIRLS International Study Center incorporated the QIRC suggestions into the TIMSS 2019 Curriculum Questionnaire and Encyclopedia chapter outline
March	2019	NRCs reviewed the proposed TIMSS 2019 Curriculum Questionnaire and *Encyclopedia* chapter outline (6th NRC meeting—Limassol, Cyprus)
March–June	2019	Northern Hemisphere countries conducted TIMSS 2019 Data Collection
April–October	2019	NRCs responded to the online TIMSS 2019 Curriculum Questionnaire
October–February	2019	NRCs submitted their *TIMSS 2019 Encyclopedia* chapters to the TIMSS & PIRLS International Study Center

CHAPTER 3

Sample Design in TIMSS 2019

Sylvie LaRoche
Marc Joncas
Pierre Foy

Introduction

TIMSS is designed to provide valid and reliable measurement of trends in student achievement in countries around the world, while keeping to a minimum the burden on schools, teachers, and students. The TIMSS program employs rigorous school and classroom sampling techniques so that achievement in the student population as a whole may be estimated accurately by assessing just a sample of students from a sample of schools. TIMSS assesses mathematics and science achievement at two grade levels and so TIMSS has two target populations—all students enrolled at the fourth grade and all students enrolled at the eighth grade, counting from the first year of primary schooling. Countries may assess either one or both student populations. In addition, at the fourth grade for the TIMSS 2019 cycle, countries for which the regular fourth grade mathematics assessment is too difficult have the option to administer a less difficult mathematics assessment, consisting of one third of the items from the regular assessment and two-thirds less difficult items. Countries availing of the less difficult mathematics option administer the regular fourth grade science assessment.

TIMSS 2019 marks the beginning of the TIMSS transition to computer based assessment, with countries having the option of administering the new computer-based version of the 2019 assessment, known as eTIMSS, or the paper-and-pencil version as in previous assessment cycles (paperTIMSS). Although the two versions were developed to be as similar in content as possible, inevitably there are some differences between them as a result of the two modes of administration. In order to control for mode effects while linking the two versions to the TIMSS achievement scales and to safeguard the measurement of trends from previous assessments, eTIMSS countries also provide a separate sample of "bridge" data. The bridge data result from administering the paper version of the trend items (eight blocks of items for each subject and grade that also were administered in 2015) to a separate, equivalent sample of students during the main data collection. These paper versions of the trend items are identical in most respects to the eTIMSS versions that are administered as part of the main eTIMSS assessment, and so comparing performance on the eTIMSS versions to performance on the paper versions administered to the bridge sample provides a bridge between the two assessment modes.

The TIMSS assessments employ a two-stage random sample design, with a sample of schools drawn as a first stage and one or more intact classes of students selected from each of the sampled schools as a second stage. Intact classes of students are sampled rather than individuals from across the grade level or of a certain age because TIMSS pays particular attention to students' curricular and instructional experiences, and these typically are organized on a classroom basis. Sampling intact classes also has the operational advantage of less disruption to the school's day-to-day business than individual student sampling.

National Sampling Plan

Each country participating in TIMSS needs a plan for defining its national target population and applying the TIMSS sampling methods to achieve a nationally representative sample of schools and students. The development and implementation of the national sampling plan is a collaborative exercise involving the country's National Research Coordinator (NRC) and TIMSS sampling experts.

Statistics Canada is responsible for advising the National Research Coordinator on all sampling matters and for ensuring that the national sampling plan conforms to the TIMSS standards. In cooperation with sampling staff from IEA Hamburg, Statistics Canada works with the NRC to select the national school sample(s) and produce all supporting documentation for tracking the sampled schools. This includes ensuring that the school sampling frame (the school population list from which the school sample is drawn) provided by the NRC is complete and satisfactory; checking that categories of excluded students are clearly defined, justified, and kept to a minimum; assisting the NRC in determining the sample size and a stratification plan that will meet both international and national objectives; and drawing a national sample of schools. When sampling has been completed and all data collected, Statistics Canada documents population coverage and school and student participation rates and constructs appropriate sampling weights for use in analyzing and reporting the results.

The TIMSS & PIRLS International Study Center, in cooperation with Statistics Canada and IEA Hamburg, provides National Research Coordinators with a series of manuals to guide them through the sampling process. More specifically, *TIMSS 2019 Survey Operations Procedures Unit 1: Sampling Schools and Obtaining their Cooperation* describes the steps involved in defining the national target population and selecting the school sample, and *TIMSS 2019 Survey Operations Procedures Unit 3: Contacting Schools and Sampling Classes for the TIMSS 2019 Data Collection* describes the procedure for sampling classes within the sampled schools and making preparations for conducting the assessments. Within-school sampling procedures for the field test are documented in *TIMSS 2019 Survey Operations Procedures Unit 2: Preparing for and Conducting the TIMSS 2019 Field Test*. More information on the Survey Operations Units can be found in Chapter 6 of this volume.

The TIMSS National Research Coordinator is responsible for providing Statistics Canada with all information and documentation necessary to conduct the national sampling, and for conducting all sampling operations in the country. In particular, the NRC is expected to identify the grade(s) that correspond to the international target population(s); create a sampling frame by listing all schools in the population that have classes with students in the target grade(s); determine national population coverage and exclusions, in accordance with the TIMSS international guidelines; work with Statistics Canada to develop a national sampling plan and identify suitable stratification variables, ensuring that these variables are present and correct for all schools; contact all sampled schools and secure their participation; keep track of school participation and the use of replacement schools; and conduct all within-school sampling of classes. As described in this chapter, each NRC is required to complete a series of sampling forms documenting the completion of each of these tasks.

A crucial feature of each international meeting of National Research Coordinators is a one-to-one meeting between each NRC and sampling staff at Statistics Canada and IEA Hamburg. At these meetings, each step of the sampling process is documented and reviewed in detail, and NRCs have the opportunity to raise issues and ask questions about their national situation and any challenges they face. Statistics Canada consults with the TIMSS & PIRLS International Study Center and the International Sampling Referee, as necessary, to resolve issues and questions. Final approval of TIMSS national sampling plans is the responsibility of the TIMSS & PIRLS International Study Center, based upon the advice of Statistics Canada and the International Sampling Referee.

Defining the Target Population

As an international study of the comparative effects of education on student achievement in mathematics and science, TIMSS defines its international target populations in terms of the amount of schooling students have received. The number of years of formal schooling is the basis of comparison among participating countries. Thus, the TIMSS international target population at the fourth grade is all students in their fourth year of formal schooling, and at the eighth grade, all students in their eighth year. UNESCO's International Standard Classification of Education (ISCED) 2011 (UNESCO, 2012) provides an internationally accepted classification scheme for describing levels of schooling across countries. The ISCED system describes the full range of schooling, from preprimary (Level 0) to the doctoral level (Level 8). ISCED Level 1 corresponds to primary education or the first stage of basic education. The first year of Level 1 "coincides with the transition point in an education system where systematic teaching and learning in reading, writing and mathematics begins" (UNESCO, 2012, p. 30). Four years after this would be the target grade for fourth grade TIMSS and is the fourth grade in most countries. Similarly, eight years after the first year of ISCED Level 1 is the target grade for eighth grade TIMSS and is the eighth grade in most countries. However, given the cognitive demands of the assessments, TIMSS wants to avoid assessing very young students. Thus, TIMSS recommends assessing the next higher grade (i.e., fifth grade for fourth grade TIMSS and ninth grade for eighth grade TIMSS) if, for fourth grade students, the average age at the time of testing would be less than 9.5 years and, for eighth grade students, less than 13.5 years.

The fourth grade and eighth grade target populations of students are defined as follows:

- **Fourth grade:** All students enrolled in the grade that represents four years of schooling counting from the first year of ISCED Level 1, providing the mean age at the time of testing is at least 9.5 years
- **Eighth grade:** All students enrolled in the grade that represents eight years of schooling counting from the first year of ISCED Level 1, providing the mean age at the time of testing is at least 13.5 years

All students enrolled in the target grade, regardless of their age, belong to the international target population and should be eligible to participate in TIMSS. Because students are sampled in two stages, first by randomly selecting a school and then randomly selecting a class from within the school, it is necessary to identify all schools in which eligible students are enrolled. Essentially, eligible schools for TIMSS are those that have any students enrolled in the target grade, regardless of type of school. All schools of all educational sub-systems that have students learning full time in the target grade are part of the international target population, including schools that are not under the authority of the national Ministry of Education.

National Target Populations

For most countries, the target grade for TIMSS is the fourth and/or eighth grade. However, because educational systems vary in structure and in policies and practices with regard to age of starting school and promotion and retention, there are differences across countries in how the target grades are labelled and in the average age of students. To ensure that the appropriate national target grades are selected, each NRC completes Sampling Form 1, which identifies the target grades, the country's name for those grades, and the average age of students in those grades at the time of data collection. An example of a completed Sampling Form 1 is presented in Exhibit 3.1.

For a variety of reasons, there are countries where students in the fifth or sixth grade are more likely to have developed the mathematics and science competencies necessary for success on the TIMSS fourth grade assessment, or in the ninth grade for the TIMSS eighth grade assessment. Such countries may choose to participate in TIMSS at either the fifth or sixth grade or in the less difficult mathematics fourth grade assessment. Similarly, some countries may choose to administer the TIMSS eighth grade assessment to their ninth grade students.

Exhibit 3.1: Example of Sampling Form 1

Sampling Form 1 — **General Information**

See Section 2 of TIMSS 2019 Survey Operations Procedures Unit 1

TIMSS 2019 Participant:	*Country X*
National Research Coordinator:	*Name of NRC*

1. Please indicate the assessment(s) in which your country plans to participate along with the target grade(s), name(s), and expected average age of students at the time of testing:

Grade 4 TIMSS Assessment — *Yes*

Target Grade	Name of the Target Grade	Average Age	Mode (Paper-TIMSS or eTIMSS)	Less Difficult Mathematics Item Blocks Option (Yes/No)
4	*Grade 4*	*9.7*	*eTIMSS*	*No*
			Select	*Select*
			Select	*Select*

Grade 8 TIMSS Assessment — *Yes*

Target Grade	Name of the Target Grade	Average Age	Mode (Paper-TIMSS or eTIMSS)
8	*Grade 8*	*13.7*	*eTIMSS*
			Select
			Select

2. Specify the usual start and end date(s) of the school year and the expected date(s) of testing for the field test and data collection.

	Start of school year: *(DD-MM-YYYY)*	End of school year: *(DD-MM-YYYY)*	Expected Testing Period
Field Test	*05/09/2017*	*22/06/2018*	*16 - 27 April 2018*
Data Collection	*01/09/2018*	*21/06/2019*	*13 - 24 April 2019*

4. Specify the language(s) in which the assessment(s) will be administered.

English

5. Describe the grade structure through ISCED Level 1 (primary education or the first stage of basic education) and ISCED Level 2 (basic or lower secondary education) in your country.

Grades 1 to 6 , Primary schools
Grades 7 to 9 , Lower secondary schools

6. Describe the age and birth date rules for entering ISCED Level 1 in your country.

Children must enter school (grade 1) in the autumn of the year in which they have their sixth birthday

National Coverage and Exclusions

TIMSS is designed to describe and summarize student achievement across the entire target grade (fourth or eighth), and so it is very important that national target populations aim for comprehensive coverage of eligible students. However, in some cases, political, organizational, or operational factors make complete national coverage difficult to attain. Thus, in some rare situations, certain groups of schools and students may have to be excluded from the national target population. For example, it may be that a particular geographical region, educational sub-system, or language group cannot be covered. Such exclusion of schools and students from the target population is referred to as reduced population coverage.

Even countries with complete population coverage find it necessary to exclude at least some students from the target population because they attend very small schools, have intellectual or functional disabilities, or are non-native language speakers. Such students may be excluded at the school level (i.e., the whole school is excluded) or within the school on an individual basis.

School-Level Exclusions. Although it is expected that very few schools will be excluded from the national target population, NRCs are permitted to exclude schools on the following grounds when they consider it necessary:

- Inaccessibility due to their geographically remote location
- Extremely small size (e.g., four or fewer students in the target grade)
- Offering a grade structure, or curriculum, radically different from the mainstream educational system
- Providing instruction solely to students in the student-level exclusion categories listed below (e.g., catering only to special needs students)

Student-Level Exclusions. The international within-school exclusion rules are specified as follows:

- Students with functional disabilities — These are students who have physical disabilities such that they cannot perform in the TIMSS testing situation. Students with functional disabilities who are able to perform should be included in the testing.
- Students with intellectual disabilities — These are students who are considered, in the professional opinion of the school principal or by other qualified staff members, to have intellectual disabilities or who have been tested as such. This includes students who are emotionally or mentally unable to follow even the general instructions of the test. Students should not be excluded solely because of poor academic performance or normal disciplinary problems. It should be noted that students with dyslexia, or other such learning disabilities, should be accommodated in the test situation if possible, rather than excluded.

- Non-native language speakers — These are students who are unable to read or speak the language(s) of the test and would be unable to overcome the language barrier in the test situation. Typically, a student who has received less than one year of instruction in the language(s) of the test should be excluded.

Because disability criteria vary from country to country, NRCs are asked to translate the TIMSS international exclusion standards into the local equivalent. Students should be considered for exclusion strictly in accordance with the international standards. If a sampled school contains a class consisting entirely of students from one of the exclusion categories, such a class is excluded prior to classroom sampling.

NRCs understand that exclusion rates must be kept to a minimum so that national samples accurately represent the national target population. Requirements for exclusion rates include the following:

- The overall number of excluded students must not account for more than 5 percent of the national target population of students in a country. The overall number includes both school-level and within-school exclusions.
- The number of students excluded because they attend very small schools must not account for more than 2 percent of the national target population of students.

To document population coverage and exclusions, each NRC completes Sampling Form 2, which lists the number of students in the national target population and the number of students excluded at both the school level and within the school for each population to be assessed. An example of a completed Sampling Form 2 is presented in Exhibit 3.2.

Exhibit 3.2: Example of Sampling Form 2

Sampling Form 2 — **Coverage and Exclusions**

See Section 3 of TIMSS 2019 Survey Operations Procedures Unit 1

TIMSS 2019 Participant : *Country X*

1. This Sampling Form refers to: *TIMSS Grade 4 Assessment*

		Number of schools	Number of students
Total enrollment in the target grade:	[a]	*822*	*56,560*

2. School-level exclusions *(if applicable)*:

	Description of exclusions	Number of schools	Number of students
1.	*Students taught in language other than English*	*8*	*630*
2.	*Special education schools*	*16*	*325*
3.	*Very small schools (less than 5 students in grade 4)*	*40*	*110*
4.			
5.			
TOTAL:	*(Sum of exclusions - Calculated automatically)* [b]	*64*	*1,065*

		schools	students
Percentage of school-level exclusions: *(Box [b] ÷ Box [a] x 100)*	**[1]**	***7.8%***	***1.9%***

3. Total enrollment after school-level exclusions: [c] *758* *55,495*
(Box [c] = Box [a] - Box [b])

Totals and percentages calculated automatically

4. Within-school exclusions *(if applicable)*:

	Description of exclusions	Number of students
1.	*Students with special education needs (based on TIMSS 2015)*	*640*
2.		
3.		
TOTAL:	*(Sum of exclusions - Calculated automatically)* [d]	*640*

		schools	students
Expected percentage of within-school exclusions: *(Box [d] ÷ Box [c] x 100)*	**[2]**	***0.0%***	***1.2%***

5. **Expected percentage of reduced coverage and exclusions:** ***7.8%*** ***3.0%***
(Box [1] + (1 - Box [1]) X Box [2])

Totals and percentages calculated automatically

6. Total enrollment in the target grade in previous school years.

Years	Number of schools	Number of students
2016/2017	*856*	*58,451*
2015/2016	*890*	*61,489*

Requirements for Sampling the Target Population

TIMSS sets high standards for sampling precision, participation rates, and sample implementation in order to achieve national samples of the highest quality and survey estimates that are unbiased, accurate and internationally comparable.

Sampling Precision and Sample Size

Because TIMSS is fundamentally a study of student achievement, the precision of estimates of student achievement is of primary importance. To meet the TIMSS standards for sampling precision, national student samples should provide for a standard error no greater than .035 standard deviation units for the country's mean achievement. This standard error corresponds to a 95% confidence interval of ±7 score points for the achievement mean and of ±10 score points for the difference between achievement means from successive cycles (e.g., the difference between a country's achievement mean on TIMSS 2015 and TIMSS 2019).[1] Sample estimates of any student-level percentage estimate (e.g., a student background characteristic) should have a confidence interval of ±3.5%.

For most countries, the TIMSS precision requirements are met with a school sample of 150 schools and a student sample of 4,000 students for each target grade. Depending on the average class size in the country, one class from each sampled school may be sufficient to achieve the desired student sample size. For example, if the average class size in a country were 27 students, a single class from each of 150 schools would provide a sample of 4,050 students (assuming full participation by schools and students). Some countries choose to sample more than one class per school, either to increase the size of the student sample or to provide a better estimate of school-level effects.

Countries transitioning to eTIMSS require an additional sample of at least 1,500 tested students for the bridge data collection. This bridge sample is obtained by selecting one additional class from a subset of the sampled schools, by selecting a distinct sample of schools, or by a combination of both strategies. The most suitable approach is developed with the sampling experts from Statistics Canada during the sampling development stage.

A school sample larger than the minimum of 150 schools may be required under the following circumstances:

- The average class size in a country is so small that, even when sampling more than one classroom per school, it is not possible to reach the student sample size requirements by selecting only 150 schools.
- Previous cycles of TIMSS showed that the sampling precision requirements cannot be met unless a larger school sample is selected.

1 The TIMSS achievement scales were established in 1995 based on the combined achievement distribution of all countries that participated in TIMSS 1995, at each grade level. To provide a point of reference for country comparisons, the scale centerpoint of 500 was located at the mean of the combined achievement distribution. The scale units were chosen so that 100 scale score points corresponded to the standard deviation of the distribution. Accordingly, one standard deviation unit is approximately 100 scale score points.

- Classes within schools are tracked by student performance (more common at eighth grade than at fourth grade). This increases variation between classes in student achievement and can reduce sampling precision. In this situation, it is advisable to sample at least two classrooms per school whenever possible, in addition to sampling more schools.
- A high level of non-response is anticipated, leading to sample attrition and reduced sample size. Note that while a larger school sample helps to maintain sample size in the face of non-response, it does not compensate for non-response bias.

Field Test Sample

Although the TIMSS field test is scheduled in the school year before the year of data collection, the school sample for the field test is drawn at the same time and from the same population of schools as the full sample. The field test sample size requirement is 200 students per field test achievement booklet, and so the total field test sample size is a function of the number of achievement booklets being field tested. For TIMSS 2019, the paperTIMSS field test has five booklets per target grade and so requires a field test sample of 1,000 students at each grade. The eTIMSS field test has five item block combinations (corresponding to the five paperTIMSS booklets) requiring 1,000 students at each grade, and also three block combinations of Problem Solving and Inquiry tasks (PSIs). Each PSI block appears in two different item block combinations, so the three block combinations require a further 300 students, for a total field test sample of 1,300 students.

Participation Rates

To minimize the potential for non-response bias, TIMSS aims for 100 percent participation by sampled schools, classrooms, and students, while recognizing that some degree of non-participation may be unavoidable. For a national sample to be fully acceptable it must have either:

- A minimum school participation rate of 85 percent, based on originally sampled schools AND
- A minimum classroom participation rate of 95 percent, from originally sampled schools and replacement schools AND
- A minimum student participation rate of 85 percent, from sampled schools and replacement schools

  OR
- A minimum combined school, classroom, and student participation rate of 75 percent, based on originally sampled schools (although classroom and student participation rates may include replacement schools)

  Classrooms with less than 50 percent student participation are deemed to be not participating.

Developing and Implementing the National Sampling Plan

Although National Research Coordinators are responsible for developing and implementing national sampling plans, Statistics Canada and the IEA Sampling team work closely with NRCs to help ensure that these sampling plans fully meet the standards set by the TIMSS & PIRLS International Study Center, while also adapting to national circumstances and requirements. National sampling plans must be based on the international two-stage sample design (schools as the first stage and classes within schools as the second stage) and must be approved by Statistics Canada.

TIMSS Stratified Two-Stage Cluster Sample Design

The basic international sample design for TIMSS is a stratified two-stage cluster sample design, as follows:

First Sampling Stage. For the first sampling stage, schools are sampled with probability proportional to their size (PPS) from the list of all schools in the population that contain eligible students. The schools in this list (or sampling frame) may be stratified (sorted) according to important demographic variables. Schools for the field test and data collection are sampled simultaneously using a systematic random sampling approach. Two replacement schools are also pre-assigned to each sampled school during the sample selection process, and these replacement schools are held in reserve in case the originally sampled school refuses to participate. Replacement schools are used solely to compensate for sample size losses in the event that the originally sampled school does not participate. School sampling is conducted for each country by Statistics Canada with assistance from the IEA Sampling Team, using the sampling frame provided by the country's National Research Coordinator.

Second Sampling Stage. The second sampling stage consists of selecting one (or more) intact class from the target grade of each participating school. Class sampling in each country is conducted by the National Research Coordinator using the Within-School Sampling Software (WinW3S) developed by IEA Hamburg and Statistics Canada. Having secured a sampled school's agreement to participate in the assessment, the NRC requests information about the number of classes and teachers in the school and enters it in the WinW3S database. Classes smaller than a specified minimum size are combined into pseudo-classes prior to sampling. The software samples one or more classes with equal probability in each school. All students in each sampled class participate in the assessment. Sampled classes that refuse to participate may not be replaced.

An additional sampling step is required for eTIMSS countries that require a bridge sample. Students in the bridge sample are administered a paper version of the trend item blocks, and it is important that this sample should mirror the main eTIMSS sample as closely as possible. For operational reasons it is not possible to administer both the eTIMSS assessment and bridge assessment in the same class, so the bridge sample should consist of an extra class from a school sampled for eTIMSS or from an additional school. In schools selected for both the eTIMSS and the bridge samples, separate classes are sampled and randomly assigned to either the eTIMSS or bridge samples using the WinW3S software.

Stratification

Stratification consists of arranging the schools in the target population into groups, or strata, that share common characteristics such as geographic region or school type. Examples of stratification variables used in TIMSS include region of the country (e.g., states or provinces); school type or source of funding (e.g., public or private); language of instruction; level of urbanization (e.g., urban or rural area); socioeconomic indicators; and school performance on national examinations.

In TIMSS, stratification is used to:

- Improve the efficiency of the sample design, thereby making survey estimates more reliable
- Apply different sample designs, such as disproportionate sample allocations, to specific groups of schools (e.g., those in certain states or provinces)
- Ensure proportional representation of specific groups of schools in the sample

School stratification can take two forms: explicit and implicit. In explicit stratification, a separate school list or sampling frame is constructed for each stratum and a sample of schools is drawn from that stratum. In TIMSS, the major reason for considering explicit stratification is disproportionate allocation of the school sample across strata. For example, in order to produce equally reliable estimates for each geographic region in a country, explicit stratification by region may be used to ensure the same number of schools in the sample for each region, regardless of the relative population size of the regions.

Implicit stratification consists of sorting the schools by one or more stratification variables within each explicit stratum, or within the entire sampling frame if explicit stratification is not used. The combined use of implicit strata and systematic sampling is a simple and effective way of ensuring a proportional sample allocation of students across all implicit strata. Implicit stratification also can lead to improved reliability of achievement estimates when the implicit stratification variables are correlated with student achievement.

National Research Coordinators consult with Statistics Canada and the IEA Sampling team to identify the stratification variables to be included in their sampling plans. The school sampling frame is sorted by the stratification variables prior to sampling schools so that adjacent schools are as similar as possible. Regardless of any other explicit or implicit variables that may be used, the school size is always included as an implicit stratification variable.

To document the stratification variables used in their sampling plans, each NRC completes Sampling Form 3, which lists the variables to be used for explicit and implicit stratification, and the number of levels of each stratification variable. An example of a completed Sampling Form 3 is presented in Exhibit 3.3. Further details on the explicit and implicit stratification variables for each country can be found in the Characteristics of National Samples section in Chapter 9: Sampling Implementation.

Exhibit 3.3: Example of Sampling Form 3

Sampling Form 3 Stratification

See Section 4 of TIMSS 2019 Survey Operations Procedures Unit 1

TIMSS 2019 Participant :	*Country X*

1. This Sampling Form refers to: *TIMSS Grade 4 Assessment*

Stratification of schools

2. List and describe the variables to be used for stratification in order of importance:
   (Please note that the choice of variables used for explicit or implicit stratification will be discussed during consultations with the TIMSS sampling experts

Stratification Variables			
	Name	**Description**	**# of levels**
1	*School type*	*public, private*	*2*
2	*Socioeconomic status*	*high, medium, low*	*3*
3			
4			
5			
6			

Include additional information if necessary:

3. If applicable, describe any additional requirements for sub-national estimates, either for reporting or analysis purposes (e.g., oversampling of specific groups of the population):

would like to have reliable estimates for students from the private schools

School Sampling Frame

One of the National Research Coordinator's most important sampling tasks is the construction of a school sampling frame for the target population. The sampling frame is a list of all schools in the country that have students enrolled in the target grade, and is the list from which the school sample is drawn. A well-constructed sampling frame provides complete coverage of the national target population without being contaminated by incorrect or duplicate entries or entries that refer to elements that are not part of the defined target population.

A suitable school measure of size (MOS) is a critical aspect of the national sampling plan, because the size of a school determines its probability of selection. The most appropriate school measure of size is an up-to-date count of the number of students in the target grade. If the number of students in the target grade is not available, total student enrollment in the school may be the best available substitute.

Sampling Form 4, presented in Exhibit 3.4, provides some basic information about the school sampling frame, including the average class size at the target grade, the number of classrooms to be sampled per school, the school measure of size (MOS) to be used for school sampling, and the school year from which the frame was constructed.

Exhibit 3.4: **Example of Sampling Form 4**

Sampling Form 4 — Classroom Information and Sampling Frame

See Section 5 of TIMSS 2019 Survey Operations Procedures Unit 1

TIMSS 2019 Participant:	*Country X*

1. This Sampling Form refers to: *TIMSS Grade 4 Assessment*

2. Specify the school measure of size (MOS) to be used.

   Click in box and on right arrow to see drop down menu

   1. Number of students in the target grade (preferred)

   Name of the MOS variable in the school frame: *GR4_STD*

   If "Other," please describe:

3. Specify the average class size (ACS) for the target grade in your schools. *24*

4. Specify how many classrooms you plan to sample per school. *(Click in box and on right arrow to see drop down menu)*

   2. More than one classroom in tracked schools

   If "Other," please describe:

5. Specify the school year for which enrollment data will be used for the school MOS. *2017/2018*

6. If a frame other than a single-level sampling frame (list of all schools) is to be used, please provide a preliminary description of the information available to construct this frame.

   Not applicable

The school sampling frame is usually a spreadsheet containing a single entry for each school. This entry includes a unique identification number and contact information (if appropriate given the country's privacy laws), the values of the stratification variables for the school, and the school measure of size. It is useful if the school entry also includes the number of classes in the school in the target grade because this provides a mechanism for predicting in advance the size of the eventual student sample. This predicted sample size may be compared with the eventual student sample size as a check on the sampling process.

Exhibit 3.5 provides an example of a partial sampling frame for a country conducting TIMSS 2019 at the eighth grade. In this example, region and urbanization are used as stratification variables.

Exhibit 3.5: Example of a Partial Sampling Frame

	A	B	C	D	E	F	G	H	I	J
1	School ID	Region	Urbani-zation	Grade 8 Students	Grade 8 Classes	School Name	School Address	Postal code	Town	Tel
2	15104	South	Rural	211	8	Campbell College	Jelly Bean Ave 23	01604	Dinsdale	040 / 5699
3	15113	North	Rural	176	7	Stromboli High School	Barracuda Street 5	01611	Lowrie	040 / 5666
4	15115	North	Rural	182	7	Central Park School	Wales Crescent 45	01600	Kristin	041 / 5599
5	15123	North	Urban	104	4	Obi Wan School	Wheel Crescent 23	01903	Curtain	040 / 5000
6	15933	North	Rural	228	9	Alfred Hitchcock High School	Dennis Street 45	01600	Tortilla Plains	041 / 5566
7	15937	North	Urban	186	7	Begonia High School	Morning Street 125	01614	Peacew	040 / 5644
8	15940	North	Urban	153	6	Calmar High School	Casey Crescent 1	01905	Waltington	040 / 5633
9	15942	North	Urban	169	7	Western High School	Travis Ave 54	01905	Waltington	040 / 5644
10	15944	North	Urban	8	1	Manhattan College	Launcaster Street 63	01614	Peacew	040 / 5577
11	15945	South	Rural	229	9	Karaoke High School	Bean Street 45	01614	Blue Lake	040 / 5700
12	15946	South	Rural	164	7	J. Oliver High Cuisine School	Cambridge Crescent 136	01905	Cinder	049 / 5777
13	15953	South	Urban	89	4	Douglas College	Douglas Drive 78	01619	Hawn	049 / 5762
14	15956	South	Urban	22	1	Emily Dickinson College	Phillip Glass Avenue 23	01619	Hawn	049 / 5645
15	15958	North	Urban	65	3	Tinsdale College	McGyver Crescent 49	01903	Curtain	040 / 5811
16	15968	South	Urban	34	1	Gualajara District High School	Strong Street 79	01615	Flowerburgh	040 / 5612
17	15970	South	Urban	188	8	Dry Creek School	Galloway Street 46	01615	Flowerburgh	040 / 5295
18	15974	South	Rural	6	1	Eagle College	Monday Street 123	01614	Candid	040 / 5774
19	15981	South	Rural	81	3	St John High School	Alec Baldwin Drive 75	01617	Holster	040 / 5511
20	15983	South	Rural	88	4	Kum Ba Yah High School	O'Malley Circuit 56	01901	Book Haven	049 / 5693
21	15984	South	Rural	54	2	La Gioconda College	Dodo Bank 45	01616	Kathleen River	049 / 5709
22	15985	South	Urban	45	2	Lake Titicaca College	Collin Benjamin Street 1	01900	Evans	049 / 5622
23	15986	South	Rural	213	9	Paul Bunyan High School	Heidelberg Street 100	01905	Charpwood	049 / 5767
24	15988	South	Rural	290	12	Lynn High School	Good Street 45	01601	Heintz	049 / 5639
25	[illegible]		[illegible]	128	5	Fruit Tree High School	[illegible]	01615	Karburetta	049 / 5611
				[illegible]	9	E. Cochr[illegible]		[illegible]	Garden Heights	[illegible]

Sampling Schools

Once the school sampling frame is structured to meet all international and national requirements, Statistics Canada can draw the school sample. If the sampling frame is explicitly stratified, it is necessary to decide how the school sample is to be allocated among the explicit strata (i.e., the number of schools to be sampled in each stratum). When this has been decided, a sample of schools is selected within each explicit stratum using systematic sampling with probabilities proportional to size (PPS). The PPS technique means that the larger schools, those with more students, have a higher probability of being sampled than the smaller schools. However, this difference in the selection probabilities of larger and smaller schools is largely offset at the second stage of sampling by selecting a fixed number of classes (usually one or two) with equal probability from the sampled school. Classes in large schools with many classes at the target grade have a lower probability of selection than classes in smaller schools that have just one or two classes. A description of the school sampling procedure is provided in Appendix 3A.

Even though the field test is scheduled in the school year before the year of data collection in most countries, the preferred approach in TIMSS is to select both samples of schools at the same time. This ensures that both the field test and data collection samples constitute random samples representative of all schools in the country, and that no school is selected for both samples.[2]

Replacement Schools. Ideally, all schools sampled for TIMSS should participate in the assessments, and NRCs work hard to achieve this goal. Nevertheless, it is anticipated that a 100 percent participation rate may not be possible in all countries. To avoid sample size losses, the sampling plan identifies, *a priori*, specific replacement schools for each sampled school. Each originally sampled school has two pre-assigned replacement schools, usually the school immediately preceding the originally sampled school on the school sampling frame and the one immediately following it. Replacement schools always belong to the same explicit stratum as the original but may come from different implicit strata if the school they are replacing is either the first or last school of an implicit stratum.

The main justification for replacement schools in TIMSS is to ensure adequate sample sizes for analysis of subpopulation differences. Although the use of replacement schools does not eliminate the risk of bias due to school nonparticipation, employing implicit stratification and ordering the school sampling frame by school size increases the chances that a sampled school's replacements would have similar characteristics. This approach maintains the desired sample size while restricting replacement schools to strata where nonresponse occurs. Since the school frame is ordered by school size, replacement schools also tend to be similar in size to the school they are designated to replace.

NRCs understand that they should make every effort to secure the participation of all of the sampled schools. Only after all attempts to persuade a sampled school to participate have failed is the use of its replacement school considered.

Common Adjustments to the TIMSS School Sampling Design

The TIMSS school sample design offers considerable flexibility to countries participating at both fourth and eighth grades to maximize or minimize the extent to which the same schools are assessed. Where fourth and eighth grade students attend the same school, some countries find it more efficient to administer TIMSS at the same school for both grades. In other cases, countries try to ensure that assessments are spread across schools and therefore prefer that TIMSS at the fourth and eighth grades are not administered at the same school and/or that TIMSS sampling avoid, when possible, selecting schools that have recently administered other national and international assessments. To provide flexibility to

2 With approval from the TIMSS & PIRLS international Study Center, the field test and full sample could be selected separately. In such cases an overlap control procedure is used to minimize the probability of selecting schools for the data collection that already had been sampled for the field test. This was the case for most eTIMSS countries due to operational constraints.

meet these requests, Statistics Canada implements modified sampling procedures—the details of which are described in Appendix 3B.

Sampling Classes

Within each sampled school, all classes with students at the target grade are listed, and one or more intact classes are selected with equal probability of selection using systematic random sampling. This procedure is implemented using the WinW3S sampling software. The selection of classes with equal probability, combined with the PPS sampling method for schools, in general results in a self-weighting student sample. If the school has multi-grade classes (i.e., the class contains students from more than one grade level), only students from the target grade are eligible for sampling.

When a country participating in eTIMSS has schools selected for both the eTIMSS and the bridging assessments, sampled classes within these schools are randomly assigned to one study or the other. This is done automatically within the WinW3S software.

Because small classes tend to increase the risk of unreliable survey estimates and can lead to reduced overall student sample size, it is necessary to avoid sampling too many small classes. Based on consideration of the size distribution of classes and the average class size, a lower class size limit or minimum class size (MCS) is specified for each country. Prior to sampling classes in a school, any class smaller than the MCS is combined with another class in the school to form a pseudo-class for sampling purposes. The procedure for sampling classes within schools is described in more detail in the Survey Operations Procedures chapter of this volume.

Sampling Weights

National student samples in TIMSS are designed to accurately represent the target populations within a specified margin of sampling error, as described previously. After the data have been collected and processed, sample statistics such as means and percentages that describe student characteristics are computed as weighted estimates of the corresponding population parameters, where the weighting factor is the sampling weight. A student's sampling weight is essentially the inverse of the student's probability of selection, with appropriate adjustments for nonresponse. In principle, the stratified two-stage sampling procedure used in TIMSS, where schools are sampled with probability proportional to school size and classes are sampled with probability inversely proportional to school size, provides student samples with equal selection probabilities. However, in practice, disproportionate sampling across explicit strata by varying the number of classes selected and differential patterns of nonresponse can result in varying selection probabilities, requiring a unique sampling weight for the students in each participating class in the study.

The student sampling weight in TIMSS is a combination of weighting components reflecting selection probabilities and sampling outcomes at three levels—school, class, and student. At each level, the weighting component consists of a basic weight that is the inverse of the probability of selection at that level, together with an adjustment for nonparticipation. The overall sampling weight for each student is the product of the three weighting components: school, class (within school), and student (within class).

Usually in TIMSS a country has only one set of sampling weights per target population (fourth and/or eighth grade). However, because of the introduction of the Problem Solving and Inquiry (PSI) tasks into the eTIMSS 2019 booklet rotation, eTIMSS countries have one set of sampling weights only for students who were assigned regular eTIMSS booklets, and a second set for all students, including those assigned PSI booklets. The first set of weights is computed in the same way for both eTIMSS and paperTIMSS countries (since the paperTIMSS countries do not use the PSI booklets) and is used for most analytic and reporting purposes in TIMSS 2019. Where necessary, these are referred to as the "TIMSS weights" to distinguish them from the second set, or "TIMSS+PSI weights."

In addition to the weights described above, countries with bridge data have a further set of weights exclusively for the bridge sample. Further details on the special weight adjustments for eTIMSS and the bridge data may be found in Chapter 9: Sampling Implementation. Regardless of whether they pertain to the regular booklet sample, the regular and PSI booklet sample, or the bridge sample, the procedure for calculating weights and nonparticipation adjustments remains the same.

School Weighting Component

Given that schools in TIMSS are sampled with probability proportional to school size, the basic school weight for the i^{th} sampled school (i.e., the inverse of the probability of the i^{th} school being sampled) is defined as:

$$BW_{sc}^{i} = \frac{M}{n \cdot m_i} \tag{3.1}$$

where n is the number of sampled schools, m_i is the measure of size for the i^{th} school, and

$$M = \sum_{i=1}^{N} m_i \tag{3.2}$$

where N is the total number of schools in the explicit stratum.[3,4]

School Nonparticipation Adjustment. If a sampled school does not participate in TIMSS and its two designated replacement schools do not participate, it is necessary to adjust the basic school weight to

3 For countries such as the Russian Federation that include a preliminary sampling stage, the basic school weight also incorporates the probability of selection in this preliminary stage. The basic school weight in such cases is the product of the preliminary stage weight and the school weight.

4 In schools selected for both the eTIMSS and the bridge samples, sampled classes are randomly assigned to either the eTIMSS or the bridge samples using the WinW3S software. If such a school has only one class, WinW3S randomly assigns the class to one of the samples (eTIMSS or bridge). In such cases, an adjustment is applied to the school weight in the corresponding explicit stratum of the non-selected sample.

compensate for the reduction in sample size. The school-level nonparticipation adjustment is calculated separately for each explicit stratum, as follows:

$$A_{sc} = \frac{n_s + n_{r1} + n_{r2} + n_{nr}}{n_s + n_{r1} + n_{r2}} \tag{3.3}$$

where n_s is the number of originally sampled schools that participated, n_{r1} and n_{r2} the number of first and second replacement schools, respectively, that participated, and n_{nr} is the number of schools that did not participate. Sampled schools that are found to be ineligible[5] are not included in the calculation of this adjustment.

Combining the basic school weight and the school nonparticipation adjustment, the final school weighting component for the i^{th} school becomes:

$$FW_{sc}^{i} = A_{sc} \cdot BW_{sc}^{i} \tag{3.4}$$

It should be noted that, as well as being a crucial component of the overall student weight, the final school weighting component is a sampling weight in its own right, and can be used in analyses where the school is the unit of analysis.

Class Weighting Component

The class weighting component reflects the class-within-school selection probability. After a school has been sampled and has agreed to participate in TIMSS, one or more classes are sampled with equal probability from the list of all classes in the school at the target grade. Because larger schools have more classes from which to sample than smaller schools, the probability of class selection varies with school size, with students in small schools more likely to have their class selected than students in large schools. This relatively greater selection probability for students in small schools offsets their lower selection probability at the first stage, where probability-proportional-to-size school sampling results in higher selection probabilities for larger schools.

The basic class-within-school weight for a sampled class is the inverse of the probability of the class being selected from all of the classes in its school. For the i^{th} sampled school, let C^i be the total number of eligible classes and c^i the number of sampled classes. Using equal probability sampling, the basic class weight for all sampled classes in the i^{th} school is:

$$BW_{cl}^{i} = \frac{C^i}{c^i} \tag{3.5}$$

For most TIMSS participants, c^i takes the values 1 or 2.

5 A sampled school is ineligible if it is found to contain no eligible students (i.e., no students in the target grade). Such schools usually are in the sampling frame by mistake or are schools that recently have closed.

Class Nonparticipation Adjustment. Basic class weights are calculated for all sampled classes in the sampled and replacement schools that participate in TIMSS. A class-level nonparticipation adjustment is applied to compensate for classes that do not participate or where the student participation rate is below 50 percent.[6] Such sampled classes are assigned a weight of zero. Class nonparticipation adjustments are applied at the explicit stratum level rather than at the school level to minimize the risk of bias. The adjustment is calculated as follows:

$$A_{cl} = \frac{\sum_{i}^{s+r1+r2} 1}{\sum_{i}^{s+r1+r2} \delta_i / c^i} \tag{3.6}$$

where c^i is the number of sampled classes in the i^{th} school, as defined earlier, and δ_i gives the number of participating classes in the i^{th} school.

Combining the basic class weight and the class nonparticipation adjustment, the final class weighting component, assigned to all sampled classes in the i^{th} school, becomes:

$$FW_{cl}^{i,j} = A_{cl} \cdot BW_{cl}^{i} \tag{3.7}$$

Student Weighting Component

The student weighting component represents the student-within-class selection probability. The basic student weight is the inverse of the probability of a student in a sampled class being selected.

In the typical TIMSS situation where intact classes are sampled, all students in the class are included, and so this probability is unity. However, under certain circumstances, students may be sampled within the class, and in this situation the probability is less than unity.

It should be noted that within-class student sampling is in effect when calculating the weights for the regular eTIMSS booklets for eTIMSS countries (the "TIMSS weight"). In this situation, students who were assigned a regular eTIMSS booklet are considered as being selected while students who received a PSI booklet are considered as not selected.

For an intact class with no student subsampling, the basic student weight for the j^{th} class in the i^{th} school is computed as follows:

$$BW_{st1}^{i,j} = 1.0 \tag{3.8}$$

6 When calculating the weights for the sample with regular TIMSS booklets only (without the PSI booklets), the nonparticipation criterion of below 50% is based on the full class, including the PSI booklets. Therefore, if 50% or more students from a class participated, regardless of the type of booklet received, the class is considered as participating when calculating the weights for the regular booklets only sample (the "TIMSS weight").

For classes with student subsampling, the basic student weight for the j^{th} class in the i^{th} school is:

$$BW_{st2}^{i,j} = \frac{n_{rg}^{i,j} + n_{bs}^{i,j}}{n_{rg}^{i,j}} \tag{3.9}$$

Where $n_{rg}^{i,j}$ is the number of students in the j^{th} class of the i^{th} school selected to participate in TIMSS and $n_{bs}^{i,j}$ is the number of students in the class not selected.

Calculating the eTIMSS weights for the regular booklet sample (the "TIMSS weight") involves student subsampling as described above, but with the added complication that the participation status is known for all the students in each sampled class. In this case, the basic student weight for the j^{th} class in the i^{th} school for this set of weights is given by:

$$BW_{st3}^{i,j} = \begin{cases} 1 & \text{for students who left school or were excluded,} \\ \dfrac{n_{rg'}^{i,j} + n_{bs'}^{i,j}}{n_{rg'}^{i,j}} & \text{for all other students who received a regular eTIMSS booklet} \end{cases} \tag{3.10}$$

where, $n_{rg'}^{i,j}$ and $n_{bs'}^{i,j}$ represent the number of students in the j^{th} class of the i^{th} school who received a regular eTIMSS booklet and the number of students in the j^{th} class of the i^{th} school who received a PSI booklet respectively, without counting students who either were excluded or left school after the class listing was completed.

Adjustment for Non-Participation. The student nonparticipation adjustment for the j^{th} classroom in the i^{th} school is calculated as:

$$A_{st1}^{i,j} = A_{st2}^{i,j} = A_{st3}^{i,j} = \frac{s_{rs}^{i,j} + s_{nr}^{i,j}}{s_{rs}^{i,j}} \tag{3.11}$$

where $s_{rs}^{i,j}$ is the number of participating students in the j^{th} class of the i^{th} school and $s_{nr}^{i,j}$ is the number of students sampled in this class who were expected to have assessment scores but did not participate in the assessment. For intact classes, the sum of $s_{rs}^{i,j}$ and $s_{nr}^{i,j}$ is the total number of students listed in the class, not counting excluded students or students who have left the school since class list was published. When calculating the "TIMSS weight" for eTIMSS countries (without the PSI booklets), the sum of $s_{rs}^{i,j}$ and $s_{nr}^{i,j}$ is the total number of students who received a regular eTIMSS booklet in the class, not counting excluded students or students who have left the school since class list was published

The final student weighting component for students in the j^{th} classroom of the i^{th} school is:

$$FW_{st}^{i,j} = A_{st\Delta}^{i,j} \cdot BW_{st\Delta}^{i,j} \tag{3.12}$$

where Δ equals 1 when there was no student subsampling (intact classes), 2 when a sample of students was drawn from the students in the class, and 3 when calculating the set of eTIMSS weights for only regular eTIMSS booklets.

Overall Student Sampling Weight. The overall student sampling weight is the product of the final weighting components for schools, classes, and students, as follows:

$$W^{i,j} = FW_{sc}^{i} \cdot FW_{cl}^{i,j} \cdot FW_{st}^{i,j} \tag{3.13}$$

Overall student sampling weights are only attributed to participating students, with non-participants weighted at 0. All student data reported in the TIMSS international reports are weighted by the overall student sampling weight, known as TOTWGT in the TIMSS international databases.

Participation Rates

Because nonparticipation can result in sample bias and misleading results, it is important that the schools, classes, and students that are sampled to participate in TIMSS actually take part in the assessments. To show the level of sampling participation in each country, TIMSS calculates both unweighted participation rates (i.e., based on simple counts of schools, classes, and students) and weighted participation rates based on the sampling weights described in the previous section. Unweighted participation rates provide a preliminary indicator that may be used to monitor progress in securing the participation of schools and classes, whereas weighted participation rates are the ultimate measure of sampling participation.

TIMSS reports weighted and unweighted participation rates for schools, classes, and students, as well as overall participation rates that are a combination of all three. To distinguish between participation based solely on originally sampled schools and participation that also relies on replacement schools, school and overall participation rates are computed separately for originally sampled schools only and for originally sampled together with replacement schools.

Unweighted School Participation Rate

The unweighted school participation rate is the ratio of the number of participating schools to the number of originally sampled schools, excluding any sampled schools found to be ineligible. A school is considered to be a participating school if at least one of its sampled classes has a student participation rate of at least 50 percent. The two unweighted school participation rates are calculated as follows:

R_{unw}^{sc-s} = unweighted school participation rate for originally sampled schools only

R_{unw}^{sc-r} = unweighted school participation rate, including originally sampled and first and second replacement schools

$$R_{unw}^{sc-s} = \frac{n_s}{n_s + n_{r1} + n_{r2} + n_{nr}} \tag{3.14}$$

$$R_{unw}^{sc-r} = \frac{n_s + n_{r1} + n_{r2}}{n_s + n_{r1} + n_{r2} + n_{nr}} \tag{3.15}$$

Unweighted Class Participation Rate

The unweighted class participation rate is the ratio of the number of sampled classes that participated to the number of classes sampled, as follows:

$$R_{unw}^{cl} = \frac{\sum_{i}^{s+r1+r2} c_*^i}{\sum_{i}^{s+r1+r2} c^i} \tag{3.16}$$

where c^i is the number of sampled classes in the i^{th} school, and c_*^i is the number of participating classes in the i^{th} school. Both summations are across all participating schools.

Unweighted Student Participation Rate

The unweighted student participation rate is the ratio of the number of selected students that participated in TIMSS to the total number of selected students that should have been assessed in the participating schools and classes. Classes where less than 50 percent of the students participate are considered to be not participating, and so students in such classes also are considered to be nonparticipants.[7] The unweighted student participation rate is computed as follows:

$$R_{unw}^{st} = \frac{\sum_{i,j} s_{rs}^{i,j}}{\sum_{i,j} s_{rs}^{i,j} + \sum_{i,j} s_{nr}^{i,j}} \tag{3.17}$$

Overall Unweighted Participation Rate

The overall unweighted participation rate is the product of the unweighted school, class, and student participation rates. Because TIMSS computes two versions of the unweighted school participation rate, one based on originally sampled schools only and the other including replacements as well as originally sampled schools, there also are two overall unweighted participation rates:

R_{unw}^{ov-s} = unweighted overall participation rate for originally sampled schools only

7 When calculating the "TIMSS weights" for eTIMSS countries (no PSI booklets), this 50% criteria is applied to all students regardless of the booklet they received.

R_{unw}^{ov-r} = unweighted overall participation rate, including originally sampled and first and second replacement schools

$$R_{unw}^{ov-s} = R_{unw}^{sc-s} \cdot R_{unw}^{cl} \cdot R_{unw}^{st} \tag{3.18}$$

$$R_{unw}^{ov-r} = R_{unw}^{sc-r} \cdot R_{unw}^{cl} \cdot R_{unw}^{st} \tag{3.19}$$

Weighted School Participation Rate

The weighted school participation rate is the ratio of two estimates of the size of the target student population. The numerator is derived from the measure of size of those sampled schools that participated in TIMSS and the denominator is the weighted estimate of the total student enrollment in the population. Weighted school participation rates are computed for originally sampled schools and for originally sampled and replacement schools combined, as follows:

R_{wtd}^{sc-s} = weighted school participation rate for originally sampled schools only

R_{wtd}^{sc-r} = weighted school participation rate, including originally sampled and first and second replacement schools

$$R_{wtd}^{sc-s} = \frac{\sum_{i,j}^{s} BW_{sc}^{i} \cdot FW_{cl}^{i,j} \cdot FW_{st}^{i,j}}{\sum_{i,j}^{s+r1+r2} FW_{sc}^{i} \cdot FW_{cl}^{i,j} \cdot FW_{st}^{i,j}} \tag{3.20}$$

$$R_{wtd}^{sc-r} = \frac{\sum_{i,j}^{s+r1+r2} BW_{sc}^{i} \cdot FW_{cl}^{i,j} \cdot FW_{st}^{i,j}}{\sum_{i,j}^{s+r1+r2} FW_{sc}^{i} \cdot FW_{cl}^{i,j} \cdot FW_{st}^{i,j}} \tag{3.21}$$

Summations in both the numerator and denominator are over all responding students and include appropriate class and student sampling weights. Note that the basic school weight appears in the numerator, whereas the final school weight appears in the denominator.

Weighted Class Participation Rate

The weighted class participation rate is computed as follows:

$$R_{wtd}^{cl} = \frac{\sum_{i,j}^{s+r1+r2} BW_{sc}^{i} \cdot BW_{cl}^{i,j} \cdot FW_{st}^{i,j}}{\sum_{i,j}^{s+r1+r2} BW_{sc}^{i} \cdot FW_{cl}^{i,j} \cdot FW_{st}^{i,j}} \qquad (3.22)$$

where both the numerator and denominator are summations over all responding students from classes with at least 50 percent of their students participating in the study, and the appropriate student-level sampling weights are used. In this formula, the basic class weight appears in the numerator, whereas the final class weight appears in the denominator. The denominator in this formula is the same quantity that appears in the numerator of the weighted school participation rate for all schools, whether originally sampled or replacement.

Weighted Student Participation Rate

The weighted student participation rate is computed as follows:

$$R_{wtd}^{st} = \frac{\sum_{i,j}^{s+r1+r2} BW_{sc}^{i} \cdot BW_{cl}^{i,j} \cdot BW_{st}^{i,j}}{\sum_{i,j}^{s+r1+r2} BW_{sc}^{i} \cdot BW_{cl}^{i,j} \cdot FW_{st}^{i,j}} \qquad (3.23)$$

where both the numerator and denominator are summations over all responding students from participating schools. In this formula, the basic student weight appears in the numerator, whereas the final student weight appears in the denominator. Also, the denominator in this formula is the same quantity that appears in the numerator of the weighted class participation rate for all participating schools, whether originally sampled or replacement.

Overall Weighted Participation Rate

The overall weighted participation rate is the product of the weighted school, class, and student participation rates. Because there are two versions of the weighted school participation rate, one based on originally sampled schools only and the other including replacement as well as originally sampled schools, there also are two overall weighted participation rates:

R_{wtd}^{ov-s} = weighted overall participation rate for originally sampled schools only

R_{wtd}^{ov-r} = weighted overall participation rate, including sampled, first and second replacement schools

$$R_{wtd}^{ov-s} = R_{wtd}^{sc-s} \cdot R_{wtd}^{cl} \cdot R_{wtd}^{st} \tag{3.24}$$

$$R_{wtd}^{ov-r} = R_{wtd}^{sc-r} \cdot R_{wtd}^{cl} \cdot R_{wtd}^{st} \tag{3.25}$$

Weighted school, class, student, and overall participation rates are computed for each TIMSS participant using these procedures.

References

Chowdhury, S., Chu, A., & Kaufman, S. (2000). Minimizing overlap in NCES surveys. In *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association. 147-179. Retrieved from http://www.asasrms.org/Proceedings/papers/2000_025.pdf

UNESCO. (2012). *International Standard Classification of Education (ISCED) 2011*. Montreal, Quebec: UNESCO Institute for Statistics. Retrieved from http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf

Appendix 3A: Sampling Schools

TIMSS employs random-start fixed-interval systematic sampling to draw the school sample, with each school selected with probability proportional to its size (PPS).

To sample schools using the PPS systematic sampling method, the schools from each explicit stratum in the sampling frame are sorted by implicit stratification variables and by their measure of size (MOS), as shown in the example in Exhibit 3.6. The MOS is accumulated from school to school and the running total (the Cumulative MOS) is listed next to each school. The cumulative MOS across the entire stratum (the Total MOS) is a measure of the size of the school population in the stratum (59,614 students in the example).

First Step: Compute the Sampling Interval

Dividing the Total MOS by the number of schools required for the sample (50 in the example) gives the sampling interval.

- **59,614 ÷ 50 = 1,192.2800**

Second Step: Generate a Random Start

Generate a random number from a uniform (0,1) distribution and multiply it by the sampling interval. The school whose cumulative MOS contains the resulting number is the first school in the sample.

- **0.5481 × 1,192.2800 = 653.4887**
- **School 1718**, with cumulative **MOS** of **690**, is the first school in the sample.

Third Step: Identify the Next School in the Sample (repeat until all schools have been sampled)

- Add the sampling interval to the number computed in the previous step.
- **653.4887 + 1,192.2800 = 1,845.7687**
- **School 0067,** with cumulative **MOS** of **1,855**, is the second school in the sample.
- Repeat until all schools have been sampled. For example, to identify the third school:
- **1,845.7687 + 1,192.2800 = 3,038.0487**
- **School 0333**, with cumulative **MOS** of **3,038**, is the third school in the sample.

Fourth Step: Identify Replacement Schools

Two replacement schools are identified for each sampled school. The first replacement (R1) is the school that immediately follows the sampled school in the sampling frame, and the second replacement (R2) the school that immediately precedes the sampled school.

Exhibit 3.6: Example of PPS Systematic Sampling—Schools

Sampling Parameters	
Total Number of Schools:	2,119
Total Measure of Size:	59,614
School Sample Size:	50
Sampling Interval:	1,192.2800
Random Start:	653.4887

First Step

Compute the Sampling Interval:

59,6914 ÷ 50 = 1,192.2800

Second Step

Generate a random start:

0.5481 × 1,192.2800 = 653.4887

Third Step (repeat until complete)

Compute the next selection numbers:

653.4887 + 1,192.2800 = 1,845.7687

1,845.7687 + 1,192.2800 = 3,038.0487

Fourth Step

Identify Replacement Schools

(R1, R2)

School Identifier	School MOS	Cumulative MOS	Sampled Schools
0829	110	110	
0552	101	211	
1802	98	309	
1288	98	407	
2043	95	502	
0974	94	596	R2
1718	94	690	✓
1807	93	783	R1
0457	93	876	
0244	93	969	
1817	91	1,060	
1741	90	1,150	
1652	89	1,239	
0121	89	1,328	
0309	89	1,417	
0032	89	1,506	
0021	89	1,595	
0609	88	1,683	
0399	86	1,769	R2
0067	86	1,855	✓
0202	86	1,941	R1
0063	86	2,027	
1467	86	2,113	
1381	86	2,199	
1043	84	2,283	
1318	84	2,367	
0659	84	2,451	
0612	83	2,534	
1696	82	2,616	
0867	82	2,698	
0537	81	2,779	
1794	80	2,859	
0695	80	2,939	
0031	80	3.019	R2
0333	79	3,098	✓
0051	79	3,177	R1
0384	79	3,256	
1361	79	3,335	
1189	79	3,414	
0731	78	3,492	
0634	78	3,570	
1230	77	3,647	

Appendix 3B: School Sampling Design Options to Accommodate Other Samples

TIMSS provides optional modifications to its sampling design for countries that want to maximize or minimize sampling overlap between schools sampled by TIMSS at the fourth and eighth grades as well as for countries that want to minimize overlap between schools sampled for TIMSS and schools sampled for other national or international assessments.

To provide options for countries in designing their school samples, Statistics Canada implements two special sampling procedures. Method A is applied when data collection occurs simultaneously for two or more populations (as is the case in 2019 with TIMSS at fourth grade and eighth grade) and the country wants to control the overlap between the schools. Method B is used primarily to ensure that the TIMSS samples avoid schools sampled for other studies, and also used when Method A is not appropriate.

Sampling Method A: Sampling Modifications for Simultaneous Data Collection

This procedure stratifies the school population according to whether schools contain students from both populations to be sampled (fourth and eighth grades, for example), or students from one population only (fourth grade only or eighth grade only) as a way of controlling sample overlap. Each school is assigned a measure of size (MOS) based on the number of students in the two populations combined (i.e., fourth grade and eighth grade combined). Schools are sampled according to the sampling design described in this chapter. When selecting schools from strata comprising students from both populations, a country can choose to maximize or minimize the number of schools to be sampled at each grade level.

The example below in Exhibit 3.7 shows a hypothetical country participating in TIMSS at both grades. For reasons of administrative efficiency, the country wants to maximize the overlap between the fourth and eighth grade school samples. The 8,805 schools from the combined school frames (fourth and eighth grades) were first split in three strata and then a school sample of 164 was drawn as shown in the exhibit.

Exhibit 3.7: Example of Method A - Allocation of School Samples in a Country Participating at Two Grade Levels

Overlap Strata	Total Sampled Schools	Allocation	
		To TIMSS Grade 4	To TIMSS Grade 8
Grade 4 only	14	14	0
Grade 8 only	14	0	14
Grade 4 & Grade 8	136	136	136
Total	**164**	**150**	**150**

Choosing as many schools as possible from the Grade 4 & Grade 8 stratum resulted in a sample of 150 schools (136+14) for each grade level, from a total of 164 sampled schools. In this case, both studies were administered in the 136 schools selected from the Grade 4 & Grade 8 stratum.

This sampling technique was most often used for TIMSS countries and benchmarking participants that had schools with students in both fourth and eighth grade populations, where there was a strong correlation between the measure of size at both grades across these schools, and when school samples could be drawn at the same time.

Sampling Method B: Sampling Modifications for Sequential Data Collection

Method B was used to minimize overlap with another study such as a national study that also samples schools, and was also used when Method A was not appropriate (e.g., low correlation between MOS for fourth grade and eighth grade, samples not drawn simultaneously). In Method B, schools were sampled using a technique described in Chowdhury, Chu, and Kaufman (2000). As explained by the authors, the method can be used to either minimize or maximize overlap amongst several samples. This method is illustrated below with an example where the aim was to minimize the overlap between a current sample of schools S_2 and a previously selected school sample S_1. (For a complete description of the method, readers are referred to the original paper).

Let RL (Response Load) be the number of times a school was sampled from previous samples. In this example, given that there is only one previous sample, RL takes the value 1 if the school was already selected and 0 otherwise.

Given that the RL variable splits the current school frame in two distinct subsets of schools, S_1 where RL=1 and $\overline{S}_1$ where RL=0, we have the following relation:

$$P_i(S_2) = P_i(S_2|S_1) \cdot P_i(S_1) + P_i(S_2|\overline{S}_1) \cdot P_i(\overline{S}_1) \tag{3.26}$$

where $P_i(S_j)$ gives the probability that school i be selected in the sample (S_j), and $P_i(S_j|S_k)$ gives the probability that school i be selected in sample (S_j) given that school i already belongs to (S_k). The idea here is to derive the conditional probabilities in such a way that the unconditional probability of selecting a school in the current sample, $P_i(S_2)$, be equal to the expected probability (as defined by the TIMSS sample design).

Note that the first term after the equal sign in equation (3.26) is related to cases where the school response load is 1, while the last term is related to cases where the school response load is 0. Therefore, minimizing the sample overlap is equivalent to zeroing the first term. In such case, equation (3.26) becomes:

$$P_i(S_2) = 0 \cdot P_i(S_1) + P_i(S_2|\bar{S}_1) \cdot P_i(\bar{S}_1) \quad (3.27)$$

and consequently,

$$P_i(S_2|\bar{S}_1) = P_i(S_2)/P_i(\bar{S}_1) \quad (3.28)$$

In other words, in the current sample S_2, schools would be selected with the following conditional probabilities:

$$\begin{cases} 0 & \text{if school } i \text{ was already selected in the first sample,} \\ P_i(S_2)/P_i(\bar{S}_1) & \text{otherwise} \end{cases} \quad (3.29)$$

However, equation (3.26) no longer holds if expression $P_i(S_2)/P_i(\bar{S}_1)$ is greater than 1. This can be avoided by setting 1 as an upper bound. We now have the following expression:

$$P_i(S_2) = P_i(S_2|S_1) \cdot P_i(S_1) + 1 \cdot P_i(\bar{S}_1) \quad (3.30)$$

and consequently

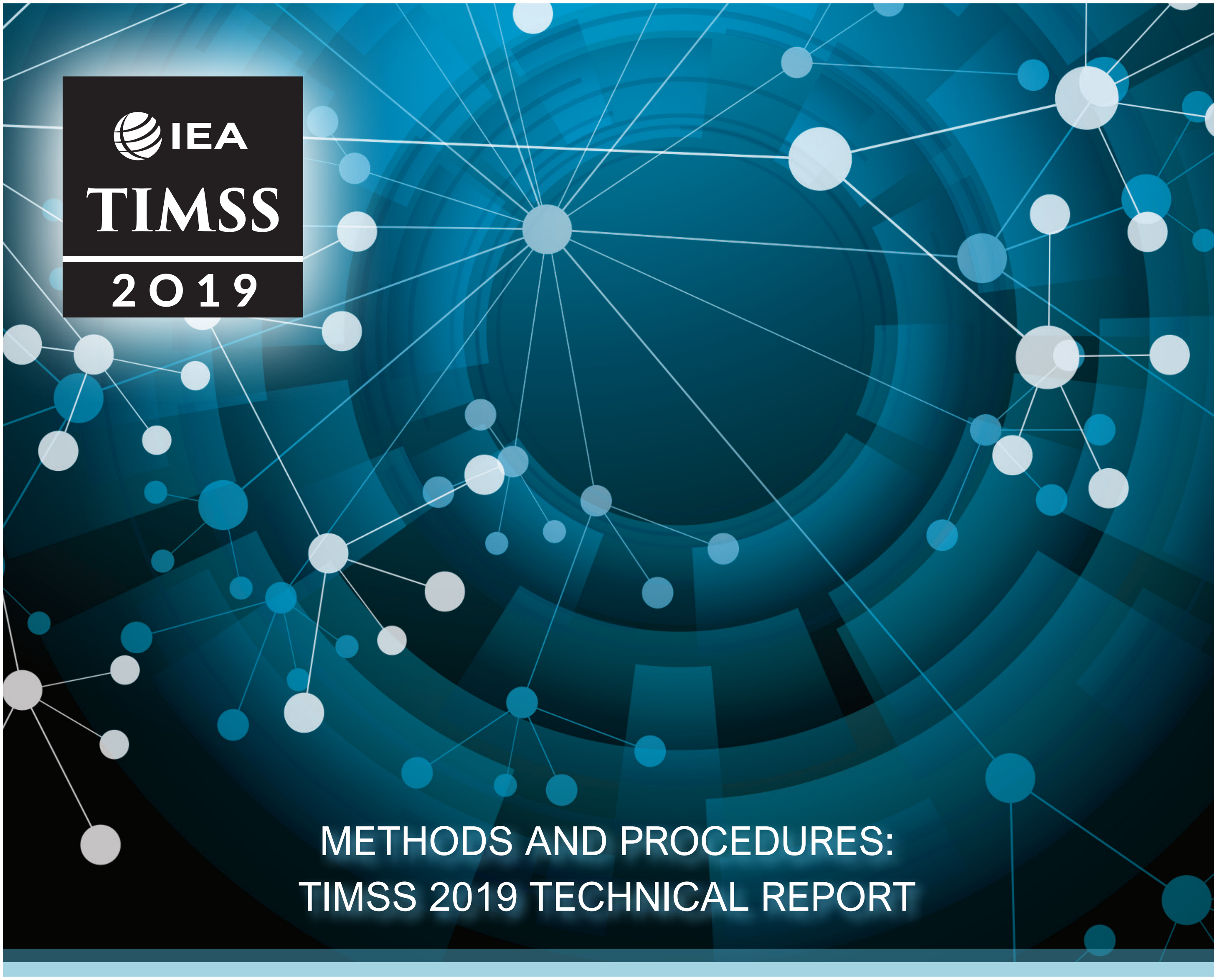
$$\frac{P_i(S_2) - P_i(\bar{S}_1)}{P_i(S_1)} = P_i(S_2|S_1) \quad (3.31)$$

Combining these two results, the conditional probabilities to use when selecting the current sample of schools are given by:

$$\begin{cases} \text{Max} \left[0, \dfrac{P_i(S_2) - P_i(\bar{S}_1)}{P_i(S_1)} \right] & \text{if school } i \text{ was already selected in the first sample,} \\ \text{Min} \left[\dfrac{P_i(S_2)}{P_i(\bar{S}_1)}, 1 \right] & \text{otherwise} \end{cases} \quad (3.32)$$

Note that maximizing rather than minimizing the overlap between two studies can be done by simply zeroing the last term of equation (3.26) rather than zeroing the first term, and following the above logic to get the conditional probabilities. The Chowdhury et al. (2000) method can be generalized to more than two samples as described in their paper.

Further details about the implementation of this method for the countries and benchmark participants can be found in the Sample Implementation chapter.



DATA COLLECTION PROCEDURES

IEA
TIMSS & PIRLS
International Study Center
Lynch School of Education
BOSTON COLLEGE

CHAPTER 4

eAssessment System for TIMSS 2019

Mark Cockle
Heiko Sibberns

Introduction

As described in Chapter 1 of this volume, TIMSS 2019 marked the beginning of the transition to eTIMSS—the digital version of TIMSS designed for computer- and tablet-based administration. eTIMSS offered an engaging, interactive, and visually attractive assessment that enabled TIMSS 2019 to better assess complex areas of the mathematics and science frameworks and increase operational efficiency in translation, assessment delivery, data entry, and scoring. Although the aim is to switch completely to the new digital mode in future assessment cycles, in recognition of the different levels of preparation and infrastructure, countries had the option in 2019 of choosing either eTIMSS or paperTIMSS.

In addition to the overarching requirements for a computer-based system that could produce attractive and engaging assessment items while being reliable, flexible, and easy to use, there were a number of other conditions that had to be taken into account in choosing the system:

- The assessment should be capable of operating on tablets as well as on personal computers.
- Assessment delivery should be via USB memory sticks or through a local server approach whereby the assessment software is installed on a local server that can be accessed by a small number of clients (no more than 30). Full internet-based administration was not a requirement for the TIMSS 2019 assessment cycle.
- Because about half the countries were administering the paperTIMSS version, it was important that items developed in the eAssessment system be as similar as possible to their corresponding paper versions, while capitalizing on interactive computer-based features such as drag-and-drop, multi-select, and drop-down menus.
- Beyond the utilization of features just mentioned in developing individual items, the system should also accommodate more extended Problem Solving and Inquiry Tasks (PSIs) designed to simulate real world or laboratory situations in which students could integrate and apply process skills and content knowledge to solve mathematics problems or conduct virtual

scientific experiments and investigations. These tasks would be tailor-made and not have any counterparts in the paperTIMSS assessment.

- Great emphasis also was placed on clarity and ease of use of the student interface, which was to be kept as simple as possible. This meant, for example, that only those tools such as rulers or calculators that were necessary for processing a specific task were available.
- Since the TIMSS assessment has to be translated and adapted to the needs of each country and language while retaining the same user experience, it was important that the system incorporate a preview functionality for checking that the assessment content appears in exactly the way it is intended for the assessment situation.

To meet all of the design requirements and constraints and to adequately take into account the workflow that has been optimized by TIMSS during the last 20 years, it was decided to develop the eTIMSS computer-based assessment system in-house instead of using an existing commercial system or having it developed by an external company. Accordingly, the eTIMSS "eAssessment system" was designed and implemented by the software team at IEA Hamburg, with input from the TIMSS & PIRLS International Study Center on the user experience/user interface and from IEA Amsterdam on translation issues.

The TIMSS 2019 eAssessment system consisted of a number of integrated software and application modules as follows:

- The **Designer** is an item authoring system used by the TIMSS & PIRLS International Study Center to develop the eTIMSS achievement items
- The **Assembler** was used to group items into item blocks and item blocks into student "item block combinations" (student booklet equivalents)
- The **Translation System** was used by National Research Coordinators (NRCs) from each country and benchmarking participant to translate the items into their language(s) of instruction and by IEA Amsterdam and the TIMSS & PIRLS International Study Center for translation and layout verification, respectively
- The assessment **Player** was used to administer the eTIMSS assessment—present the items on tablet or computer, record students' responses, and upload the data to the IEA servers
- The **Data Monitor** was used by NRCs and test administrators to check the status of uploaded material and progress of the data collection
- The **Scoring System** was used by NRCs and their scoring staff to review students' written responses to constructed-response items and score them according to the eTIMSS scoring guides.

Design and Architecture of the eTIMSS Modules

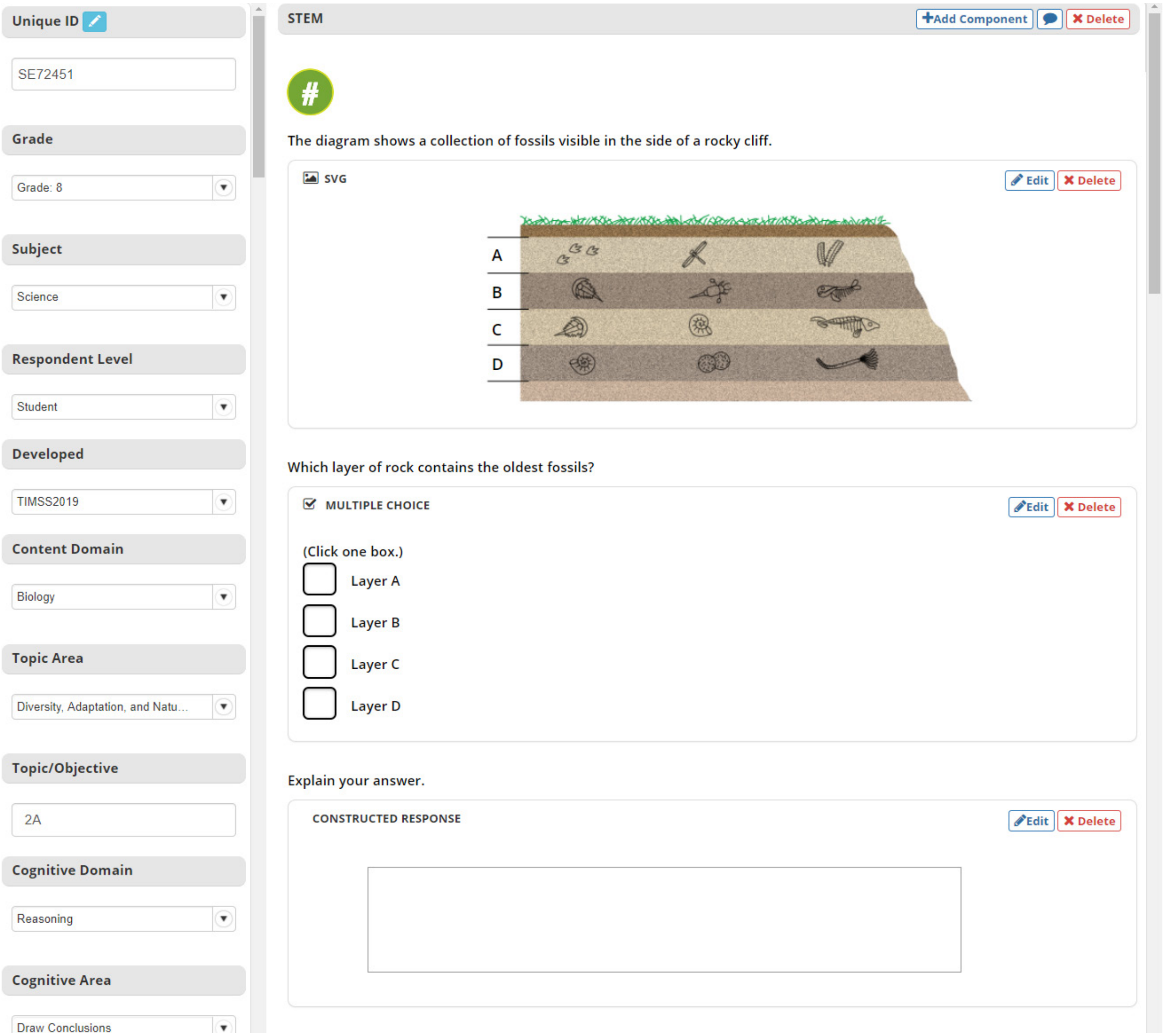
In considering the description of the TIMSS eAssessment system it is helpful to differentiate among three distinct subsystems: 1) the production system for creating assessment content (the Designer, the Assembler, and the Translation System); 2) the delivery system for administering the test in the test session (the assessment Player); and 3) the retrieval and processing system, for upload of the test data to the IEA servers, scoring, and further data processing.

The following is an overview of the various system components and their interaction.

Designer

The designer was used by staff at the TIMSS & PIRLS International Study Center to create the digital versions of the standard (non-PSI) items. Exhibit 4.1 shows part of the screen used for the creation of items. In this example, the item includes three separate elements: a Scalable Vector Graphic (SVG) image, a multiple-choice option, and a constructed response field. On the left is a column for item properties, including, amongst other information, the item ID number, testing grade, testing subject, and content domain assessed by the item. These were used for the selection and identification of the item later in the process of assembly and booklet creation, as well as within the Player.

Exhibit 4.1: eAssessment Designer Input Screen for Example Item



The objective in developing the Designer was to create an item authoring system that encompassed existing TIMSS paper item formats (multiple-choice, constructed response etc.), including the stem text, images and so on, but also new item types unique to the electronic environment. These included drop-down menus, drag & drop, selection (boxes or images), and sorting (boxes or images) item types (see *TIMSS 2019 Item Writing Guidelines*). To accommodate items where the student had to draw shapes or lines, a line-drawing grid also was introduced.

The Designer included various features that could be used by item developers in creating or customizing items, which was particularly important when dealing with trend items where a close match between the electronic item and the paper version was required. For example, the application of labels to images could be made above, below, or to the left or right of an image, or tables could be inserted with invisible lines in order to place objects within columns to obtain a more precise layout.

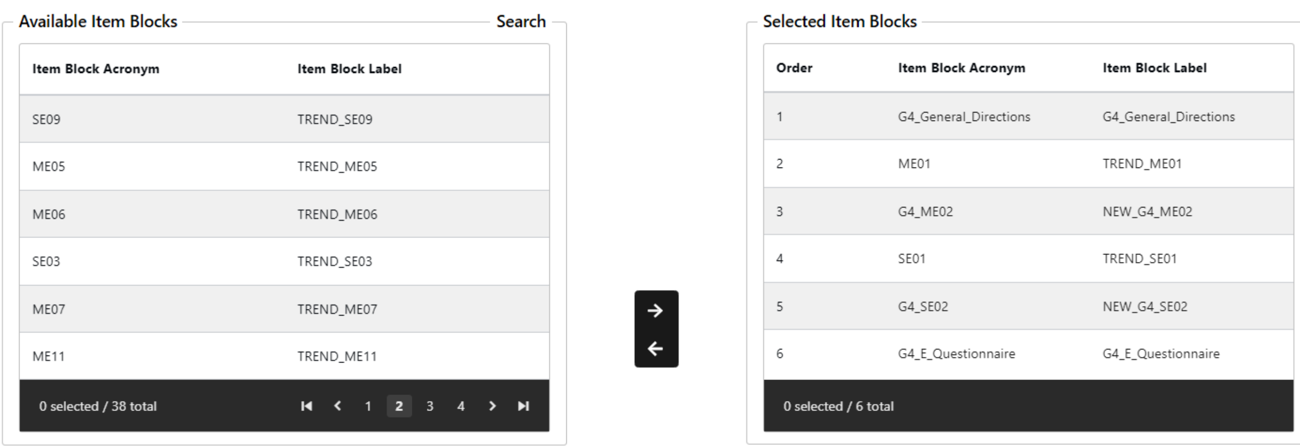
To accommodate items that included images with overlaid text, which are very common in TIMSS, an SVG feature was introduced. This provided great versatility in working with images such as line or bar charts that had overlaid text (e.g., axes labels) that later had to be translated.

It should be noted that the extended Problem Solving and Inquiry Tasks (PSIs) were substantially more complex and interactive than the standard eTIMSS items, and so were constructed independently of the Designer and subsequently combined into item block combinations (or eAssessment "booklets") by the Assembler.

Assembler

The Assembler module was used by IEA Hamburg and the TIMSS & PIRLS International Study Center to combine assessment items into blocks, and then item blocks or PSI tasks into item block combinations along with the assessment directions and eTIMSS questionnaire, in accordance with the TIMSS 2019 matrix-sampling booklet design (see TIMSS 2019 Assessment Design). It was also within this module that the allocation of the booklet number to the instrument was made. Exhibit 4.2 shows part of the Assembler window, and illustrates how a completed booklet has been constructed with six elements; on the left are available blocks that were not included in the construction of this booklet.

Exhibit 4.2: Combining Item Blocks and Directions in an Assembler Window



Once items were assembled into item blocks they could be released country-by-country into the Translation System, where countries and benchmarking participants could begin translation. Item blocks could be “released” for translation one by one as they were completed in the Designer. However, all items in the block had to be complete before the block could be released. If any item in the block was still in the state “In Progress,” the release was prevented to ensure that only the approved material would be presented to those using the Translation System. A consequence of this “whole block” approach was that if a minor change had to be made to an item after the block had been released, the entire block had to be withdrawn from the released state until such time as the change had been made and the item could be reassigned to the block.

Translation System

The Translation System was a critical part of the eAssessment system and was used by multiple parties. These included IEA Amsterdam and IEA Hamburg during the setup, NRCs and translators from the participating countries, translation verifiers employed by IEA Amsterdam, and layout verifiers from the TIMSS & PIRLS International Study Center. Although the primary purpose of the system was to enable translation, verification, and documentation of any deviations from the original international English source version, the system also enforced the appropriate workflow by a combination of user rights applicable at different stages of the process leading to the final, approved, translations.

Exhibit 4.3 shows an example of the translation window, where the eTIMSS Questionnaire title has been translated into German. Note that three fields are displayed—the original English source text in the top field, the current translation in the bottom field (only this field is directly editable), and a “track changes” field to show the changes in a color-coded manner. Additional fields could be shown, for example, to compare a current translation with an earlier version submitted for translation verification.

Exhibit 4.3: Example Translation from the Translation System

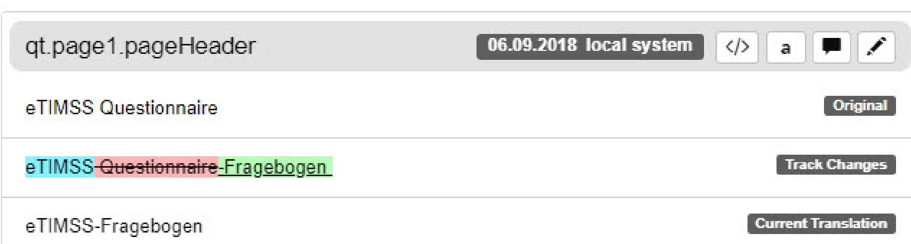
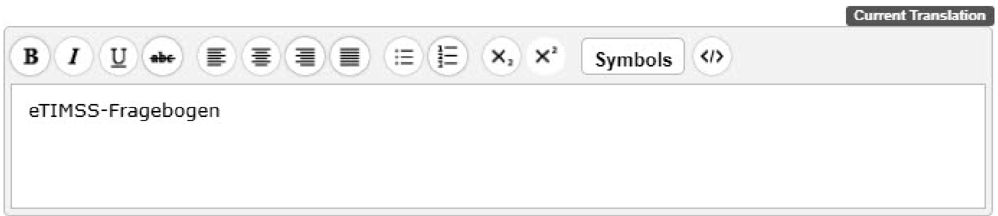


Exhibit 4.4 shows the translation editor, which enabled making a range of layout and font changes as well as inserting HTML commands, symbols, or mathematical structures such as fractions.

Exhibit 4.4: Translation System Editor



Translators also had the option to export the international source text elements in an XLIFF format, which could be used in standard translating programs for increased efficiency. This required the translations to be imported back into the Translation System and formatted for translation verification and layout verification.

The guiding principle in designing the Translation System was to define a process similar to that used in translating and verifying paperTIMSS assessment instruments. The process began by releasing a copy of the international English version of the achievement items into a separate language-specific folder for each country, followed by a workflow consisting of a series of status indicators indicating the progress of the translation and verification that were set by the various parties involved in the process. These status indicators were defined by IEA Hamburg and the TIMSS & PIRLS International Study Center as follows:

- **In Translation & Adaptation**: The initial, default status following release to the country
- **Ready for Translation Verification**: After completing translations, all blocks had to be set to this status when the translated materials were ready for translation verification
- **In Translation Verification**: Set by IEA Amsterdam when translation verification began, which locked the system for editing during the process
- **In NRC Translation Approval**: On completion of translation verification, the system was unlocked to allow the NRC to apply edits based on feedback from the translation verifier
- **Ready for Layout Verification**: After translation verification was complete and all edits applied, all blocks were set to this status to submit materials for layout verification
- **In Layout Verification**: Set by the TIMSS & PIRLS International Study Center when layout verification began, which locked the system for editing during the process
- **In NRC Layout Approval**: On completion of layout verification, the system was unlocked to allow further editing by the NRC based on feedback from the TIMSS & PIRLS International Study Center

- **Instrument Finalized**: The final status, assigned by the NRC, indicated that the materials had completed all verification steps and were ready for assessment Player production.

At each step along the way, comments could be left to document the process. The idea was to mimic the workflow and fields used in the National Adaptation Forms for the paper version. In the particular case of translation verification, additional labels were available to indicate specific errors or deviations found/corrected, including a "severity code" assigned by verifiers to each deviation to assist the NRC in deciding whether to accept or reject suggestions made by the verifier (see Instrument Translation and Layout Verification). In general, comments were labeled so as to indicate for whom they were intended. For example, a layout verifier could leave comments in the system for the attention of the NRC, in which case the label would have been "Layout."

A preview feature was available for all users to display items exactly as they would appear within the final Player. This was especially useful during layout verification, allowing as it did comparison with a preview of the original (untranslated) source version.

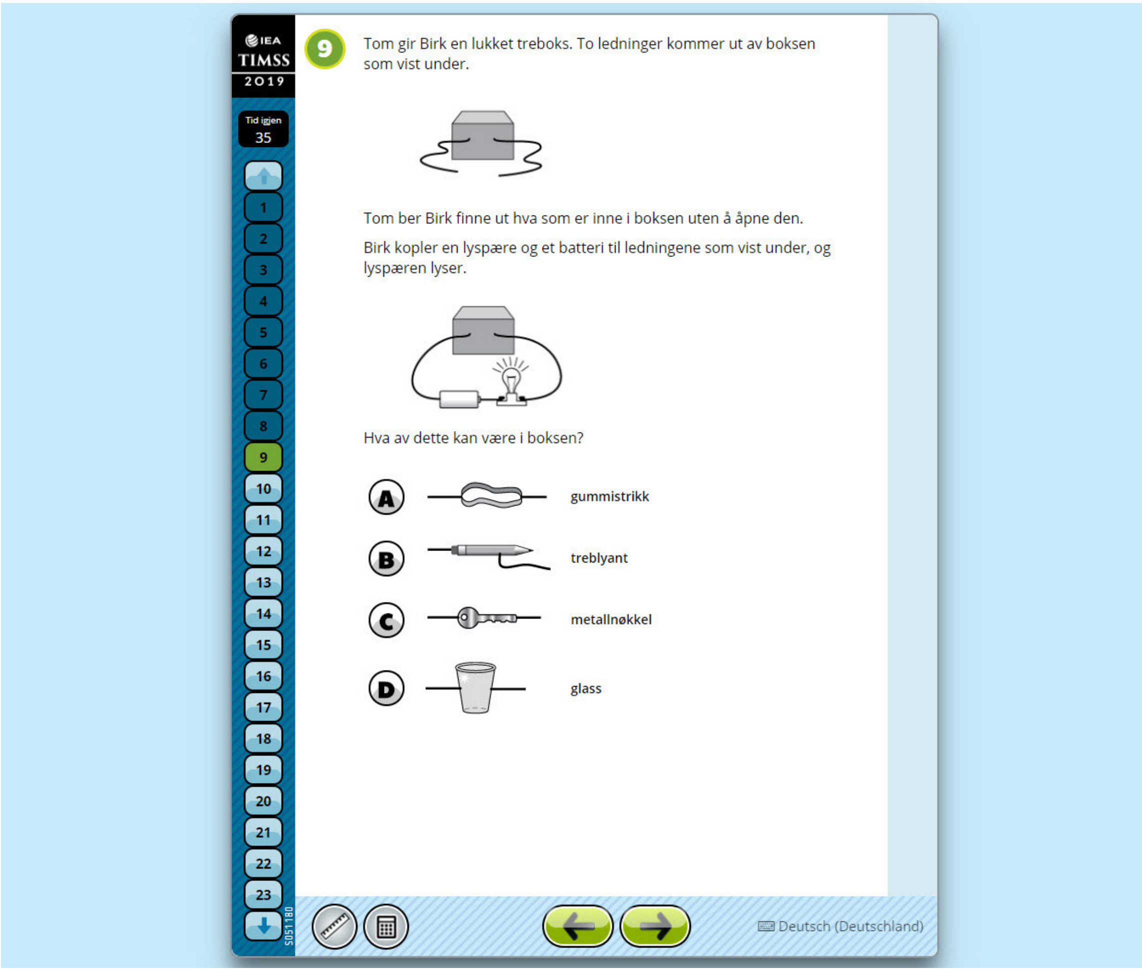
Player

The assessment Player is the software that the student interacts with while taking the eTIMSS assessment. The Player presents the assessment items to the student and uploads the student response data to IEA's data servers. After translation and layout verification were successfully completed, a customized version of the Player was produced for each language of instruction in each country. This sometimes required last-minute adjustments to the layout by IEA Hamburg before supplying the Player to the countries.

To access the Player, the student or the test administrator entered the login credentials assigned to that student. These consisted of a unique ID number and password which incorporated a two-digit code that determined the specific assessment item block combination assigned to each student. After entering the correct login credentials, the test administrator read aloud a test administration script that instructed students to enter a four-digit code to begin the test directions introducing students to the various types of items. After working through the directions, students were instructed to enter another four-digit code to begin working on the first part of the assessment. Following a short break, a third four-digit code provided access to the second part of the assessment. Finally, a fourth four-digit code allowed access to the short eTIMSS questionnaire.

Exhibit 4.5 shows the Player user interface, with an example science item. The students navigated through the assessment using the green forward/backward arrow buttons or via the navigation bar on the left side of the screen. The navigation bar records the students' progress through the assessment, showing which items have been completed and which have been omitted or not yet attempted. There also is a timer showing the remaining time.

Exhibit 4.5: eTIMSS Assessment Player User Interface



The software development criteria for the Player encompassed multiple, sometimes competing elements. The key elements were speed of operation, security, and consistency of user experience.

Speed of Operation

Several design factors were involved in producing a Player that responded smoothly and produced the minimum delay when navigating between items:

- The content (directions and items) to be displayed to the student was preloaded directly after the login screen, so that only response storage processes took place during the test session
- The Player database was mirrored in memory for faster performance
- The format of the data saved was as parsimonious as possible to maximize performance when saving and uploading data.

Security

To ensure the security of the student data and test items:

- The Player was configured to run within a ”sandbox”—a virtual space in which software can be run securely–to isolate the Player in a restricted memory range
- Contents of the sandbox were automatically deleted after the testing session.

Consistency of User Experience

To ensure the Player operated the same way for all countries and languages:

- The Player exhibited a close to identical display on Firefox or Chrome browsers or with either the Android or USB Players
- Country- and language-specific CSS files were available to make final layout and font adjustments
- Right-to-left languages had automatically reversed layout, with the ability to revert individual elements back to left-to-right format.

The Player software consisted of an executable file and two or three additional files: 1) a country-specific “Player Model” SQLite database containing the translations and the item block combination structure; 2) a template SQLite database file as the basis of the results database; and, optionally, 3) a CSS file in case layout or font changes were required.

The results database was created for each student at login time, and included information about the particular culture (country/language combination), as well as the student ID and a reference to the country in the name of the database file itself. It should be noted that the results database is the repository of not only the students’ responses to the items but also the timestamped events that reflect the process of working on the assessment, such as navigating between screens, using interface tools, and changing responses to items.

To upload data from the Player to the IEA servers, a menu option in the Player opened a separate upload page. A list of all the results databases in the default location (the same folder level as the Player executable) was displayed, along with buttons to “Upload Data” and ”Refresh.” For those using a Player to upload data from multiple USBs, it was possible to add additional results databases to the list. Clicking “Upload Data” triggered the upload process to start, and a color-coded bar showed the number of successful and unsuccessful uploads.

On completion of an upload, an acknowledgement was sent back to the Player client performing the upload confirming a successful (or rarely, unsuccessful) data transfer. Databases successfully uploaded were moved to an “uploads” subfolder and flagged to ensure that they would not be uploaded again. Databases not successfully uploaded remained in the list for a further attempt to be made.

Data Monitor

The Data Monitor was provided to enable NRCs and test administrators to further check the status of uploaded data and to monitor overall progress during the data collection. As shown in Exhibit 4.6, the Data Monitor enabled all records for a specific grade and country to be viewed, including information regarding the student ID, the record creation time (the time the student logged into eTIMSS Player), and the time of uploading.

Exhibit 4.6: Example Information from the Data Monitor

Country	
Test Country ZZA 2019-06-11T10:48:31	3
Test Country ZZB 2019-04-04T05:00:35	864
Test Country ZZC 2019-03-29T08:07:55	8
Test Country ZZD 2019-04-04T03:10:41	24

Culture	IDSCHOOL	IDCLASS	IDSTUD	Module	File Size	Record Created	Record Uploaded
en-ZZB	9998	999820	99982014	25	53 kB	3/27/2019 5:40:27 PM	3/27/2019 5:46:32 PM
en-ZZB	9998	999820	99982011	22	64 kB	3/27/2019 5:35:55 PM	3/27/2019 5:46:33 PM
en-ZZB	9998	999820	99982006	17	70 kB	3/27/2019 5:30:51 PM	3/27/2019 5:46:32 PM
en-ZZB	9998	999820	99982005	16	72 kB	3/27/2019 5:27:26 PM	3/27/2019 5:46:33 PM
en-ZZB	9998	999810	99981001	12	43 kB	3/27/2019 12:57:53 PM	3/27/2019 5:09:04 PM
en-ZZB	9998	999810	99981001	12	43 kB	3/27/2019 12:57:53 PM	3/27/2019 5:09:03 PM
en-ZZB	9998	999810	99981001	12	43 kB	3/27/2019 12:57:53 PM	3/27/2019 5:09:04 PM

Scoring System

The IEA CodingExpert software, consisting of an Administration Module and a CodingExpert Client, was the online scoring system used by NRCs and their scoring staff to score the eTIMSS constructed response items. The Administration Module enabled scoring administrators from each country and benchmarking participant to activate scorer accounts, assign scorers to items, set up and distribute training materials, distribute student answers, and monitor the progress and quality of the scoring. Scorers used the CodingExpert Client to score the student item responses assigned to them by the scoring administrator.

The Scoring System was an independent online system, working in tandem with local client software that supplied the students' responses to the scorers along with contextual information such as the translated item stem. In addition to the standard constructed response questions familiar from paper scoring, the eTIMSS Scoring System had to accommodate responses from unique, digitally-enhanced item types in the Problem Solving and Inquiry Tasks, as well as display screenshot images from the line-drawing items.

Preparing Data for Scoring and Processing

Some pre-processing steps were required to prepare data in a suitable format for import into the Scoring System and to enhance the efficiency of the human-scoring process. Data uploads from the eTIMSS Players were processed at IEA Hamburg by several data servers that received and then extracted the raw data from the uploaded SQLite databases into the "central" SQL database for all countries. This new structure contained a separate database for each country and grade, including all data from the original

SQLite databases with the addition of identifiers relating to the import of data and additional fields for scoring purposes.

Although scoring supervisors controlled the distribution of responses to scorers within countries, the responses themselves became available in the system soon after upload (with some delay due to the asynchronous handling of the import to the central database and thence to the scoring system). To avoid unnecessary scoring, therefore, it was essential that any duplicates in the central database were dealt with before import to the scoring system. In addition to measures to prevent a database from being uploaded a second time from the client side, checks were made to the results database creation date and content to ensure any possible duplicates were flagged before import. There were, however, some kinds of duplicate records that could be legitimate. Two databases with the same student ID but with different creation times could have originated in several scenarios. For example, this could be simply a case of the test administrator mistakenly using the same ID twice for two different students, or an interruption in the assessment may have led to part 1 being conducted from one USB stick and part 2 from a second. Such cases needed to be reconciled by IEA Hamburg's data processing procedures.

When scoring was completed, the student response data were transferred to tables prepared for import into the data processing system (DPE) employed at IEA Hamburg for all large-scale international assessments. Here data from the various other TIMSS sources, such as the student questionnaire or online context questionnaires, were merged together, using the IDs from the WinW3S database as the key. Following an intensive series of quality control checks to identify and reconcile any inconsistencies, the data were exported to SPSS and SAS data files for distribution to countries as part of the International Database (see Chapter 8: Creating the TIMSS 2019 International Database).

eTIMSS Assessment Delivery Methods

Countries participating in eTIMSS could choose from the following three methods for delivering the assessment:

- **USB delivery** involved running an executable file from a USB flash drive preloaded with the eTIMSS Player
- **Tablet delivery** involved running the assessment Player directly from an application on an Android-based tablet
- **Local server method** involved the use of the same Player as for the USB delivery, but run from a server on a local area network, with the client computers running a browser to connect to the server.

In practice, most countries focused on one standard method of delivery, with other options only used as exceptions due to special circumstances within schools.

USB Delivery

USB delivery involved running an executable file from a USB flash drive preloaded with the eTIMSS Player. Test administrators were instructed to run the executable, which would open the program on a main menu. On clicking the menu "Start eTIMSS" the program would present a login screen in "kiosk" full screen mode—a semi-locked-down state where some key strokes are blocked and students are unable to access or see the browser address bar.

The suitability of computers for this mode of delivery was determined by running a "system check" program, which returned a clear yes/no indication on parameters based on screen resolution, operating system, CPU speed and available memory, as well as a USB transfer rate check. This system check was provided as a stand-alone program for checking computer compatibility ahead of administration, but was also a module of the USB eTIMSS Player itself, for use on the day of testing.

Following the test session, the test administrator could use an escape code to return to the main menu in order to upload the results. It was recommended to perform the upload as soon as possible following the assessment, but it was also possible to conduct consecutive test sessions for several students and then upload these together at once. Further, it was possible to copy the results databases from several student USB flash drives to one single drive and use that to perform the upload function.

Tablet Delivery

Tablet delivery involved running the assessment Player directly from an application on an Android-based tablet. This application needed to be first installed on the tablet from an .apk (Android Application Package) file. Once installed, the application was available from the tablet home screen. The icons for these were labeled in such a way that fourth grade and eighth grade versions could be distinguished.

On tapping the appropriate application icon, a login screen would appear in full screen. In contrast to the USB version, it was not possible to prevent operation of the home button due to restrictions of the Android operating system.

A system check for tablets was provided via the Google Play Store as a separate application; there was no system check within the standard application. Minimum requirements comprised screen resolution (identical to the minimum for the USB application), version of operating system (Android 5.0.2 or higher), available storage, CPU speed, and available memory.

Following the test session, the application would return to the login screen. A button at the bottom of the screen allowed the upload of the assessment data to take place given the entry of a four-digit password. If multiple students had taken the eTIMSS assessment since the last upload, data for all these students would be transmitted.

Local Server Method

The local server method was a feature of the USB Player. With the Player program stored on a local drive of a PC, it was possible to start the program as with the standard USB from the main menu. An option then enabled the test administrator to set up the PC as a server and enable computers connected to the local network to connect to this, displaying the assessments in a browser (the Chrome browser was the preferred option).

The minimum specifications for the server computer were above those for standard USB delivery, and no system check was available to test suitability. Therefore the following minimum requirements for the server PCs were defined in order to determine if a machine was able to run the Player successfully:

- OS: Windows 8 or higher
- Processor speed: 2.2 GHz
- Memory: 8GB
- Available storage space: 10GB on SSD drive
- Administrator rights.

The upload procedure was similar to the USB method, with the additional step of stopping the server-client service. Once this was done, the results from all students could be uploaded at once.

Description of eAssessment Data

The assessment Player recorded student item responses as well as other actions taken by the student and the data were stored in a SQLite database. Student actions were broken down into timestamped events that recorded process data such as navigation behavior and tool use, but also messages to the student that were created by the system (e.g. time remaining towards the end of the test). The student responses and event data were stored separately, with the item responses in a "response table" and the events in an "event table." There were also auxiliary tables containing the student ID together with the language in which the assessment was administered and information about whether the data had already been uploaded to the IEA server. Other tables were used for error handling.

Each item response or event was stored with both general attributes and attributes specific to that response or event. The following general attributes were recorded:

- Two timestamp parts: The first recorded events and item responses in Unix time and gave the elapsed time in seconds since January 1, 1970. Since a more precise time information was needed for event data, the second timestamp added the milliseconds.
- A sequential number recording the correct sequence of actions: This number reflected the exact order of events and responses and had to coincide with the sequence obtained using the timestamp information.

- A screen ID number: This number indicated the specific screen (or item) on which the response was saved or event occurred.
- A page identifier: Due to the rotation of item blocks within booklets, an item could be displayed in different positions in the assessment. Therefore it was necessary to also include a "page number" as a general attribute.
- An item ID number: For recording responses, the item identifier referred to the particular item or item input (e.g., keyboard field) on the screen. This number corresponded to a given "raw variable name" specified by the TIMSS & PIRLS International Study Center.
- An event-type ID number: For recording events, using ID numbers instead of names helped to minimize data traffic during the assessment administration. A separate reference look-up table held the actual event names that corresponded to the event-type ID numbers.
- A response ID number: For recording responses, this identifier indicated if a response was changed later during the response process. It showed the sequential number (ID) under which the subsequent answer was saved. The final answer the student gave to an item was marked with a "NULL" value for this field.

Item Responses

In the response table, each response was stored in a separate record. The response table held the entire response history of each item the student worked on. All item responses were stored as one or more records with string of characters indicating the student response. This could be a single number, but also an extended string containing information about drawn lines or the dragging and dropping of objects. In addition, the student response table contained typed student responses that were later transferred to the Scoring System for human scoring, along with screenshot images of responses from the line-drawing tool. Responses that did not need human scoring were machine scored. For these responses, a set of detailed scoring rules provided by the TIMSS & PIRLS International Study Center were incorporated in a scoring algorithm and applied to each response to determine the appropriate score.

Event-Specific Attributes

In addition to the general attributes, attributes specific to each event were stored as JSON objects. JSON objects in general hold for each attribute the name of the attribute (property) and the value of the property. Exhibit 4.7 shows an example extract from the event table for the "UI:IsLoaded" event type. This event indicates that the appropriate test form was loaded with the first item presented to the student. The event-specific attribute is the "index" which is set to zero for the first page of the test, stores as the JSON object {"index":0}.

Exhibit 4.7: Extract from the Event Table for Event Type "UI:IsLoaded"

Event-Type Id	Screen ID	Page Identifier	Information
26	13617	0	{"index":0}

Results, Challenges, and Lessons Learned

In retrospect, it was the right decision to set up the eTIMSS system modularly and to differentiate between the phases of content and item creation, translation, instrument assembly, assessment delivery, monitoring of the data retrieval, and scoring of the responses. In each phase, different roles with the corresponding rights were required. Administration was comparatively easy due to the modular structure.

For the translation and translation verification, it turned out to be very helpful that the eTIMSS system supported the XLIFF format. With the help of XLIFF exports, translators could easily import the texts to be translated into standard translation programs and thus carry out the translations very efficiently.

The preview function, which made it possible to display the translated content as it is displayed in the specific assessment situation, was of great help. In this way, it was possible to react very early if the space allotted for the translation was not sufficient and translations were not displayed at all or incorrectly. In these situations, often manual intervention was necessary through CSS files.

In particular, the right to left (RtL) languages (Arabic, Hebrew) presented multiple challenges. A lot of effort went into producing a standard RtL template that could be applied on request in the Translation System. This template had for example certain images flipped or moved to fit to the style. Despite this template, a lot of manual work needed to be done at IEA Hamburg for adjustments of texts, images and input boxes by way of CSS files to finalize players.

The large number of Players that had to be produced in a very short possible time posed a particular challenge. In total, more than 100 player variants were created, all of which had to be tested before distribution. This work was all managed conforming to the timelines for producing paperTIMSS assessment materials.

CHAPTER 5

Instrument Translation and Layout Verification for TIMSS 2019

David Ebbs
Erin Wry
Jan-Philipp Wagner
Andrea Netten

Overview

The TIMSS & PIRLS International Study Center developed the international versions of the TIMSS 2019 assessment instruments, context questionnaires, and procedural manuals in English. Then, using the international source versions, the participating countries translated the materials into their languages of instruction and adapted them to their cultural contexts as necessary. For many countries, identifying the language of instruction, referred to as the "target" language, was relatively straightforward because there is a primary language used in the education system. However, some countries use more than one language of instruction in their education systems, and in these cases, they translated the TIMSS 2019 instruments into multiple languages. These multilingual countries also translated the context questionnaires and test administration scripts for each language assessed. In addition, some countries also translated the home questionnaire into additional languages in order to make the questionnaire more accessible to parents from different backgrounds. The complete scope of the verification process for TIMSS 2019 is fully documented in a subsequent section of this chapter.

As an additional complication for TIMSS 2019, it was the first cycle of TIMSS' two-cycle transition to digital assessment. About half the countries administered TIMSS as a digitally-based assessment (eTIMSS) and the rest as a paper-based assessment (paperTIMSS). To maintain international comparability, the TIMSS items were designed to be as identical as possible between eTIMSS and paperTIMSS. For the eTIMSS achievement materials, the procedures for translation and verification took place in the eTIMSS Online Translation system, part of IEA Hamburg's eAssessment system (see Chapter 4). The translation system was designed to mimic the same overarching procedures of paperTIMSS but also contained additional features for accommodating eTIMSS. These features included a "player preview" mode that displayed how each item would appear in the player software, an SVG editor to edit images and/or labels on images, a button to duplicate translations that appear in more than one item, and a feature to add comments or document national adaptations.

In addition, countries who participated in eTIMSS administered paper booklets of their trend items from TIMSS 2015 to a subsample of schools, to provide a "bridge" between the two administration modes. Substantial effort was required to maintain consistent procedures for verifying the three types of TIMSS 2019 instruments—eTIMSS, paperTIMSS, and bridge booklets.

To ensure a fair basis for comparing mathematics and science achievement across countries, languages, and contexts; the participating countries followed standardized internationally agreed-upon procedures to translate and prepare their national instruments for data collection (see Chapter 6: Survey Operations Procedures). This process included two stages: translation verification and layout verification. As part of the translation verification process, each country's national instruments underwent formal external review by linguistic and assessment experts. During translation verification, verifiers compared the national text to the international text and provided detailed feedback to improve the accuracy and comparability of the national translations. Once the verification was completed, the National Research Coordinators (NRCs) reviewed the feedback, revised their national materials as needed, and documented their changes. Following translation verification, countries submitted their national instruments to the TIMSS & PIRLS International Study Center for layout verification. During layout verification, verifiers checked to ensure that all national instruments conformed to the international format and that any national adaptations made to the TIMSS 2019 instruments did not unduly influence their international comparability.

The process of translation verification and layout verification was carried out once for the field test materials and a second time prior to data collection. Before data collection, the process involved verifying any changes made to field test materials and checking the trend materials to make document any changes. In the interest of measuring trends in student achievement over time, the overall process of instrument translation and verification remains consistent from one TIMSS cycle to the next.

The following TIMSS 2019 instruments underwent verification:

- Student achievement items and directions
- Context questionnaires, covers, and directions for the student, home, teacher, and school questionnaires
- Online questionnaire items, covers, and directions (for countries administering questionnaires to parents, teachers, and/or schools online)
- Paper bridge booklets (for eTIMSS countries).

Providing the Instruments to the Countries for Translation and Adaptation

The TIMSS & PIRLS International Study Center provided NRCs with the TIMSS 2019 assessment materials based on each country's mode of administration. For countries that administered paperTIMSS, NRCs received electronic files consisting of the paperTIMSS achievement materials, guidelines for adaptation, and National Adaptation Forms for documenting each step of the adaptation, translation, and verification processes. For countries that administered eTIMSS, the NRCs received digital versions of all achievement items via the eTIMSS Online Translation System, as well as PDF versions of the eTIMSS achievement blocks, guidelines for adaptation, and instructions and tutorial videos on using the eTIMSS Online Translation System. Additionally, trend countries participating in eTIMSS also received electronic files consisting of the TIMSS 2019 "bridge booklet" production files and instructions on applying their trend translations to their national bridge booklets.

As part of the TIMSS assessment design, each "block" of assessment items appeared in two achievement booklets or two eTIMSS "item block combinations" at each grade level. Therefore, the component parts of the booklets/item block combinations (item blocks and directions) were prepared as separate files for translation and translation verification. This approach allowed countries to translate each component only once. Following translation verification, countries were required to assemble their national paperTIMSS or "bridge" blocks, covers, and directions into booklets to be reviewed during layout verification. To assist in this process, the TIMSS & PIRLS International Study Center provided NRCs with detailed manuals and instructional videos, support materials for right-to-left languages, and instructions for booklet assembly. For eTIMSS, digital item block combinations were assembled through IEA's eAssessment System.

In addition to the achievement materials, all countries also received electronic files consisting of the international versions of the context questionnaires, guidelines for context questionnaire adaptation, and National Adaptation Forms for documenting the translation, adaptation, and verification processes for the questionnaires. For countries that chose to administer the home, teacher, or school questionnaires online, IEA Hamburg provided access and instructions for using the Online SurveySystem (OSS) to create, administer, and monitor online versions of the questionnaires.

Guidelines for Translation and Adaptation

The TIMSS & PIRLS International Study Center provided guidelines for translating and adapting the TIMSS 2019 instruments. The purpose of the guidelines was to ensure that, when countries translated and adapted the international versions, the meaning and difficulty level of the instruments remained the same. All participating countries were expected to follow these guidelines, including countries that administered the TIMSS 2019 instruments in English or used the Arabic source versions.

In accordance with the guidelines, translators and reviewers ensured that:

- The translated texts had the same register (language level and degree of formality) as the source texts
- The translated texts had correct grammar and usage (e.g. subject/verb agreement, prepositions, verb tenses, etc.)
- The translated texts did not remove text from the source text and did not clarify or add more information
- The translated texts had equivalent qualifiers and modifiers appropriate for the target language
- Idiomatic expressions were translated appropriately, not necessarily word for word
- Spelling, punctuation, and capitalization in the target texts were appropriate for the target language and the country's national context.

After the field test, the TIMSS & PIRLS International Study Center provided NRCs with a list of changes made to the international versions that they could refer to while preparing their assessment instruments for the main data collection. This information helped minimize the translation burden by highlighting the necessary changes to the translations before data collection.

TIMSS 2019 Arabic International Reference Version

As has been the practice since 2007, Arabic reference versions of the TIMSS 2019 instruments were made available to participating Arabic-speaking countries to serve as a starting point for preparing their national instruments. The Arabic reference versions were first created for the field test, and then updated by the same team of experts for the main data collection. This was done both for paper and digital versions of the assessment.

In TIMSS 2019, Arabic reference versions were offered for the following materials:

- Grade 4 achievement instruments
- Grade 4 less difficult mathematics achievement booklets
- Grade 8 achievement instruments
- Grade 4 context questionnaires for students, parents, teachers, and schools
- Grade 8 context questionnaires for students, teachers, and schools

The initial translation of the TIMSS 2019 instruments into Arabic was conducted in accordance with the general guidelines for translation and adaptation. The translation was produced by a team of linguists (two expert translators, one reconciler, and one proofreader) from BranTra, an independent translation agency based in Brussels, Belgium. The translators produced two separate translations that were reviewed and compared against one another. In the case of differences between the two translations, the reconciler selected the most appropriate translation for use in the field test instruments. The resulting draft versions then underwent a second review by experienced NRCs to assess the content and terminology used in specific school subjects at the target grades in a variety of Arabic-speaking countries. Upon completion of the content review, the recommendations were taken into consideration and the translations were revised accordingly. The final translations were then sent to the TIMSS & PIRLS International Study Center to produce the right-to-left Arabic reference materials.

The TIMSS & PIRLS International Study Center used the Middle Eastern Version of Adobe® InDesign® software to create the paper Arabic-reference production files with CopyFlow Gold® to import the translation from rich-text format (RTF) into InDesign. After importing the translations, the TIMSS & PIRLS International Study Center applied fonts, styles, and graphics to the instruments and reviewed the materials to ensure that the translations and layout resembled the international version aside from the right-to-left format. Before the release of the TIMSS 2019 Arabic paper reference versions, an additional optical check was performed to verify the layout of the Arabic version and eradicate any remaining errors or issues that occurred during the import process. The multiple stages of translation and review of the Arabic reference instruments ensured that they were an adequate starting point for Arabic-speaking countries to use in preparing their national versions.

For eTIMSS, the Arabic reference translations were imported into the eTIMSS Online Translation System for Arabic-speaking countries that requested to start with this source version. All graphics were automatically flipped and countries were given instructions on how to revert this if they required certain graphics to be viewed left-to-right in their national education context. Further assistance was provided to the eTIMSS Arabic-speaking countries during layout verification for any right-to-left issues that NRCs were not able to adjust themselves.

Blocks of Achievement Items Designated to Measure Trends

According to the TIMSS design, about two-thirds of the items are carried over from one cycle to the next for the purpose of measuring changes in student achievement over time. Therefore, TIMSS 2019 included some items previously used in TIMSS 2015 and 2011. To ensure the quality of measuring TIMSS trends, the trend items must identical from cycle to cycle. For countries that previously participated in TIMSS 2011 or TIMSS 2015, the TIMSS 2019 trend blocks (including paper bridge booklets) were reviewed during translation and layout verification in comparison with those from the last cycle in which the country participated. If a country determined that changes to an item in a trend block were absolutely necessary (e.g., in order to correct a mistranslation discovered in a previous version), they were instructed to document the change for further review during the verification process. Trend items that underwent changes were not included in the scaling process or the estimation of the achievement scores for that country.

National Adaptations Forms

Each country prepared one National Adaptations Form (NAF) for each set of paperTIMSS achievement instruments and/or set of questionnaires in each language in which they were administered. NAFs are Excel documents formatted to contain the translations, adaptations, and verification history of each set of national instruments administered on paper. When countries translated and adapted their national paper instruments, the NAFs were filled out by the translators, reviewers, and NRCs. Documenting an adaptation in the NAF requires entering the identifying information (location and/or question number), an English back translation of the adaptation, and recoding instructions (if applicable). During verification, the verifiers reviewed the documentation in the NAFs and recorded any feedback. NRCs were responsible for updating the documentation within the NAFs after each round of international verification. To ease the process of documentation and review, the NAFs include designated areas for each stage of instrument preparation and verification.

For eTIMSS, NAFs were not external worksheets but instead built into the eTIMSS Online Translation System. All national adaptations and documentation for the eTIMSS instruments, as well as feedback from the verifiers was recorded directly into the eTIMSS Online Translation System. For archiving purposes, the translation system had a function to export all documentation including translations, adaptations, and comments from the translators, verifiers, and NRCs.

Countries administering eTIMSS were also provided with Bridge Verification Forms for the paper bridge booklets. Because the bridge booklets were comprised of each country's trend blocks and did not contain any new translations or adaptations, the Bridge Verification Forms were a simplified version of the NAF. These forms did not need to be filled out by NRCs but, rather, were used by the verifiers to document any deviations from trend and any layout issues noted during verification.

Scope of Translation and Layout Verification in TIMSS 2019

For many countries, identifying the language of assessment, referred to as the "target" language, was relatively straightforward because there is a primary language used in the education system. However, some countries use more than one language of instruction in their education systems, and in these cases, they translated the TIMSS 2019 instruments into multiple languages. These multilingual countries also translated the context questionnaires and test administration scripts for each language assessed. In addition, some countries also translated the home questionnaire into additional languages in order to make the questionnaire more accessible to parents from different backgrounds.

For TIMSS 2019, 64 countries and 8 benchmarking participants prepared a total of 144 sets of achievement instruments and 145 sets of background questionnaires in 50 languages.[1] The instruments

1 Counts may be inconsistent with Exhibits 5.1, 5.2, and 5.3 due to omission of benchmarking entities that share instruments with the national country participant and did not require additional translation and layout verification.

were translated into 50 different languages across 58 participating countries and 6 benchmarking entities at the fourth grade, and across 39 countries and 7 benchmarking entities at the eighth grade. Of these participants, 31 countries and 4 benchmarking entities administered the TIMSS 2019 instruments in more than one language. The most common languages used were English (24 countries) and Arabic (10 countries).

Exhibits 5.1, 5.2, and 5.3 lists the target languages used for the TIMSS 2019 fourth grade assessment, the fourth grade less difficult mathematics assessment, and the eighth grade assessment, respectively.

Exhibit 5.1: Languages Used for the TIMSS 2019 Grade 4 Assessment Instruments

Country	Language	Instruments				
		Achievement Test	Student Questionnaire	Teacher Questionnaire	School Questionnaire	Home Questionnaire
Armenia	Armenian	●	●	●	●	●
Australia	English	●	●	●	●	
Austria	German	●	●	●	●	●
Azerbaijan	Azeri	●	●	●	●	●
	Russian	●	●	●	●	●
Bahrain	English	●	●	●	●	●
	Arabic	●	●	●	●	●
Belgium (Flemish)	Dutch	●	●	●	●	●
Bulgaria	Bulgarian	●	●	●	●	●
Canada	English	●	●	●	●	●
	French	●	●	●	●	●
Chile	Spanish	●	●	●	●	●
Chinese Taipei	Traditional Chinese	●	●	●	●	●
Croatia	Croatian	●	●	●	●	●
	Italian		●			●
	Serbian		●			●

Exhibit 5.1: Languages Used for the TIMSS 2019 Grade 4 Assessment Instruments (continued)

Country	Language	Instruments				
		Achievement Test	Student Questionnaire	Teacher Questionnaire	School Questionnaire	Home Questionnaire
Cyprus	Greek	●	●	●	●	●
	English	●	●	●	●	●
Czech Republic	Czech	●	●	●	●	●
Denmark	Danish	●	●	●	●	●
England	English	●	●	●	●	
Finland	Finnish	●	●	●	●	●
	Swedish	●	●	●	●	●
France	French	●	●	●	●	●
Georgia	Georgian	●	●	●	●	●
Germany	German	●	●	●	●	●
Hong Kong SAR	English	●	●	●	●	●
	Traditional Chinese	●	●	●	●	●
Hungary	Hungarian	●	●	●	●	●
Iran, Islamic Rep. of	Farsi	●	●	●	●	●
Ireland	English	●	●	●	●	●
	Irish	●	●	●	●	●
Italy	Italian	●	●	●	●	●
Japan	Japanese	●	●	●	●	●
Kazakhstan	Kazakh	●	●	●	●	●
	Russian	●	●	●	●	●
Korea, Rep. of	Korean	●	●	●	●	●

Exhibit 5.1: Languages Used for the TIMSS 2019 Grade 4 Assessment Instruments (continued)

Country	Language	Instruments				
		Achievement Test	Student Questionnaire	Teacher Questionnaire	School Questionnaire	Home Questionnaire
Latvia	Latvian	●	●	●	●	●
	Russian	●	●			●
Lithuania	Lithuanian	●	●	●	●	●
	Polish	●	●			
	Russian	●	●			
Malta	Maltese					●
	English	●	●	●	●	●
Netherlands	Dutch	●	●	●	●	
New Zealand	English	●	●	●	●	●
Northern Ireland	English	●	●	●	●	●
Norway (5)	Bokmål	●	●	●	●	●
	Nynorsk	●	●			
Oman	Arabic	●	●	●	●	●
	English	●	●	●	●	●
Poland	Polish	●	●	●	●	●
Portugal	Portuguese	●	●	●	●	●
Qatar	Arabic	●	●	●	●	●
	English	●	●	●	●	●
Russian Federation	Russian	●	●	●	●	●
Serbia	Serbian	●	●	●	●	●

Exhibit 5.1: Languages Used for the TIMSS 2019 Grade 4 Assessment Instruments (continued)

Country	Language	Instruments				
		Achievement Test	Student Questionnaire	Teacher Questionnaire	School Questionnaire	Home Questionnaire
Singapore	English	●	●	●	●	●
	Traditional Chinese					●
	Tamil					●
	Malay					●
Slovak Republic	Slovak	●	●	●	●	●
	Hungarian	●	●			●
Spain	Spanish	●	●	●	●	●
	Catalan	●	●	●	●	●
	Valencian	●	●	●	●	●
	Galician		●	●	●	●
	Basque	●	●	●	●	●
Sweden	Swedish	●	●	●	●	●
Turkey (5)	Turkish	●	●	●	●	●
United Arab Emirates	Arabic	●	●	●	●	●
	English	●	●	●	●	●
United States	English	●	●	●	●	

Exhibit 5.2: Languages Used for the TIMSS 2019 Grade 4 Less Difficult Mathematics Assessment Instruments

Country	Language	Instruments				
		Achievement Test	Student Questionnaire	Teacher Questionnaire	School Questionnaire	Home Questionnaire
Albania	Albanian	●	●	●	●	●
Bosnia and Herzegovina	Bosnian	●	●	●	●	●
	Croatian	●	●	●	●	●
	Serbian	●	●	●	●	●
Kosovo	Albanian	●	●	●	●	●
Kuwait	Arabic	●	●	●	●	●
	English	●	●	●	●	●
Montenegro	Montenegrin (Cyrillic)	●	●	●	●	●
Morocco	Arabic	●	●	●	●	●
North Macedonia	Macedonian	●	●	●	●	●
	Albanian	●	●	●	●	●
Pakistan	Urdu	●	●	●	●	●
	English	●	●	●	●	●
	Sindhi	●	●	●	●	●
Philippines	English	●	●	●	●	●
Saudi Arabia	Arabic	●	●	●	●	●
	English	●	●	●	●	●
South Africa (5)	Afrikaans	●	●	●	●	●
	English	●	●	●	●	●

Exhibit 5.3: Languages Used for the TIMSS 2019 Grade 8 Assessment Instruments

Country	Language	Instruments			
		Achievement Test	Student Questionnaire	Teacher Questionnaires	School Questionnaire
Australia	English	●	●	●	●
Bahrain	English	●	●	●	●
	Arabic	●	●	●	●
Canada[1]	English	●	●	●	●
	French	●	●	●	●
Chile	Spanish	●	●	●	●
Chinese Taipei	Traditional Chinese	●	●	●	●
Cyprus	Greek	●	●	●	●
	English	●	●	●	●
Egypt	Arabic	●	●	●	●
	English	●			
England	English	●	●	●	●
Finland	Finnish	●	●	●	●
	Swedish	●	●	●	●
France	French	●	●	●	●
Georgia	Georgian	●	●	●	●
	English	●	●	●	●
Hong Kong SAR	Traditional Chinese	●	●	●	●
Hungary	Hungarian	●	●	●	●
Iran, Islamic Rep. of	Farsi	●	●	●	●
Ireland	English	●	●	●	●
	Irish	●	●	●	●

1 Canada only participated at Grade 8 in the benchmarking regions of Ontario and Quebec.

Exhibit 5.3: Languages Used for the TIMSS 2019 Grade 8 Assessment Instruments (continued)

Country	Language	Instruments			
		Achievement Test	Student Questionnaire	Teacher Questionnaires	School Questionnaire
Israel	Hebrew	●	●	●	●
	Arabic	●	●	●	●
Italy	Italian	●	●	●	●
Japan	Japanese	●	●	●	●
Jordan[2]	Arabic	●	●	●	●
	English[2]	●			
Kazakhstan	Kazakh	●	●	●	●
	Russian	●	●	●	●
Korea, Rep. of	Korean	●	●	●	●
Kuwait	Arabic	●	●	●	●
	English	●	●	●	●
Lebanon	English	●	●	●	●
	French	●	●	●	●
Lithuania	Lithuanian	●	●	●	●
	Polish	●	●		
	Russian	●	●		
Malaysia	Malay	●	●	●	●
	English	●			
Morocco	Arabic	●	●	●	●
	French	●			
New Zealand	English	●	●	●	●

2 For Jordan, the Grade 8 Achievement Test in English did not undergo international adaptation/translation verification.

Exhibit 5.3: **Languages Used for the TIMSS 2019 Grade 8 Assessment Instruments** (continued)

Country	Language	Instruments			
		Achievement Test	Student Questionnaire	Teacher Questionnaires	School Questionnaire
Norway (9)	Bokmål	●	●	●	●
	Nynorsk	●	●		
Oman	Arabic	●	●	●	●
	English	●	●	●	●
Portugal	Portuguese	●	●	●	●
Qatar	Arabic	●	●	●	●
	English	●	●	●	●
Romania	Romanian	●	●	●	●
Russian Federation	Russian	●	●	●	●
Saudi Arabia	Arabic	●	●	●	●
	English	●	●	●	●
Singapore	English	●	●	●	●
South Africa (9)	English	●	●	●	●
	Afrikaans	●	●	●	●
Sweden	Swedish	●	●	●	●
Turkey	Turkish	●	●	●	●
United Arab Emirates	Arabic	●	●	●	●
	English	●	●	●	●
United States	English	●	●	●	●

Translation and Translation Verification

Translators and Reviewers

All countries and benchmarking participants were advised to hire highly qualified translators and reviewers well suited to the task of working with the TIMSS materials.

Essential qualifications for translators and reviewers included:

- Excellent knowledge of English
- Excellent knowledge of the target language
- Experience in the country's cultural context
- Experience translating texts in the subject areas related to the TIMSS assessment (mathematics and science).

The primary responsibility of the reviewer was assessing the readability and accuracy of the translation for the target population. In addition to excellent language skills and knowledge of the country's cultural context, reviewers were expected to have experience with students in the target grade (preferably as a school teacher).

In cases where several translators and reviewers were needed to distribute the work, NRCs were responsible for maintaining the consistency of the translations within and across instruments. Countries that administered the assessment in more than one language were advised to employ translators and reviewers that were highly proficient in the various languages to ensure the consistency of the translations and adaptations across different language versions.

Translation and Adaptation of the Achievement Instruments

One of the main challenges in translating TIMSS achievement blocks is finding appropriate terms and expressions in the target language(s) that convey the same meaning and style of text as the international version. When adapting and translating expressions with more contextually appropriate terms, translators ensured that the meaning and difficulty of the item remained the same as the international version. In particular, it was important that adaptation/translation did not simplify or clarify the text in such a way as to provide a hint or definition of the meaning of a question. Translators also ensured the consistency of adaptations and translations from item to item. For multiple-choice items, translators were instructed to pay particular attention to the literal and synonymous matches of text in both the question stem and answer options. Any matches in the international version were required to be maintained in the translated national version.

Although NRCs were strongly advised to keep adaptations to a minimum, some adaptations were necessary in order to prevent students from facing unfamiliar contexts or vocabulary that could hinder their ability to read and understand the item. For example, a reference to the working week as Monday to Friday might be adapted according to national customs. Similarly, a word such as "flashlight" in American English would be adapted to "torch" in British English. In TIMSS 2019, most of the adaptations were in respect to national conventions of measurement (e.g. metric vs imperial units), mathematical notation (e.g. decimal separator, multiplication sign), punctuation, and expressions of date and time. In addition, fictional names of characters and places were modified to similar names in the target language. When adapting the names of fictional cities or towns, translators were instructed not to use real names of places to prevent student responses' from being influenced by their perceptions and knowledge of the real locations.

Within the TIMSS items, some terms were not to be changed or adapted beyond translation. Examples included proper names of actual people and places, as well as the fictional currency "zed" which is used in TIMSS items to denote currency. To aid in the standardization of common adaptations across countries, the TIMSS & PIRLS International Study Center provided a list of specific examples of acceptable and unacceptable adaptations, including a list of measurement conversions.

Translation and Adaptation of the Context Questionnaires

Translation procedures for the questionnaires differed from the achievement blocks in that participating countries were required to adapt some terms to ensure that questions were appropriate for the national context and education system. The terms requiring adaptation were listed in angle brackets in the international version with a description of what country-specific information was needed. For example, <language of test> and <fourth grade> would be adapted to the actual language and grade in which the assessment is administered—in the Netherlands, these terms would be replaced by equivalents "Nederlands" (Dutch) and "groep 6" (grade 4).

The guidelines for translation and adaptation contained detailed descriptions of the required questionnaire adaptations, including the intent of each adaptation to help translators select the appropriate national term or expression to convey the intended meaning. For TIMSS 2019, the main difficulties encountered in adapting the questionnaires involved terminology, specific educational contexts, and, for a few countries, consistency across multiple languages of administration.

Countries were permitted to add a limited number of questions to the questionnaires that were of national interest. To avoid influencing responses to the international questions, NRCs were advised to place these national questions at the end of the corresponding module or questionnaire and to ensure these questions adopted the same format as the rest of the questionnaire. All national questions required approval by the TIMSS & PIRLS International Study Center before inclusion in the final questionnaires.

International Translation Verification

After the instruments were translated and adapted, they were submitted to IEA Amsterdam for translation verification. For TIMSS 2019, the international translation verifiers were responsible for reviewing and documenting the quality of the national instruments and their comparability to the international instruments.

The required qualifications for international translation verifiers were:

- Fluency in English
- Mother tongue proficiency in the target language
- Formal credentials as translators working in English
- University-level education and (if possible) familiarity with the subject area
- Residency in the target country, or close contact with the country and its culture.

IEA Amsterdam in collaboration with cApStAn Linguistic Quality Control trained the international translation verifiers and provided them with a comprehensive set of instructional materials to support their work. For TIMSS 2019, web-based seminars were used to train and provide verifiers with information about TIMSS and the assessment instruments. Each verifier received a document containing the description of the adaptation and translation guidelines, the relevant manuals and instruments, and a document with the directions and instructions for reviewing the national instruments and documenting deviations from the international version.

The Translation Verification Process

The instruction and training given to the verifiers emphasized the importance of maintaining the same meaning and difficulty level of the translations as in the international versions and ensuring that translations and adaptations were adequate and consistent within and across national instruments. The translation verification process involved:

- Checking the accuracy, linguistic correctness, and comparability of the translation and adaptations of the achievement items and questionnaires
- Documenting any deviations between the national and international versions, including additions, deletions, and mistranslations
- Suggesting an alternative translation/adaptation to improve the accuracy and comparability of the national instruments.

Verifiers provided feedback on the quality of the translated and adapted texts directly in the instruments, in the accompanying NAFs, and/or in the eTIMSS Online Translation System. Verifiers were asked to correct the text of the assessment items and questionnaires and/or to add notes specifying errors using either "Sticky Notes" in Adobe PDFs, "Track Changes" and "New Comment" functions in Microsoft Word or the "Add comment" button in the eTIMSS Translation System.

For paper-based instruments, all comments viewed by the verifiers as major issues or deviations in the adaptation/translation were entered in the NAF. For eTIMSS achievement materials all verifier comments were recorded in the eTIMSS Online Translation System. All verifier comments were accompanied by a code to help NRCs understand the severity and type of deviation of the translated text from the international version (see Exhibit 5.4). Translation verifiers were also instructed to review all adaptations and check whether or not the adaptations were correctly documented and implemented for review by the NRC and the TIMSS & PIRLS International Study Center.

Exhibit 5.4: Translation Verification Feedback Codes for TIMSS 2019

The criteria for coding are as follows:	
CODE 1 indicates a major change or error. Examples include the omission or addition of a question or answer option; incorrect translation that changes the meaning or difficulty of the item or question; and incorrect order of questions or answer options in a multiple-choice question. If in any doubt, verifiers are instructed to use **CODE 1?** so that the error can be referred to the TIMSS & PIRLS International Study Center for further consultation	**CODE 2** indicates a minor change or error, such as a spelling or grammar error that does not affect comprehension. **CODE 3** indicates that while the translation is adequate, the verifier has a suggestion for an alternative wording. **CODE 4** indicates that an adaptation is acceptable and appropriate.

Translation Verification of the Trend Assessment Blocks

For countries assessing changes of student achievement over time, the international verification procedures included a so-called trend check of the achievement instruments to ensure that the trend items had not changed. For countries administering eTIMSS, this included a check of the bridge booklets against the national trend versions.

As part of the trend check process, translation verifiers checked that each of the trend items used in the current TIMSS cycle remained identical to the trend items as they were administered in the previous cycle and documented any differences in content or wording.

The verifiers were instructed to record any discrepancies found in the trend items in the NAF, eTIMSS Online Translation System, or Bridge Verification Form. NRCs were required to carefully review all discrepancies and discuss any proposed changes with the TIMSS & PIRLS International Study Center.

Review of International Translation Verification Feedback

Upon completion of international translation verification, the NRCs were responsible for responding to the translation verifiers' feedback by either accepting, modifying or rejected suggested changes to the adapted and/or translated text. Some of the typical errors identified by the verifiers during translation verification included mistranslations, omissions/additions of text, inconsistent translations, gender agreement, and grammar. Some of the domain-specific concepts in mathematics and science were a particular challenge to translate for some languages. The constructive feedback from the verifiers aided NRCs in revising the materials and in improving the quality of their national versions in line with the translation guidelines for TIMSS 2019.

Layout Verification

Following translation verification, all national instruments were required to undergo layout verification by the TIMSS & PIRLS International Study Center. Layout verification is the final external review and ratification of each participating country's assessment instruments, questionnaires, and corresponding documentation. During layout verification, staff at the TIMSS & PIRLS International Study Center reviewed all national instruments to ensure international comparability of layout structure and proper documentation of any national adaptations.

In particular, layout verification focused on the following:

- Reviewing the national achievement materials and context questionnaires against the international versions for acceptable layout structure
- Reviewing national adaptations to the achievement materials and context questionnaires with respect to how they may influence the international comparability of the data
- Reviewing the online questionnaires against their corresponding paper versions (where applicable)
- Reviewing trend materials and bridge booklets against the previous national versions for consistency across cycles.

Layout Verification of Achievement Materials

The primary goal of layout verification of achievement materials is to ensure that students in different countries experience the assessment instruments in the same way. Thus, the national versions of the paperTIMSS 2019 achievement booklets or eTIMSS item blocks were checked against the appropriate international versions to identify any deviations from the international format. For paperTIMSS instruments, layout verification was conducted on printed versions of each set of national booklets compared to printed versions of the international booklets. For eTIMSS materials, layout verification was conducted directly in the eTIMSS Online Translation System using the player preview mode. To accommodate on-screen verification, the translation system included a feature to view both the national preview and the international preview so that verifiers could compare each national item to the international version.

Due to differences in languages, the TIMSS national assessment instruments varied slightly in length and format across countries. The international versions, however, were designed with this in mind. For paperTIMSS materials, extra space was provided in the margins of the pages to facilitate the use of longer text and different paper sizes (letter versus A4) without necessitating extensive changes to the layout of each page. For eTIMSS, the layout of the items was designed to run vertically to minimize scrolling in longer languages. In addition, specific layout adjustments were made to national eTIMSS items, as needed, to accommodate things such as special characters, longer languages, and country-specific right to left requirements.

During layout verification of paperTIMSS instruments, verifiers reviewed the national booklets against the international versions with respect to pagination, page breaks, headers, footers, stop signs, item sequence, scoring boxes, response options, text formats, and graphics. For countries administering paperTIMSS in right-to-left languages this included ensuring that no elements were incorrectly altered in adjusting the alignment and conventions for graphics were implemented consistently throughout all booklets. Any layout deviations or errors, as well as any concerns of international incomparability of assessment items, were documented by the verifiers in the NAFs. Following layout verification, the NAFs containing the verifiers' comments were sent back to the National Research Coordinators for consideration. The NRC's were asked to confirm that each suggested change was implemented or provide an explanation for not implementing the suggested change.

During layout verification of eTIMSS materials, the verifiers reviewed the layout of all items, directions, system login pages, on-screen alerts, and eTIMSS system components including navigation tools, number pad, ruler, and calculator (8th grade only). The verifiers checked the eTIMSS materials for comparability to the international versions as well as on-screen readability, minimal scrolling, item sequence, response format, text format and graphics. For countries with right-to-left languages the verifiers checked that no elements were incorrectly altered in adjusting the alignment and conventions for graphics were implemented consistently throughout all of the items. As an additional step for eTIMSS layout verification, the verifiers also checked the basic functionality of the items and eTIMSS system components. Any technical issues were reported to the IEA Hamburg software unit to be fixed prior to development of the national Player software.

For eTIMSS, the verifiers entered their comments regarding layout deviations or errors, as well as any concerns of international incomparability directly in the eTIMSS Online Translation System. Comments from the verifiers included a reference to the text element, whether the comment was related to an adaptation or layout issue, and a button to "accept" or "reject" the comment. Following the completion of layout verification, the NRC's were asked to review the verifier's feedback and accept or reject each comment. If the NRC rejected a comment they were required to provide an explanation for not implementing the suggested change.

Layout Verification of Context Questionnaires

As with the achievement booklets, the context questionnaires were checked against the international versions to identify any potential layout issues as well as to ensure the international comparability of the questionnaire data. During layout verification of questionnaires, the verifiers took into consideration any national adaptations documented by the NRCs. Instances of internationally incomparable adaptations or errors were recorded by the verifiers in the NAFs along with recommendations for recoding or rewording.

In an effort to make the questionnaires general enough for international analyses but appropriate for each intended audience, participating countries were required to adapt certain phrases and designations in the text of the questionnaires. For example, items asking about levels of education were expressed in terms of the current version of the *International Standard Classification of Education ISCED 2011* (UNESCO Institute for Statistics, 2012), and required adaptation to the nationally equivalent educational terms by each participating country. These items were reviewed during layout verification in comparison to the ISCED level classifications, and if deemed internationally comparable, suggestions were made by the verifier to revise or recode their education categories.

The verifiers ensured that all items requiring adaptations were accompanied by proper English back translations. The documentation for these universally adapted questionnaire items was intended for later use in the National Adaptations Database. The database is a compilation of each country's questionnaire adaptations, to be used during data processing by IEA Hamburg (see Chapter 8). The information included in the database is reported as a supplement to the *TIMSS 2019 User Guide for the International Database*.

For countries that chose to administer the home, teacher, or school questionnaires online using the IEA Online SurveySystem (OSS), layout verification of the online questionnaires was conducted in the OSS environment. All countries that administered online questionnaires were also required to create paper directions containing information on accessing the online questionnaire and the purpose and use of the information being collected. The paper directions were reviewed by the layout verifiers in conjunction with the online questionnaires. For countries that administered any of the questionnaires in both paper and online, the layout verifiers compared the paper version to the corresponding online version to ensure consistency across the two forms. Feedback for both online and paper questionnaires were entered into the questionnaire NAFs and sent back to the NRCs for consideration. The NRCs were asked to confirm that each suggested change was implemented or provide an explanation for not implementing the suggested change.

Layout Verification of Trend Materials and Bridge Booklets

For countries that previously participated in TIMSS 2015 or TIMSS 2011, the national TIMSS 2019 trend blocks were also reviewed against the versions from the last cycle in which the country participated. During layout verification of trend materials, the verifiers ensured that the layout structure and adaptations in the national TIMSS 2019 instruments were consistent with countries' trend versions. In the event a country needed to make a change to their trend materials due to an error in previous cycles or a change in curriculum, the TIMSS & PIRLS International Study Center documented approval of the change or requested more information in the "trend check" section of the NAF.

For eTIMSS achievement materials, the change in mode of administration from paper to digital necessitated slight changes to the layout of some trend items. During layout verification the verifiers ensured that all conventions and adaptations in the eTIMSS 2019 materials were consistent with the trend versions and any changes beyond adjustments for digital administration were properly documented.

In addition to the eTIMSS 2019 achievement items, countries participating in eTIMSS also produced paper bridge booklets for use in the TIMSS 2019 bridge study. The bridge booklets were reviewed during layout verification alongside the corresponding national trend blocks from previous cycles. The verifiers also ensured that the pagination, page breaks, block sequence, headers, footers, graphics, covers, and directions of each bridge booklet matched the international versions of the TIMSS 2019 bridge booklets.

Review of Final Instruments

Upon completion of layout verification, the NRCs were responsible for finalizing their national TIMSS 2019 instruments. This included making any necessary adjustments to the materials and responding to all the feedback from the layout verifiers. Once the materials were reviewed and finalized, NRCs were required to submit their materials to the TIMSS & PIRLS International Study Center for a final review. In the final review of paper-based instruments (paperTIMSS, context questionnaires, and bridge booklets), the layout verifiers checked to see that all issues had been addressed, comments in the NAFs had been answered, and all of the compiled booklets and questionnaires had been submitted. Once, the TIMSS & PIRLS International Study Center confirmed the materials were finalized, the country was permitted to begin printing the paper-based instruments. In the final review of eTIMSS achievement materials, the layout verifiers checked that all issues had been addressed in the eTIMSS Online Translation System, comments from verification had been answered, and all materials had been set to the status "Instrument Finalized." This status indicated that no further changes would be made to the materials and the country was now ready to receive their national Player software.

Outcomes and Summary for TIMSS 2019

TIMSS 2019 followed stringent procedures for translation, adaptation, and verification. The ultimate goal of the translation and verification process was to create national versions of the TIMSS 2019 instruments that accommodated national languages and context while maintaining international comparability. The TIMSS & PIRLS International Study Center provided countries and benchmarking entities with comprehensive guidelines and procedural manuals outlining the various steps of instrument preparation and verification.

The feedback from translation verification helped NRCs to improve the quality and comparability of their national instruments. Similarly, the feedback from the layout verification provided NRCs with explanations for the adjustments requested and helped ensure the international comparability of instruments across countries. Ultimately, the stringent procedures applied in TIMSS 2019 resulted in high quality instruments that allowed for comparisons in student achievement across all participating countries and benchmarking entities.

Reference

UNESCO. (2012). *International Standard Classification of Education ISCED 2011*. Montreal: UNESCO Institute of Statistics. Retrieved from http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf

CHAPTER 6

Survey Operations Procedures for TIMSS 2019

Ieva Johansone

Overview

As data-based indicators of countries' student achievement profiles and learning contexts, TIMSS assessments are crucially dependent on the quality of the data collected by each participating country and benchmarking entity. Whereas the development of the assessments is an intensely collaborative process involving all of the partners in the enterprise, the process of administering the assessments and collecting the data is uniquely the responsibility of each individual country or benchmarking participant.

To ensure the consistency and uniformity of approach necessary for high-quality, internationally comparable data, all participants are expected to follow a set of standardized operations procedures. These procedures have been developed through a partnership involving the TIMSS & PIRLS International Study Center, IEA Amsterdam, IEA Hamburg, Statistics Canada, and National Research Coordinators (NRCs) from participating countries. The major steps of the operations and procedures are similar from one assessment cycle to the next. However, with each assessment cycle the operations procedures are updated to enhance efficiency and accuracy and reduce burden, making use of developments in information technology to automate routine activities wherever possible.

Each new assessment cycle also brings something new and unique requiring the operations and procedures to be adapted. For example, the 2019 cycle of TIMSS began the transition to digital assessment (known as eTIMSS) with about half of the participating countries switching from the previous paper-based version (known as paperTIMSS) to the new digital format. Adapting operational procedures for this new assessment mode and integrating the workflow into the existing TIMSS operations was a significant undertaking. In order to control for any assessment mode effects, in addition to the usual nationally representative sample, countries transitioning to eTIMSS were required to administer "bridge" paper instruments to an extra, equivalent sample of students, which also required integrating operations and procedures into the overall TIMSS 2019 assessment administration.

In each country or benchmarking entity, the National Research Coordinator was responsible for the implementation of TIMSS 2019. Internationally, National Research Coordinators provided the country's perspective in all international discussions, represented the country at international meetings, and were the responsible contact persons for all project activities. Locally, National Research Coordinators were responsible for implementing all the internationally agreed-upon procedures and facilitating all of the national decisions regarding TIMSS, including any adaptations for the national context.

The daily tasks of the National Research Coordinators varied over the course of the TIMSS 2019 cycle. In the initial phases, National Research Coordinators participated in the TIMSS 2019 assessment frameworks and assessment development process (see Chapter 1). and collaborated with Statistics Canada and IEA Hamburg in developing a plan to implement the TIMSS 2019 sampling design within the country or benchmarking entity (see Chapter 3).

Following the development of the draft achievement items and context questionnaires, countries conducted a full-scale field test of all instruments and operational procedures in March through May 2018 in preparation for the TIMSS 2019 data collection, which took place in October through December 2018 in Southern Hemisphere countries, and in March through June 2019 in Northern Hemisphere countries. As well as providing crucial data to support finalization of the assessment instruments (achievement items and questionnaires), the field test enabled the National Research Coordinators and their staff to become acquainted with the operational activities. The feedback they provided was used to improve the procedures for the data collection. As expected, the field test resulted in some enhancements to survey operations procedures, especially for eTIMSS which was new for the 2019 assessment cycle and contributed to ensuring the successful execution of TIMSS 2019.

As part of ongoing efforts to improve operations, the National Research Coordinators were asked to complete a Survey Activities Questionnaire (SAQ), which sought feedback on all aspects of their experience conducting TIMSS 2019. The feedback solicited in the SAQ included an evaluation of the quality of the assessment materials and the effectiveness of the operations procedures and documentation. The results of the TIMSS 2019 Survey Activities Questionnaire are presented in the final section of this chapter.

TIMSS 2019 Survey Operations Units, Manuals, and Software

To support the National Research Coordinators in conducting the TIMSS 2019 assessments, the TIMSS & PIRLS International Study Center provided step-by-step documentation of all operational activities. Organized into a series of units, the *TIMSS 2019 Survey Operations Procedures* were made available at critical junctures of the project to ensure that National Research Coordinators had all the tools and information necessary to discharge their responsibilities. Also, the procedures units were accompanied by a series of manuals for use by School Coordinators and Test Administrators that National Research

Coordinators could translate and adapt to their local situations. Often, separate versions of the units and manuals were provided for paperTIMSS and for eTIMSS. The TIMSS & PIRLS International Study Center and IEA Hamburg also provided National Research Coordinators and their staff with intensive training in constructed response item scoring and data management.

IEA Hamburg was responsible for the development of the eTIMSS software system, or "eAssessment System" (see Chapter 4). Hosted on IEA Hamburg's servers, the eAssessment System consisted of an integrated series of software modules for authoring achievement items (eTIMSS Item Designer), translating and verifying assessment instruments (eTIMSS Online Translation System), checking the suitability of computers for eTIMSS (eTIMSS System Check Program), administering the assessment to students (eTIMSS Player), monitoring the upload of student response and process data (eTIMSS Online Data Monitor), and scoring constructed response items (eTIMSS Online Scoring System, also known as IEA's CodingExpert Software).

In addition to the eAssessment System and consistent with the goal of automating and streamlining procedures wherever possible, IEA Hamburg provided National Research Coordinators in both eTIMSS and paperTIMSS countries with a range of custom-built software products to support project activities. These included the Windows® Within-School Sampling Software (WinW3S) for sampling and tracking classes and students; the IEA Online SurveySystem (OSS) for administering school, teacher, and home questionnaires online; the IEA CodingExpert Software for documenting scoring reliability; and the IEA Data Management Expert (DME) software for creating and checking data files.

The *TIMSS 2019 Survey Operations Procedures* units were crucial resources for the National Research Coordinators as the units described in detail the tasks the NRCs were responsible for conducting. In the event that some of these tasks were contracted out to other people or organizations, the units ensured that the NRCs had sufficient knowledge of these matters to supervise the activities of the people contracted to conduct aspects of the assessment in their countries.

The following units, manuals, and software systems were provided for administering TIMSS 2019:

- *TIMSS 2019 Survey Operations Procedures Unit 1: Sampling Schools and Obtaining their Cooperation*
- *TIMSS 2019 Survey Operations Procedures Unit 2: Preparing for and Conducting the TIMSS 2019 Field Test*

  Unit 2 consisted of the following sections: Sampling Classes and Field Test Administration, Preparing the Field Test Instruments (paper or electronic), Scoring the Field Test Constructed Response Items, and Creating and Submitting the Field Test Databases. An eTIMSS supplement describing online scoring of the eTIMSS constructed response items also was included.

Unit 2 was accompanied by field test versions of the School Coordinator and Test Administrator Manuals for paperTIMSS and eTIMSS, instructions on "Preparing Computers and/or Tablets for eTIMSS," and a National Quality Control Monitor Manual.

In addition to the manuals, IEA Hamburg provided field test versions of the WinW3S within-school sampling software, the OSS online survey system for questionnaire administration, and the DME data management software.

eTIMSS countries also were provided with field test versions of the following systems: eTIMSS System Check Program, eTIMSS Online Translation System, eTIMSS Player, eTIMSS Online Data Monitor, and eTIMSS Online Scoring System (IEA's CodingExpert Software).

- *TIMSS 2019 Survey Operations Procedures Unit 3: Contacting Schools and Sampling Classes for the TIMSS 2019 Data Collection*

  Unit 3 was accompanied by the main data collection versions of the School Coordinator Manual and the WinW3S within-school sampling software and its manual. eTIMSS countries also received the eTIMSS System Check Program and instructions on "Preparing Computers and/or Tablets for eTIMSS," which provided the necessary information and tools for countries to test their devices for eTIMSS compatibility and prepare them for eTIMSS data collection.

- *TIMSS 2019 Survey Operations Procedures Unit 4: Preparing the TIMSS 2019 Assessment Instruments*

  Separate versions of Unit 4 were provided for paperTIMSS and eTIMSS countries; the latter also received a manual on preparing the paper "bridge" booklets. The eTIMSS version provided access to the eTIMSS Online Translation System, which enabled National Research Coordinators to translate the eTIMSS achievement items into their language(s) of instruction. The translated materials were available online for translation and layout verification by IEA Hamburg and the TIMSS & PIRLS International Study Center (see Chapter 5).

  Unit 4 was accompanied by the main data collection version of the OSS online survey system for online administration of the school, teacher, and home (Early Learning Survey) questionnaires.

- *TIMSS 2019 Survey Operations Procedures Unit 5: Conducting the TIMSS 2019 Data Collection*

  Unit 5 was accompanied by the main data collection versions of the Test Administrator Manuals for paperTIMSS and eTIMSS, the National Quality Control Monitor Manual, and the International Quality Control Monitor Manual.

  eTIMSS countries also received the eTIMSS Player for administering the eTIMSS assessment to students and the eTIMSS Online Data Monitor for monitoring the uploading of the data from the player to the IEA Hamburg data server. Each country's eTIMSS Player was customized to contain the country's translations of the eTIMSS assessment items.

- *TIMSS 2019 Survey Operations Procedures Unit 6: Scoring the TIMSS 2019 Constructed Response Items*

  Unit 6 was accompanied by the main data collection versions of the TIMSS 2019 scoring guides and IEA's CodingExpert Software (online scoring system) and manuals. The CodingExpert Software was used to facilitate eTIMSS online scoring and the trend and cross-country reliability scoring tasks.

- *TIMSS 2019 Survey Operations Procedures Unit 7: Creating and Submitting the TIMSS 2019 Databases*

  Unit 7 was accompanied by the main data collection versions of the DME data management software, codebooks, and manual. The DME software is used for data entry and data verification.

TIMSS 2019 Survey Tracking Forms

TIMSS uses a series of tracking forms to document class sampling procedures, assign assessment instruments, and track school, teacher, and student information, including the participation status of the respondents. The tracking forms also facilitate the data collection and data verification process. Four different tracking forms were used for TIMSS 2019:

- Class Listing Form: This form was completed by each sampled school, listing the eligible classes and providing details about the classes, such as the class stream (if applicable), the number of students, and the names of teachers.
- Student-Teacher Linkage Form: This form was completed for each class sampled, listing the names of the students and their teachers, student birth dates, gender, exclusion codes, and linking the students to their teachers.
- Student Tracking Form: This form was created for each class assessed and was completed by the Test Administrators during test administration. The Test Administrators used this form to verify the assignment of survey instruments to students and to indicate student participation.
- Teacher Tracking Form: This form was completed by each sampled school to indicate the completion of the Teacher Questionnaires.

Operations for Data Collection

The following sections describe the major operational activities coordinated by the National Research Coordinators:

- Contacting schools and sampling classes
- Overseeing translation and preparing assessment instruments
- Managing the TIMSS 2019 assessment administration
- Scoring the constructed response items
- Creating the TIMSS 2019 data files

Two other major TIMSS 2019 operational activities are described in separate chapters of this publication—sampling schools (Chapter 3) and verifying translation and layout of the assessment instruments (Chapter 5).

Contacting Schools and Sampling Classes

Exhibit 6.1 illustrates the major steps in working with schools to sample classes and prepare for the TIMSS assessment administration. Once the school samples were drawn, National Research Coordinators were tasked with contacting schools and encouraging them to take part in the assessments. Depending on the national context, this could involve obtaining support from national or regional educational authorities. Survey Operations Procedures Unit 1 included suggestions on ways to encourage schools to participate in the assessment.

Exhibit 6.1: Diagram of the Sampling Procedures and Preparations for the Assessment Administration Implemented by National Centers and Schools

National Center	Schools

Contacting and Tracking Schools
- Contact sampled schools
- Get started in WinW3S (complete project information, import school sample database, translate/adapt tracking forms)
- Complete/adapt school information
- Record school participation
- Print Class Listing Forms and send them to School Coordinators for completion

List all fourth grade and/or eighth grade classes and their teachers on the Class Listing Form

Class Sampling and Tracking;
Preparing Computers/Tablets for eTIMSS Administration
- Enter school and class information from Class Listing Forms into WinW3S
- Sample classes
- Enter teacher information from Class Listing Forms into WinW3S
- Print Student-Teacher Linkage Forms and send them to School Coordinators for completion
- If school computers/tablets are to be used for eTIMSS administration, send the “Preparing Computers and/or Tablets for eTIMSS” instructions and the eTIMSS System Check Program to School Coordinators

List student information on the Student-Teacher Linkage Forms.
If applicable, run the eTIMSS System Check Program on all available computers/tablets.

Student and Teacher Tracking;
Preparing Instruments for Assessment Administration
- If applicable, confirm with School Coordinators the method for delivering the eTIMSS Player to students
- Enter student information from Student-Teacher Linkage Forms into WinW3S
- Update teacher information and enter student-teacher linkage information from Student-Teacher Linkage Forms into WinW3S
- Assign achievement booklets / item block combinations to students
- Print tracking forms
- Print instrument labels
- Send tracking forms and labeled assessment materials to schools

ASSESSMENT ADMINISTRATION

In cooperation with school principals, National Research Coordinators were responsible for identifying and training School Coordinators for all participating schools. A School Coordinator could be a teacher or guidance counselor in the school, or National Research Coordinators could appoint a member of the national center to fill this role. In some countries, a School Coordinator from the national center was responsible for several schools in an area. School Coordinators were provided with a School Coordinator Manual describing their responsibilities. The School Coordinator Manual was prepared by the TIMSS & PIRLS International Study Center and translated/adapted by national center staff in each country.

The responsibilities of the School Coordinators included providing the national center with information on the school; coordinating the dates, times, and places for testing; identifying and training Test Administrators to administer the assessments; coordinating the completion of the tracking forms; distributing questionnaires; and when necessary obtaining parental permission. If school computers were used for eTIMSS administration, School Coordinators were provided with the “Preparing Computers and/or Tablets for eTIMSS” instructions and the eTIMSS System Check Program in order to test the computers for eTIMSS compatibility and prepare the compatible computers for testing. School Coordinators also confirmed receipt of all assessment materials, oversaw the security of the assessment materials, and ensured the return of the assessment materials to the national center following assessment administration.

School Coordinators also played a critical role in providing information for the sampling process, providing the national center with data on eligible classes in the school. With this information, the national centers used the WinW3S within-school sampling software to sample class(es) within the school. WinW3S tracked school, teacher, and student information and generated the necessary tracking forms and instrument labels used to facilitate both the assessment administration process and data checking during the data cleaning process.

As TIMSS samples intact classes, one of the roles of the School Coordinator was to ensure that every student in the school was listed in one and only one class. This was necessary to ensure that the sample of classes resulted in a representative sample of students, and that every student at the target grade had a chance of being selected. At the fourth grade in most countries, students are taught mathematics and science in the same classroom and therefore the fourth grade classroom was designated as the sampling unit. At the eighth grade, however, students are grouped differently for mathematics and science instruction in many countries, so that a student may take mathematics with one group of students and science with a different group of students. As the sampling required one set of students who could be considered a classroom, eighth grade classrooms usually were defined on the basis of mathematics instruction for the purposes of sampling.

Overseeing Translation and Preparing Assessment Instruments

National Research Coordinators also were responsible for preparing the assessment instruments (paperTIMSS achievement booklets, eTIMSS item block combinations, "bridge" booklets, if applicable, and context questionnaires) for their countries—a process that included overseeing the translation of the assessment instruments. The overarching goal of assessment instrument preparation was to create internationally comparable instruments that were appropriately adapted for the national context of each participating country.

As described in the TIMSS 2019 Assessment Design there were 14 blocks of assessment items for each subject and grade, and these were assembled into 14 TIMSS achievement booklets/item block combinations per grade, with two blocks of mathematics items and two blocks of science items in each booklet/block combination. eTIMSS had two additional block combinations per grade, incorporating the new Problem Solving and Inquiry Tasks (PSIs). Each block/PSI had to be translated only once, even though it was included in two different booklets/item block combinations. For paperTIMSS, countries used Adobe® InDesign® software to link the translated and adapted assessment blocks to the appropriate booklets. Automating this process through InDesign decreased the chances of human error in the production process.

In addition to the main eTIMSS assessment, countries transitioning to eTIMSS had to prepare eight "bridge" booklets for each grade, which were paper versions of eight eTIMSS item block combinations. The bridge booklets were composed entirely of the eight trend item blocks that were previously used in TIMSS 2015 and kept secure for TIMSS 2019. For the bridge booklets, countries also used InDesign software to link their translated and adapted assessment blocks from TIMSS 2015 to the appropriate bridge booklets.

In addition to the 16 trend blocks at each grade level from previous assessments (eight in mathematics and eight in science), twelve new assessment blocks were developed for TIMSS 2019 at each grade level (six mathematics and six science). The new assessment blocks replaced those released after the previous assessment cycle. Also, all four PSIs (two mathematics and two science) for each grade level were newly developed for eTIMSS 2019. Countries administering paperTIMSS 2019 at the fourth grade had the option of administering a less difficult mathematics assessment. The less difficult assessment consisted of nine item blocks previously administered in 2015 in TIMSS or TIMSS Numeracy and five blocks newly developed for TIMSS 2019.

All participating countries and benchmarking entities translated and/or adapted the item blocks into their language(s) of instruction. Countries that participated in the 2015 or 2011 assessment cycles were required to use the same translations that they used in those cycles for the trend assessment blocks.

Similarly, all context questionnaires (school, teacher, student, and, for fourth grade, home questionnaires) were translated/adapted and field tested by all participating countries and evaluated following the field test to gauge the validity and reliability of the various questionnaire scales.

In preparation for translation for both the field test and main data collection, the participating countries received the international version (English) of the achievement booklets/item block combinations and context questionnaires with all the necessary instrument production files, including fonts and graphics files. For the eTIMSS assessment, this was done via the eTIMSS Online Translation System. Instructions on how to use the materials to produce high-quality, standardized instruments were included in the corresponding Survey Operations Procedures units and manuals. IEA Amsterdam and the TIMSS & PIRLS International Study Center also provided a generic Arabic source version of the TIMSS 2019 assessment booklets/item block combinations and context questionnaires. Individual countries adapted the generic source version to local usage.

Once translated and/or adapted, first for the field test and then again for the main data collection, the achievement items and context questionnaires were submitted to IEA Amsterdam for translation verification (see Chapter 5). IEA Amsterdam worked with independent translators to evaluate each country's translations and, when deemed necessary, suggested changes to the text.

After the translations had been verified by IEA Amsterdam, National Research Coordinators assembled the paper-based achievement booklets and context questionnaires using InDesign software, and print-ready copies of the instruments were sent to the TIMSS & PIRLS International Study Center for layout verification and a review of national adaptations. For eTIMSS this also was achieved via the eTIMSS Online Translation System. This review checked that the instruments conformed to the international format and that any adaptations made to the instruments did not unduly influence their international comparability.

Documenting National Adaptations

While preparing national achievement items and context questionnaires, countries sometimes by necessity made adaptations to the international versions. paperTIMSS countries documented all their national adaptations using the National Adaptations Forms (NAFs). eTIMSS countries documented their national adaptations to the achievement test via the eTIMSS Online Translation System and adaptations to the context questionnaires using the National Adaptations Forms.

Separate NAFs were provided for the paper achievement booklets and for the context questionnaires (per grade/assessment). During the translation verification and layout review, the verifiers checked whether the national adaptations were likely to influence the ability to produce internationally comparable data for the items involved. Any questions raised were directed to the NRC for consideration via the NAFs. Bridge booklets had their own Bridge Verification Forms, which were used to track any changes to the national version of the 2015 cycle and to document any layout issues noted during the layout verification.

The documentation was completed and reviewed at various stages of preparing national assessment instruments. Version I of the forms and online documentation was completed during the internal translation and review process and sent along with the rest of the materials for international translation verification. After translation verification, the documentation (Version II) was updated in response to the translation verifier's comments, reflecting any changes resulting from the verification, and sent along with the national assessment instruments for layout and adaptations verification. Following layout verification, the national instruments and documentation were finalized (Version III) and submitted to IEA and the TIMSS & PIRLS International Study Center.

Managing the Administration of the TIMSS 2019 Assessments

Preparing and distributing assessment materials to the participating schools required careful organization and planning on the part of the National Research Coordinators. The assessment materials were packaged and sent to the School Coordinators prior to testing, giving ample time for the School Coordinators to confirm the receipt and correctness of the materials. The school and teacher questionnaires were then distributed, and the other instruments were kept in a secure room until the testing date.

Each sampled class was assigned a Test Administrator who followed procedures described in the Test Administrator Manual to administer the assessments and student questionnaire. Test Administrators were in most cases chosen and trained by School Coordinators, and in some cases, the School Coordinator doubled as the Test Administrator.

Test Administrators were responsible for distributing materials to the appropriate students, reading the instructions provided in the Test Administrator Manual to the students, and timing the sessions. WinW3S systematically assigned achievement booklets/eTIMSS item block combinations and produced labels to facilitate the distribution of the assessment, and Test Administrators used the Student Tracking Form and these labels to distribute the assessment instruments (devices for eTIMSS) to the correct students) and to document student participation. When a class had a participation rate below 90 percent, it was the School Coordinator's responsibility to hold a makeup session for the absent students before returning all of the testing materials to the national center. Using the Test Administration Form, the Test Administrators documented the timing of the testing sessions and information about anything out of the ordinary that took place during assessment administration.

The achievement booklets consisted of two sections and the time allotted for each section of the assessment was standardized and strictly enforced by the Test Administrator. The TIMSS assessment consisted of two parts with each containing two item blocks. To complete each part of the TIMSS achievement test, fourth grade students were allowed 36 minutes and eighth grade students were allowed 45 minutes. For eTIMSS countries, the eTIMSS Player automatically logged students out of the system once the time allowed had expired. There was a required break between the two parts of assessment administration. The break was not to exceed 30 minutes. Students who completed part 1 or part 2 of the assessment before the allotted time were not allowed to leave the testing room and were asked to review their answers or read quietly. Some Test Administrators provided activity sheets for these students.

Following the administration of the TIMSS assessment, students were provided 30 minutes to complete the student questionnaire with extra time provided to students who needed it. Following the administration of the eTIMSS assessment, students also took a short computer-based questionnaire about their experiences and attitudes toward using a computer. During administration of the fourth grade student questionnaire, Test Administrators were permitted to read the questionnaire items aloud together with the students.

eTIMSS was mostly administered via individual USB sticks on individual eTIMSS compatible computers or via Android tablets. Sometimes, the server method was used via a Local Area Network (LAN), which entailed a single eTIMSS compatible computer being used as a local server and students using individual devices connected to the server computer. For eTIMSS, the Test Administrators and School Coordinators submitted/uploaded the eTIMSS data after each testing session. Due to computer shortages, sometimes multiple eTIMSS testing sessions were needed for each class.

Linking Students to their Teachers and Classes

Exhibit 6.2 illustrates the hierarchical identification system codes that were used to link the data among schools, classes, students, and teachers. The school, class, and student IDs were strictly hierarchical, with classes nested within schools and students nested within classes.

Exhibit 6.2: Hierarchical Identification System Codes Used to Link Schools, Classes, Students, and Teachers

Participant	ID Components	ID Structure	Numeric Example
School	School	CCCC	0001
Class	School + Class within the school	CCCCKK	000101 000102
Student	School + Class within the school + Student within the class	CCCCKKSS	00010101 00010201
Teacher	School + Teacher within the school + Linkage number to the sampled class	CCCCTTLL	00010101 00010201

Each teacher was assigned a teacher identification number consisting of the four-digit school number followed by a two-digit teacher number. Since the same teacher could be teaching more than one class within a school, it was necessary to have a unique identification number for each teacher linked to a class. This is achieved by adding a two-digit link number to the six digits of the teacher identification number to create a unique eight-digit identification number.

Online Administration of the School, Teacher, and Home Questionnaires

Countries could choose to administer the school, teacher, and home questionnaires online. The benefits of administering the questionnaires online included saving money and time in printing, and improving the efficiency of questionnaire distribution, data entry, and data cleaning.

For the online administration of the questionnaires, IEA Hamburg provided its IEA Online SurveySystem (OSS) Software that incorporates design, presentation, and monitoring components.

The design component, known as the Designer, supports the preparation of the online surveys, data management, and data output to IEA Hamburg. Through the OSS Designer, national centers could tailor the online questionnaires to their national language. To facilitate translation and adaptation, the Designer concurrently stored the original English question text and the translations and/or national adaptations. It also stored the variable names and data validation rules. If a national center decided not to administer a particular international question or option, it could be disabled in the Designer and not administered during the online questionnaire administration. The Designer also included an integrated preview function to allow for a visual side-by-side comparison of the paper/PDF and online versions of the questionnaires, facilitating the layout verification process.

For the online data collection, the OSS Web Component presented the questionnaires to the respondents. The navigation capabilities of the Web Component allowed respondents to pick and choose their order of response. Buttons marked "next" and "previous" facilitated navigation between adjacent pages, so users could browse through the questionnaire in the same way that they flip through the pages of the paper questionnaire. A hyperlinked interactive "table of contents" allowed the respondents to fluidly navigate to specific questions. Overall, these two functions permitted respondents to answer questions in the order of their choosing. Also, the online questionnaires could be accessed through any standard internet browser on all standard operating systems without any additional software.

Finally, the OSS Monitor component allowed NRCs to monitor the survey responses in real time. Many national centers made extensive use of the Monitor to follow-up with non-respondents.

IEA Hamburg followed a stringent set of procedures to safeguard the confidentiality of the respondents and maintain the integrity of the data. Each respondent received a statement of confidentiality, and information on how to access the online questionnaire. For most countries, the online questionnaire administration was hosted on the IEA Hamburg customized high performance server. This server allowed for the 24-hour availability of the questionnaires during the data collection period, and it also ensured backup and recovery provisions for the data.

Scoring the Constructed Response Items

Constructed response items represent a substantial portion of the TIMSS assessments, and because reliable and valid scoring of these items is critical to the assessment results, the TIMSS & PIRLS International Study Center provided explicit scoring guides for each individual item and extensive training in their use. Also, the Survey Operations Procedures units specified a procedure for efficiently organizing and implementing the scoring activity. Scoring the eTIMSS constructed response items was done online via IEA's CodingExpert Software, which incorporated the IEA standards and reliability procedures.

International scoring training sessions (one for the field test and two for the main data collection—one for Southern Hemisphere countries and another for Northern Hemisphere countries) were conducted where all National Research Coordinators (or country representatives appointed by the National Research Coordinators) were trained to score each of the constructed response items. At these training sessions, the scoring guide for each item was reviewed and applied to a set of example student responses that had already been scored. These example papers were chosen to represent a range of response types and to demonstrate the guides as clearly as possible. Following the example papers, the training participants applied the scoring guides to a different set of student responses that had not yet been scored. The scores to these practice papers were then shared with the group and any discrepancies were discussed.

Following international scoring training, national centers trained their scoring staff on how to apply the scoring guides for the constructed response items. National Research Coordinators were encouraged to create additional example papers and practice papers from student responses collected in their country.

Documenting Scoring Reliability

Because reliable scoring of the constructed response items is essential for high quality data, it is important to document the reliability of the scoring process. A high degree of scorer agreement is evidence that scorers have applied the scoring guides in the same way. The procedure for scoring the TIMSS constructed response items provided for documenting scoring reliability within each country (within-country reliability scoring), over time (trend reliability scoring), and across countries (cross-country reliability scoring) (see results in Chapter 10).

The method for establishing the reliability of the scoring within each country was for two independent scorers to score a random sample of 200 responses for each constructed response item. The degree of agreement between the scores assigned by the two scorers is a measure of the reliability of the scoring process. In collecting the within-country reliability data, it was vital that the scorers independently scored the items assigned to them, and each scorer did not have prior knowledge of the scores assigned by the other scorer. The within-country reliability scoring was integrated within the main scoring procedure and ongoing throughout the scoring process.

The purpose of the trend reliability scoring was to measure the reliability of the scoring from one assessment cycle to the next (i.e., from TIMSS 2015 to TIMSS 2019). The trend reliability scoring required scorers of TIMSS 2019 to score student responses collected in 2015. The scores from 2019 were then compared with the scores awarded in 2015. Trend reliability scoring was conducted using IEA's CodingExpert Software provided by IEA Hamburg.

Student responses included in the trend reliability scoring (200 responses per item) were actual student responses to 22 fourth grade items (13 items for the less difficult mathematics assessment) and/or 27 eighth grade items (4 item blocks) from the TIMSS trend assessment blocks collected during the TIMSS 2015 assessment administration in each country and benchmarking entity. These responses were scanned and provided to each participating country and benchmarking entity, and were scored with IEA's CodingExpert Software. All scorers who scored the trend assessment blocks in 2019 were required to participate in the trend reliability scoring. If all scorers were trained to score all trend items, the software divided the student responses equally among the scorers. If scorers were trained to score specific assessment blocks, National Research Coordinators were able to specify within the software which scorers would score particular blocks, and the software allocated the student responses accordingly. Similar to the within-country reliability scoring, the trend reliability scoring had to be integrated within the main scoring procedure.

Finally, cross-country reliability scoring gave an indication about how consistently the scoring guides were applied from one country to the next. The cross-country reliability scoring also was conducted using IEA's CodingExpert Software. Student responses included in the cross-country reliability scoring (200 responses per item) were student responses to 22 fourth grade items (17 items for the less difficult mathematics assessment) and/or 27 eighth grade items. The same items were used for the trend scoring reliability study. Student responses were collected from the English-speaking countries during the TIMSS 2015 assessment administration. All scorers who could score student responses written in English were required to participate in the cross-country reliability scoring, and the student responses were equally divided among the participating scorers in each country. In most countries, the scoring exercise was completed immediately after all other scoring activities.

Creating the TIMSS 2019 Databases

The data entry process took place from March to May 2018 for the field test, from December 2018 to March 2019 following data collection in the Southern Hemisphere, and June to September 2019 following data collection in the Northern Hemisphere. The procedure for creating the TIMSS 2019 databases included entering sampling and assessment administration information into WinW3S and adding responses from the context questionnaires and achievement booklets using IEA's Data Management Expert (DME) software. IEA Hamburg provided the DME software to accommodate keyboard data entry from the paper instruments. The DME software also offers data and file management capabilities, a convenient checking and editing mechanism, interactive error detection, and quality control procedures.

The eTIMSS achievement test data were captured automatically by submitting them to the IEA Hamburg eTIMSS server immediately after the assessment administration. Countries were provided with the eTIMSS Online Data Monitor to monitor the data submission. The eTIMSS constructed response scoring took place directly in the online database and thus did not require any manual data entry. For the TIMSS 2019 teacher, school, and home questionnaires administered online through the Online SurveySystem (OSS) via the IEA Hamburg server, the data were directly accessible by IEA Hamburg and no further data entry was required.

For manual data entry using the DME software, IEA Hamburg provided international codebooks describing all variables and their properties to ensure that data files produced with this system met the internationally defined rules and standards for data entry. Before being used, however, the international codebooks had to be updated to accommodate any national adaptations to the data collection instruments. These adapted national codebooks then were used to create the TIMSS 2019 data files in each country, with the responses to the context questionnaires, achievement booklets, and Reliability Scoring Sheets keyed into the DME database.

Quality control throughout the data entry process was essential to maintain accurate data. Therefore, National Research Coordinators were responsible for performing periodic reliability checks during data entry and for applying a series of data verification checks provided by both WinW3S and DME systems prior to submitting the databases to IEA Hamburg. To ensure the reliability of the data entry process, data entry staff was required to independently reenter at least 5 percent of the records from each instrument type. An error rate of 1 percent or less was acceptable for the questionnaire files. An error rate of 0.1 percent or less was required for the student achievement files and the reliability scoring files. If the required agreement was not reached, retraining of the key punchers was required.

Both WinW3S and DME systems offered a data verification module that checked for a range of problems, such as inconsistent identification codes, inconsistencies between participation status information and achievement and/or questionnaire data availability, and out-of-range or otherwise invalid codes. The data verification module also verified the integrity of the linkage between the students, teachers, and schools entered into the DME database and tracking of information for those specified in WinW3S. For data captured online (i.e., eTIMSS achievement data and context questionnaires administered online), it was possible to export data availability information and apply data verification to check for inconsistencies via the WinW3S and DME data verification modules.

When all data files had passed the quality control checks, they were submitted to IEA Hamburg, along with data documentation, for further checking and processing. For information on data processing at IEA Hamburg, please refer to Chapter 8 of this publication.

TIMSS 2019 Survey Activities Questionnaire

The Survey Activities Questionnaire was designed to elicit information about National Research Coordinators' experiences in preparing for and conducting the TIMSS 2019 data collection. The questionnaire was composed of six sections and focused on the following:

- Sampling schools and classes
- Translating, adapting, and producing the assessment instruments
- Administering the assessments
- Implementing the National Quality Control Program
- Preparing for and scoring the constructed response items
- Creating and submitting the databases and documentation

All items in the Survey Activities Questionnaire included accompanying comment fields, in which NRC respondents were encouraged to explain their responses, provide additional information, and suggest improvements for the process.

The TIMSS 2019 Survey Activities Questionnaire was administered online via the OSS system and was completed by a total of 65 NRCs, 31 for paperTIMSS and 34 for eTIMSS. The following sections summarize information gathered from the Survey Activities Questionnaire.

Sampling Schools and Classes

The first section of the Survey Activities Questionnaire asked National Research Coordinators about the Survey Operations Procedures units for sampling both schools and classes within the sampled schools. As shown in Exhibit 6.3, 59 National Research Coordinators considered Survey Operations Procedures Unit 1 to be clear and sufficient, and 63 considered Unit 3 to be clear and sufficient. Eight countries reported deviating from the basic TIMSS sampling design. Their reasons for these modifications to the sampling procedures included a change in the way a country identified schools, adjustments for classes based on gender, special needs, or mixed grade levels, identification of schools for the field test and the main data collection at separate times, and the need to oversample for enhanced reporting. Statistics Canada, in cooperation with IEA Hamburg, selected the school samples for all countries and benchmarking participants.

Exhibit 6.3: Survey Activities Questionnaire, Section One—Sampling (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "TIMSS 2019 Survey Operations Procedures Unit 1: Sampling Schools and Obtaining their Cooperation" clear and sufficient?	59	5	1
Were there any conditions or organizational constraints that necessitated deviations from the basic TIMSS sampling design described in the "Survey Operations Procedures Unit 1"?	8	56	1
Did you use the Within-School Sampling Software (WinW3S) to sample classes?	63	0	2
If you answered "yes", did you experience any problems when using the WinW3S software?	17	45	3
Was the information provided in the "TIMSS 2019 Survey Operations Procedures Unit 3: Contacting Schools and Sampling Classes" clear and sufficient?	63	1	1
Did you follow the procedures outlined in "TIMSS 2019 Survey Operations Procedures Unit 3: Contacting Schools and Sampling Classes" for working with the schools to sample classes (e.g., using the appropriate tracking forms in the proposed order to obtain information from School Coordinators)?	53	10	2

Almost all of the National Research Coordinators reported using the Windows® Within-School Sampling Software (WinW3S) provided by IEA Hamburg to select classes within the sampled schools. National Research Coordinators reported experiencing problems using the WinW3S software. Among the issues reported were the slow processing speed, difficulty in accepting 2019 dates, and difficulties created by the status of excluded students.

Ten National Research Coordinators applied some modifications to the procedures outlined in the Survey Operations Procedures Unit 3. For example, some National Research Coordinators used an online survey or online form to gather information from School Coordinators. Some National Research Coordinators did not use the Class Listing Forms because a class-level database was available from the ministry or national center, and one country did not use the Teacher Tracking Forms because there was only one teacher per class in every school. All modifications were reviewed and approved by the TIMSS & PIRLS International Study Center.

Translating, Adapting, and Producing Assessment Instruments

The second section of the Survey Activities Questionnaire asked National Research Coordinators about translating, adapting, assembling, and printing the test materials, as well as issues related to checking the materials and securely storing them. Some eTIMSS-specific questions were asked in this section related to using the eTIMSS Online Translation System, receiving the eTIMSS Player, and preparing USBs in order to deliver eTIMSS to schools and students.

As reported in Exhibit 6.4, almost all National Research Coordinators found the instructions on preparing achievement booklets, context questionnaires, and eTIMSS item block combinations to be clear and sufficient. However, ten countries reported experiencing some problems using the paper-based survey instrument production materials. These problems mostly included issues with fonts and special characters (e.g., for Cyrillic alphabet) and difficulties due to changes in staff between the field test and main data collection. The 13 National Research Coordinators who reported issues with the eTIMSS Online Translation System noted the difficulty in editing the format of some text and images, in adjusting for font-related issues, particularly regarding character-based languages, and in using some shared text across grades. All of the identified problems were resolved either by specialists at the national center or with assistance from IEA Hamburg and the TIMSS & PIRLS International Study Center.

All but three National Research Coordinators reported applying corrections to their survey instruments as suggested by the external translation verifier or the layout verifier. When suggestions were rejected it was because the language suggested was not the most appropriate for the age group or was not consistent with styles used in trend items, because of the National Research Coordinator's strong preference, or due to time constraints.

Exhibit 6.4: Survey Activities Questionnaire, Section Two—Translating, Adapting, and Producing Assessment Instruments (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "TIMSS 2019 Survey Operations Procedures Unit 4: Preparing the Assessment Instruments" clear and sufficient?	61	2	2
Did you encounter any major problems using the assessment instrument InDesign/RTF production/translation materials (used for preparing the paper context questionnaires and achievement booklets)?	10	53	2
Did you encounter any major problems using the eTIMSS Translation System for preparing the eTIMSS achievement test?	13	20	1
After the translation verification (IEA Amsterdam), did you correct your translations/adaptations as suggested by the verifier in the majority of cases?			
paperTIMSS achievement booklets	29	0	2 *(Not Answered)* 34 *(Not Applicable)*
eTIMSS bridge booklets	31	0	3 *(Not Answered)* 31 *(Not Applicable)*

Exhibit 6.4: Survey Activities Questionnaire, Section Two—Translating, Adapting, and Producing Assessment Instruments (Numbers of NRC Responses) (continued)

Question	Yes	No	Not Answered
Context questionnaires	61	0	1 *(Not Answered)* 3 *(Not Applicable)*
eTIMSS achievement test	30	2	2 *(Not Answered)* 31 *(Not Applicable)*
After the layout verification (TIMSS & PIRLS International Study Center), did you correct your assessment instruments as noted by the verifier in the majority of cases?			
paperTIMSS achievement booklets	29	0	2 *(Not Answered)* 34 *(Not Applicable)*
eTIMSS bridge booklets	30	1	3 *(Not Answered)* 31 *(Not Applicable)*
Context questionnaires	60	0	1 *(Not Answered)* 3 *(Not Applicable)*
eTIMSS achievement test	31	1	2 *(Not Answered)* 31 *(Not Applicable)*
Did you apply any quality control measures to check paper assessment instruments during the printing process (e.g., checking for missing pages, upside down pages, text too bright or too dark)?	58	4	3
Did you experience any problems receiving the eTIMSS Player(s) from IEA Hamburg and preparing the eTIMSS USB sticks and/or tablets?	2	31	1
Did you apply quality control measures to check random eTIMSS USBs (e.g., number of files, size of the files, initiating the eTIMSS Player) before they were provided to schools?	26	1	7
Did you take measures to protect the security of the assessment instruments during the preparing and duplicating process?	61	3	1
Did you detect any potential breaches in security of the assessment instruments?	0	64	1
Did you encounter any problems preparing the Online SurveySystem files for administering the school, teacher, and/or home (Early Learning Survey) questionnaires online?	6	23	1 *(Not Answered)* 35 *(Not Applicable)*

Nearly all of the countries conducted the recommended quality control checks during the process of printing the testing materials for paperTIMSS and preparing devices for eTIMSS. Samples of the printed material were checked for any missing pages, pages in the wrong order, upside down pages, and text being too dark or too light. For eTIMSS, countries randomly sampled USB sticks/tablets to ensure the size of the files and/or that they were operating properly.

Six countries reported that they experienced problems with the IEA's Online SurveySystem (OSS). They reported issues with adding national questions and adding skip-logic to some questions. These problems were solved with assistance and support from IEA Hamburg.

Assessment Administration

The third section of the Survey Activities Questionnaire addressed the extent to which National Research Coordinators were notified about errors in the testing materials sent to schools. As shown in Exhibit 6.5, a small number of errors were found in the materials. Almost half of such errors were corrected before distributing the materials to the respondents. Errors found after distribution were mostly minor, and were either fixed by School Coordinators or replacement materials were provided. The cases where the errors could not be remedied were reported to the TIMSS & PIRLS International Study Center, where decisions were made about setting the problematic data to "not administered."

Exhibit 6.5: Survey Activities Questionnaire, Section Three—Assessment Administration (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "TIMSS 2019 Survey Operations Procedures Unit 5: Conducting the Data Collection" clear and sufficient?	62	2	1
Were any errors detected in any of the following assessment materials after they were sent to schools?			
paperTIMSS achievement booklets	9	21	1 *(Not Answered)* 34 *(Not Applicable)*
paperTIMSS achievement booklet ID labels	4	26	1 *(Not Answered)* 34 *(Not Applicable)*
eTIMSS bridge booklets	5	26	3 *(Not Answered)* 31 *(Not Applicable)*
eTIMSS bridge booklet ID labels	2	29	3 *(Not Answered)* 31 *(Not Applicable)*
eTIMSS files on USB sticks/tablets	3	28	3 *(Not Answered)* 31 *(Not Applicable)*
Student Questionnaire	5	58	1 *(Not Answered)* 1 *(Not Applicable)*
Student Questionnaire ID labels	2	61	1 *(Not Answered)* 1 *(Not Applicable)*
Learning to Read Survey	2	47	1 *(Not Answered)* 15 *(Not Applicable)*

Exhibit 6.5: Survey Activities Questionnaire, Section Three—Assessment Administration (Numbers of NRC Responses) (continued)

Question	Yes	No	Not Answered
Learning to Read Survey ID labels	1	48	1 *(Not Answered)* 15 *(Not Applicable)*
Student Tracking Forms	3	59	1 *(Not Answered)* 2 *(Not Applicable)*
Teacher Questionnaires	3	59	1 *(Not Answered)* 2 *(Not Applicable)*
Teacher Tracking Forms	0	59	1 *(Not Answered)* 5 *(Not Applicable)*
School Questionnaire	0	63	1 *(Not Answered)* 1 *(Not Applicable)*
School Coordinator Manual(s)	3	57	1 *(Not Answered)* 4 *(Not Applicable)*
Test Administrator Manual(s)	0	61	2 *(Not Answered)* 2 *(Not Applicable)*
If any errors were detected, did you correct the error(s) before the testing began?	19	22	4 *(Not Answered)* 20 *(Not Applicable)*
Did you provide access to the Data Protection Declaration (provided by IEA and/or prepared by your country) to respondents in your country?	30	34	1
Does your country have a confidentiality policy that restricts putting respondents' names on tracking forms and assessment instrument covers?	16	48	1
Did you encounter any problems translating and/or adapting the School Coordinator Manual(s)?	6	58	1
Did you encounter any problems translating and/or adapting the Test Administrator Manual(s)?	6	57	1
Were most/all School Coordinators appointed from within the participating schools?	56	8	1
Did you hold formal training session(s) for School Coordinators?	37	27	1
Were most/all Test Administrators trained by School Coordinators within the participating schools?	37	27	1
Did the Test Administrators document any problems or special circumstances that occurred frequently during the assessment administration (please refer to the completed Test Administration Forms)?	33	31	1

Exhibit 6.5: Survey Activities Questionnaire, Section Three—Assessment Administration (Numbers of NRC Responses) (continued)

Question	Yes	No	Not Answered
If you administered school, teacher, and/or home (Early Learning Survey) questionnaires online, did any of the respondents in your country encounter any problems responding to the online questionnaires?	12	17	36
Who did the devices used for eTIMSS testing belong to?			
Participating schools	10	–	–
Outsourced company	3	–	–
National center	6	–	–
A combination of above	15	–	–
If you used personal computers, did you use the individual USB sticks or the local server method to administer eTIMSS in your country?			
Individual computers/USB sticks	16	–	–
Local server method	3	–	–
Both methods were used	10	–	–
Not applicable, only tablets were used	5	–	–
Did you require/suggest/provide an additional person to help the Test Administrators during the eTIMSS testing sessions?	26	7	1
Did you experience any software-specific problems with the eTIMSS Player(s)?	16	17	1
Did you have a sufficient number of computers/tablets available for all/most schools to test all of the selected students (the whole class) at the same time?	22	11	1

In May 2018, a new General Data Protection Regulation (GDPR) was implemented in the European Union law on data protection and privacy for all individuals within the European Union and the European Economic Area. In order for the TIMSS study to comply with the requirements of the law, IEA provided countries with templates of the Data Protection Declaration for each of the TIMSS 2019 context questionnaires, specifically reflecting the content of each questionnaire. The provided templates were fully compliant with the GDPR of Europe. All European countries prepared a Data Protection Declaration, complying with the GDPR and country-specific amendments to the law, and provided it along with each of the TIMSS 2019 national context questionnaires. Some non-European participating countries also adapted and adopted the declaration as required by law in those countries. Altogether 30 National Research Coordinators responded that they prepared and provided Data Protection Declaration along with national context questionnaires.

Six National Research Coordinators reported difficulties translating the School Coordinator Manual and/or the Test Administrator Manual. Primarily, problems arose when the manual(s) had to be reorganized or adapted and the standardized procedures were modified (e.g., no Class Listing Forms or Teacher Tracking Forms were used). Countries administering both eTIMSS and bridge booklets also had two sets of manuals to prepare.

In 56 countries, School Coordinators were appointed from within the participating schools. In the remaining countries, School Coordinators were from the national center or were contracted externally. In most countries, the National Research Coordinators organized centralized training sessions for School Coordinators. In others, training was conducted through webinars, regional meetings, and online and written materials. In 37 countries, Test Administrators were trained by the School Coordinators within the participating schools. In the remaining countries, Test Administrators were trained by members of the national center staff.

Although the TIMSS administration mostly went well, Test Administrators occasionally reported difficulties. Among the problems documented by Test Administrators were the following: loud noises outside the classroom, some disruptive students, some students being unfamiliar with some of the subject material, some students having difficulty with the language of the test, some technical problems with eTIMSS administration, the length of the student questionnaire in some countries, and some commenting that the test was too long or that there was not enough time to complete it.

Less than half the countries that administered the school, teacher, and/or home questionnaires online reported issues. The great majority of these issues related to typos or user error when typing in the URL or login information. For some countries, the problem was easily solved by providing direct links to the correct web address.

In most countries administering eTIMSS, an additional person helped the Test Administrators during the eTIMSS testing sessions. This was usually the classroom teacher, School Coordinator, or an information technology consultant/expert. Several countries added two people per classroom to help with computer set up as well as any technical issues that arose during the testing session.

In about half the eTIMSS countries, some software-specific problems occurred. In the early sessions, there were some issues with initiating the software that were promptly addressed by IEA Hamburg. Other problems included the system sometimes crashing during testing, timer disabling for special needs students not working properly, inability to close the program, difficulty in using the ruler, unintentionally moving out of the test on tablet touchscreens, and some issues with submitting the data. In all but a few cases, eTIMSS was successfully administered despite the need to resolve the above reported issues.

Twenty-two of the 34 countries administering eTIMSS had enough computers or tablets to test all the selected classes at the same time. The rest of the schools held multiple sessions, from two to nine sessions per school. Two countries reported providing extra computers to schools specifically for the testing sessions.

National Quality Control Program

The fourth section of the Survey Activities Questionnaire addressed the National Quality Control Program that each country implemented during data collection (see Chapter 7). As part of national quality assurance activities, National Research Coordinators were instructed to send National Quality Control Observers to ten percent of the participating schools to observe both TIMSS and eTIMSS test administration and to document compliance with the prescribed procedures. The national program was in addition to the program of International Quality Control visits conducted by IEA. Some countries did not use national monitors due to the additional cost or planning time needed for the program. Others made additional efforts when training Test Administrators or used phone calls, surveys and National Resource Center staff to gather information.

As shown in Exhibit 6.6, when applicable, almost all of the national centers conducted their quality assurance program using the National Quality Control Monitor Manual provided by the TIMSS & PIRLS International Study Center. Among the documented problems detected by the national monitors were eTIMSS technical issues where students needed to change computers during the test, schools saying the fourth grade assessment was too long for students, a high absentee rate due to flu season, and in one country, issues with poor testing facilities.

Exhibit 6.6: Survey Activities Questionnaire, Section Four—National Quality Control Program (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Did you conduct a national quality control program that observed the data collection in the participating schools?	56	8	1
Did you use the National Quality Control Monitor (NQCM) Manual and the Classroom Observation Record provided by the TIMSS & PIRLS International Study Center to conduct your national quality control program?	51	6	8 *(Not Applicable)*
Did your national quality control monitors (NQCMs) document any major problems or special circumstances that occurred frequently during the assessment administration?	9	48	8

Preparing for and Scoring the Constructed Response Items

Exhibit 6.7 provides data on responses to items asking National Research Coordinators about their experiences preparing for and scoring the constructed response items. Almost all National Research Coordinators found the scoring procedures as explained in the Survey Operations Procedures Unit 6: Scoring the Constructed Response Items to be clear and sufficient. Countries reporting problems with the scoring training materials asked for more “borderline” examples, including more detailed explanations within the scoring guides. Almost half of National Research Coordinators reported creating their own national examples and practice papers for training their scorers, as suggested by the TIMSS & PIRLS International Study Center.

Exhibit 6.7: Survey Activities Questionnaire, Section Five—Preparing for and Scoring the Constructed Response Items (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "TIMSS 2019 Survey Operations Procedures Unit 6: Scoring the Constructed Response Items" clear and sufficient?	60	3	2
Did you encounter any major problems using the scoring training materials, provided by the TIMSS & PIRLS International Study Center?	8	55	2
Did you create national scoring training materials in addition to the international scoring training materials?	31	32	2
Did you scan any paper achievement booklets for electronic image scoring?	3	27	1 *(Not Answered)* 34 *(Not Applicable)*
Did you encounter any major procedural problems during the TIMSS 2019 constructed response item scoring in your country?	2	61	1 *(Not Answered)* 1 *(Not Applicable)*
Did you encounter any major problems with the Online Scoring System (IEA's CodingExpert Software)?	12	51	2 *(Not Answered)* 0 *(Not Applicable)*
Did all your scorers participate in scoring student responses of the trend items, including the Trend Reliability Scoring?	36	18	1 *(Not Answered)* 10 *(Not Applicable)*
Did all your scorers participate in the Cross-country Reliability Scoring?	26	34	5 *(Not Answered)* 0 *(Not Applicable)*

Three countries scanned their TIMSS achievement booklets and scored student responses electronically. A small number of countries reported some minor problems using the Online Scoring System (IEA's CodingExpert Software), which was used for all eTIMSS scoring and also for the trend and cross-country reliability scoring for both paper and eTIMSS countries. The reported problems included software-related issues that were addressed early in the process by IEA Hamburg, difficulty assigning items to scorers, and problems with scanned images.

Because English was used for the cross-country reliability scoring task, not all scorers were able to participate. Only one country reported no participation, while the majority reported at least two or more scorers participating. For the countries that did not participate in the previous cycle of TIMSS, the question on the trend reliability scoring procedures did not apply.

Creating and Submitting the Databases and Documentation

The last section of the Survey Activities Questionnaire addressed data entry of the paper assessment instruments, administration data entry, and data quality control activities. As shown in Exhibit 6.8, almost all of the National Research Coordinators found the instructions in Survey Operations Procedures Unit 7: Creating and Submitting the TIMSS 2019 Databases to be clear and sufficient. Some National Research Coordinators reported issues when using WinW3S, mainly related to import and export functions. For example, the participation status of excluded students created an issue when importing data, and time/date data needed to be entered manually by some countries. IEA Hamburg was able to provide support to countries as needed.

Exhibit 6.8: Survey Activities Questionnaire, Section Six—Creating and Submitting the Databases and Documentation (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "TIMSS 2019 Survey Operations Procedures Unit 7: Creating and Submitting the TIMSS 2019 Databases" clear and sufficient?	59	4	2
Did you encounter any problems entering test administration information and exporting your WinW3S database(s)?	21	42	2
Who primarily entered the test administration information and paper instrument data for your country?			
National center staff	26	–	–
Temporarily hired data entry staff	9	–	–
An external data entry firm	4	–	–
Combination of the above	22	–	–
Other	3	–	–
Did you use manual (key) data entry to enter paper instrument data for your country?			
paper achievement booklets	26	2 *(optical scanning)*	3 *(Not Answered)* 34 *(Not Applicable)*
eTIMSS bridge booklets	21	9 *(optical scanning)*	4 *(Not Answered)* 31 *(Not Applicable)*
Context questionnaires	52	10 *(optical scanning)*	1 *(Not Answered)* 2 *(Not Applicable)*
Did you encounter any major problems using the IEA's Data Management Expert (DME) software?	2	61	1 *(Not Answered)* 2 *(Not Applicable)*

Exhibit 6.8: Survey Activities Questionnaire, Section Six—Creating and Submitting the Databases and Documentation (Numbers of NRC Responses) (continued)

Question	Yes	No	Not Answered
If you entered paper data manually, did you enter 5% of each assessment instrument twice as a quality control measure?	21	5	3 *(Not Answered)* 34 *(Not Applicable)*
Did you apply all the data quality checks described in the "TIMSS 2019 Survey Operations Procedures Unit 7: Creating and Submitting the TIMSS 2019 Databases" before submitting your data and documentation to IEA Hamburg?	63	1	2
Have you stored all the assessment instruments in a secure storage area until the original documents can be destroyed?	62	0	3

In 26 countries, the national center staff entered data from the paper instruments and 22 countries used a combination of national center staff, temporarily hired staff, and an external data entry firm. Some countries used optical scanning instead of manual data entry. All countries but one reported applying all required data quality checks. All countries reported having securely stored their original assessment instruments until all data are processed and reported, and these materials can be destroyed. The non-responses here correspond to the benchmarking participants for whom data entry and instrument storage was done centrally for the whole country.

CHAPTER 7

International Quality Assurance Program for TIMSS 2019

Ieva Johansone
Susan Flicop

Standardized assessment materials and survey operations procedures were developed and adapted from previous cycles so that the TIMSS 2019 data collection met the highest standards. To document data collection activities and verify that the standardized procedures were followed, the TIMSS & PIRLS International Study Center, working with IEA Amsterdam, developed and implemented an International Quality Assurance Program, whereby International Quality Control Monitors visited a sample of schools in each country and observed the TIMSS 2019 assessment administration. The purpose of this chapter is to provide an overview of the International Quality Assurance Program and report on the data collected through this program.

Overview

The International Quality Assurance Program was implemented by independent International Quality Control Monitors (IQCMs) appointed by IEA Amsterdam. The major task of the IQCMs was to conduct site visits during the data collection process. In each country, the IQCM visited a sample of 15 participating schools at each grade during the assessment administration. When there were one or more benchmarking participants from the same country and only one centrally organized national center responsible for all aspects of data collection, the IQCM visited five additional schools in each benchmarking entity in addition to the schools visited for the country as a whole. In countries transitioning to eTIMSS during the 2019 assessment cycle, three additional schools per grade were visited for the paper "bridge" booklet administration.

In each school visited, IQCMs observed the testing sessions and recorded their observations, noting any deviations from the standardized administration script, timing, and procedures. They also interviewed the School Coordinators about their experiences coordinating the assessment. For paperTIMSS, the ICQMs verified that the suggestions made by the international translation and layout verifiers had been integrated into the final national versions of both the paper achievement booklets and context questionnaires, as documented in the National Adaptation Forms. This was not necessary for digital instruments as the eTIMSS Translation System was able to track all translation and layout verification comments and subsequent changes.

Prior to beginning their assignments, the IQCMs attended a mandatory training session conducted by the TIMSS & PIRLS International Study Center. There were two training sessions, one for Southern Hemisphere countries (September 2018) and one for Northern Hemisphere countries (January 2019). During the training, IQCMs were introduced to the [TIMSS 2019 Survey Operations Procedures], the assessment design, and context questionnaires. IQCMs were also supplied with a manual detailing their role and responsibilities as well as the necessary materials for completing the quality control tasks.

An important aspect of the International Quality Assurance Program is the independence of the IQCMs from the national centers. In most participating countries and benchmarking entities, IEA Amsterdam recruited IQCMs who had served in the same role in previous IEA assessments. For the remaining countries, National Research Coordinators assisted IEA Amsterdam in nominating an International Quality Control Monitor. The nominated person could not be a member of the national center, a family member, or personal friend of the National Research Coordinator. Often, this person was a school inspector, ministry official, or retired schoolteacher. The IQCM was required to be fluent in both English and the language(s) spoken in the country.

When necessary, the IQCMs were permitted to recruit assistants to effectively cover the territory and testing timetable. For TIMSS 2019, a total of 71 IQCMs were trained across the 64 participating countries and 6 benchmarking participants. In addition, the IQCMs trained more than 200 assistant monitors.

International Quality Control Monitors observed 493 paperTIMSS (including bridge booklet administration) fourth grade testing sessions, 471 eTIMSS fourth grade testing sessions, 322 paperTIMSS (including bridge booklet administration) eighth grade testing sessions, and 383 eTIMSS eighth grade testing sessions. Altogether, IQCMs observed 1,669 testing sessions for TIMSS 2019. The results of the TIMSS 2019 IQCM observations are reported in the following sections of this chapter.

Quality Control Observations of the TIMSS 2019 Data Collection

International Quality Control Monitors (IQCMs) conducted site visits during the assessment administration to a sample of schools in each country. For each school visit, the IQCMs completed the Classroom Observation Record. The records were completed online via the IEA's Online SurveySystem.

The observation records were organized into the following sections:

- Section A—Documentation of the TIMSS Testing Session
- Section B—Summary Observations of the TIMSS Testing Session

- Section C—Student Questionnaire Administration and Distribution of the Early Learning Survey
- Section D—Interview with the School Coordinator

Documentation and Summary Observations of the TIMSS 2019 Testing Sessions

Sections A and B of the Classroom Observation Record addressed activities that took place during the testing sessions. The assessments were administered in two parts with a break of up to 30 minutes between each part. During test administration, IQCMs were asked to observe the activities of the Test Administrator, such as distributing, collecting, and securing the testing materials, following the assessment administration script, and timing the testing sessions.

The percentages of IQCM responses on these activities are reported in Exhibit 7.1 for paperTIMSS fourth grade testing sessions, Exhibit 7.2 for eTIMSS fourth grade, Exhibit 7.3 for paperTIMSS eighth grade, and Exhibit 7.4 for eTIMSS eighth grade. IQCMs reported that the assessments were conducted in accordance with the international procedures.

Exhibit 7.1: Observations of paperTIMSS 2019 Fourth Grade Administration Sessions – 493 Observations (Percentage of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Did the Test Administrator distribute test booklets according to the booklet assignment on the Student Tracking Form and booklet labels?	98	2	0
Was the total testing time for Part 1 of the testing session equal to the time allowed?	95	5	0
Did the Test Administrator announce, “You have 10 minutes left” prior to the end of Part 1 of the testing session?	93	7	0
Were there any other “time remaining” announcements made during Part 1 of the testing session?	29	71	0
Was the total time for the break between Part 1 and Part 2 of the testing session equal to or less than 30 minutes?	97	3	0
Were the booklets left unattended or unsecured during the break?	4	96	0
Was the total testing time for Part 2 of the testing session equal to the time allowed?	94	6	0
Did the Test Administrator announce “you have 10 minutes left” prior to the end of Part 2 of the testing session?	91	9	0
Were there any other “time remaining” announcements made during Part 2 of the testing session?	28	72	0

Exhibit 7.1: Observations of paperTIMSS 2019 Fourth Grade Administration Sessions – 493 Observations (Percentage of IQCM Responses) (continued)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Did any students finish either Part 1 or Part 2 of the assessment early (before the time allowed was up)?	85	15	0
Did the Test Administrator have a timer (watch with a seconds hand, a stopwatch, a timer, or a phone with timer) for accurately timing the testing session?	98	2	0
Were the booklets collected and secured after the testing session?	97	3	0

Exhibit 7.2: Observations of eTIMSS 2019 Fourth Grade Administration Sessions – 471 Observations (Percentage of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Did the Test Administrator make sure that students were seated at their assigned computers/tablets (logged into the eTIMSS Player with his/her Student ID and password) according to the Student Tracking Form?	98	2	0
Did the Test Administrator announce, “You have 10 minutes left” prior to the end of Part 1 of the testing session?	82	17	1
Were there any other “time remaining” announcements made during Part 1 of the testing session?	22	77	1
Was the total time for the break between Part 1 and Part 2 of the testing session equal to or less than 30 minutes?	94	5	1
Were the computers and USB sticks or tablets kept secure during the break (e.g., the Test Administrator or a teacher remained in the classroom)?	95	4	1
Did the Test Administrator announce “you have 10 minutes left” prior to the end of Part 2 of the testing session?	80	19	1
Were there any other “time remaining” announcements made during Part 2 of the testing session?	22	77	1
Did the Test Administrator submit the data from each computer/tablet students used for the eTIMSS testing session directly after the testing session?	77	22	1
Did any students finish either Part 1 or Part 2 of the assessment early (logged out before the time was up)?	89	11	0

Exhibit 7.3: Observations of paperTIMSS 2019 Eighth Grade Administration Sessions – 322 Observations (Percentage of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Did the Test Administrator distribute test booklets according to the booklet assignment on the Student Tracking Form and booklet labels?	98	2	0
Was the total testing time for Part 1 of the testing session equal to the time allowed?	92	8	0
Did the Test Administrator announce, “You have 10 minutes left” prior to the end of Part 1 of the testing session?	87	13	0
Were there any other “time remaining” announcements made during Part 1 of the testing session?	33	67	0
Was the total time for the break between Part 1 and Part 2 of the testing session equal to or less than 30 minutes?	96	4	1
Were the booklets left unattended or unsecured during the break?	5	95	0
Was the total testing time for Part 2 of the testing session equal to the time allowed?	93	7	0
Did the Test Administrator announce “you have 10 minutes left” prior to the end of Part 2 of the testing session?	91	9	0
Were there any other “time remaining” announcements made during Part 2 of the testing session?	29	71	0
Did any students finish either Part 1 or Part 2 of the assessment early (before the time allowed was up)?	78	22	0
Did the Test Administrator have a timer (watch with a seconds hand, a stopwatch, a timer, or a phone with timer) for accurately timing the testing session?	95	5	0
Were the booklets collected and secured after the testing session?	94	6	0

Exhibit 7.4: Observations of eTIMSS 2019 Eighth Grade Administration Sessions – 383 Observations (Percentage of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Did the Test Administrator make sure that students were seated at their assigned computers/tablets (logged into the eTIMSS Player with his/her Student ID and password) according to the Student Tracking Form?	98	1	1
Did the Test Administrator announce, “You have 10 minutes left” prior to the end of Part 1 of the testing session?	82	17	1
Were there any other “time remaining” announcements made during Part 1 of the testing session?	22	77	1

Exhibit 7.4: Observations of eTIMSS 2019 Eighth Grade Administration Sessions – 383 Observations (Percentage of IQCM Responses) (continued)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Was the total time for the break between Part 1 and Part 2 of the testing session equal to or less than 30 minutes?	90	7	3
Were the computers and USB sticks or tablets kept secure during the break (e.g., the Test Administrator or a teacher remained in the classroom)	95	2	3
Did the Test Administrator announce “you have 10 minutes left” prior to the end of Part 2 of the testing session?	82	17	1
Were there any other “time remaining” announcements made during Part 2 of the testing session?	21	78	1
Did the Test Administrator submit the data from each computer/tablet students used for the eTIMSS testing session directly after the testing session?	66	33	1
Did any students finish either Part 1 or Part 2 of the assessment early (logged out before the time was up)?	81	18	1

In those sessions where the total testing time for a part of the paperTIMSS administration was not equal to the time allowed, many IQCMs reported that it was because students completed their work a few minutes before the allotted time had elapsed. When a few minutes over the time allowed were reported, it was usually the result of a discrepancy in timekeeping. Most classes received a 10-minute announcement, while 29 percent of paperTIMSS and 22 percent of eTIMSS classes also received at least one more timing announcement, most frequently a 5-minute or 2-minute announcement. When the break exceeded 30 minutes, it was often due to schools deciding to follow their regular break schedule. These extended breaks were usually reported to be 35 to 45 minutes in duration.

In accordance with the procedure at the end of the testing session for paperTIMSS, Test Administrators were asked to collect and secure the test booklets. The IQCMs reported that in 97 percent of the fourth grade testing sessions and in 94 percent of the eighth grade sessions this occurred. After each eTIMSS session, Test Administrators were asked to upload the data to IEA’s eTIMSS server, which received, stored securely, and time-stamped all uploads. The international monitors reported observing 77 percent of fourth grade Test Administrators and 66 percent of eighth grade Test Administrators submitting the data directly after the testing sessions. In the remaining sessions, the data was either uploaded via the server computer when the server method was used, or the USBs were removed with the data to be uploaded at a later time.

Exhibits 7.5, 7.6, 7.7, and 7.8 report on the activities conducted during the assessment sessions for fourth grade paperTIMSS, fourth grade eTIMSS, eighth grade paperTIMSS, and eighth grade eTIMSS, respectively. To standardize test administration, all Test Administrators were instructed to read the script in the Test Administrator Manual to the students. IQCMs reported that in 74 percent of fourth grade and 83 percent of eighth grade paperTIMSS observations, the Test Administrators followed the script exactly. For eTIMSS, 63 percent of both fourth grade and eighth grade Test Administrators followed the script exactly. When the Test Administrator deviated from the script, nearly all modifications were reported to be "minor."

Exhibit 7.5: paperTIMSS Fourth Grade Test Administrators Following the Test Administration Script – 493 Observations (Percentage of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
In your opinion, had the Test Administrator familiarized himself or herself with the test administration script prior to the testing?	93	4	3 *(I Cannot Answer)* 0 *(Not Answered)*
Did the Test Administrator follow the test administration script in the ***Test Administrator Manual***?	74	23 *(Minor changes)* 3 *(Major changes)*	0
If the Test Administrator made changes to the script, how would you describe them?			
Additions	16	8	76 *(Not Answered)* 0 *(Not Applicable)*
Revisions	11	12	77 *(Not Answered)* 0 *(Not Applicable)*
Deletions	10	13	77 *(Not Answered)* 0 *(Not Applicable)*
In your opinion, did the Test Administrator address students' questions appropriately?	98	2	0

Exhibit 7.6: eTIMSS Fourth Grade Test Administrators Following the Test Administration Script – 471 Observations (Percentage of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
In your opinion, had the Test Administrator familiarized himself or herself with the test administration script prior to the testing?	94	4	2
Did the Test Administrator follow the test administration script in the ***Test Administrator Manual***?	63	32 *(Minor changes)* 4 *(Major changes)*	1
If the Test Administrator made changes to the script, how would you describe them?			
Additions	26	7	67 *(Not Answered)* 0 *(Not Applicable)*
Revisions	19	14	67 *(Not Answered)* 0 *(Not Applicable)*
Deletions	11	22	67 *(Not Answered)* 0 *(Not Applicable)*
In your opinion, did the Test Administrator address students' questions appropriately?	98	2	0

Exhibit 7.7: paperTIMSS Eighth Grade Test Administrators Following the Test Administration Script – 322 Observations (Percentage of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
In your opinion, had the Test Administrator familiarized himself or herself with the test administration script prior to the testing?	88	7	5 *(I Cannot Answer)* 1 *(Not Answered)*
Did the Test Administrator follow the test administration script in the ***Test Administrator Manual***?	83	15 *(Minor changes)* 2 *(Major changes)*	0
If the Test Administrator made changes to the script, how would you describe them?			
Additions	8	6	86 *(Not Answered)* 0 *(Not Applicable)*
Revisions	6	8	86 *(Not Answered)* 0 *(Not Applicable)*
Deletions	9	6	85 *(Not Answered)* 0 *(Not Applicable)*
In your opinion, did the Test Administrator address students' questions appropriately?	98	2	0

Exhibit 7.8: eTIMSS Eighth Grade Test Administrators Following the Test Administration Script – 383 Observations (Percentage of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
In your opinion, had the Test Administrator familiarized himself or herself with the test administration script prior to the testing?	92	4	3 *(I Cannot Answer)* 1 *(Not Answered)*
Did the Test Administrator follow the test administration script in the ***Test Administrator Manual***?	63	31 *(Minor changes)* 5 *(Major changes)*	0
If the Test Administrator made changes to the script, how would you describe them?			
Additions	21	11	1 *(Not Answered)* 67 *(Not Applicable)*
Revisions	17	16	1 *(Not Answered)* 66 *(Not Applicable)*
Deletions	14	18	1 *(Not Answered)* 67 *(Not Applicable)*
In your opinion, did the Test Administrator address students' questions appropriately?	97	2	1

Exhibits 7.9 and 7.10 summarize observations on student compliance with instructions and overall cooperation during assessment administration for the fourth grade and eighth grade, respectively. The first two questions in each exhibit apply only to the paperTIMSS assessment since the timing and access to the eTIMSS test was controlled on the computer. According to the IQCM's observations, in almost all the paperTIMSS sessions for both grades, students complied well or very well with the instruction to stop work at the end of both part 1 and part 2. As evidenced in the third question in each exhibit for both paperTIMSS and eTIMSS, the IQCMs described the students as extremely or moderately orderly and cooperative during most of the testing sessions.

Exhibit 7.9: Fourth Grade Student Cooperation During Assessment Administration – 493 paperTIMSS Observations and 471 eTIMSS Observations (Percentage of IQCM Responses)

Question	Very Well (%)	Fairly Well (%)	Not well at all (%)	Not Answered or Not Applicable (%)
When the Test Administrator ended Part 1 of the testing session, how well did the students comply with the instructions to stop work (close their booklets and put their pens down)?	84	16	0	0
When the Test Administrator ended Part 2 of the testing session, how well did the students comply with the instructions to stop work (close their booklets and put their pens down)?	88	12	0	0

Question	Extremely (%)	Moderately (%)	Somewhat (%)	Hardly (%)	Not Answered or Not Applicable (%)
To what extent would you describe the students as orderly and cooperative?	67	30	3	0	0

Exhibit 7.10: Eighth Grade Student Cooperation During Assessment Administration – 322 paperTIMSS Observations and 383 eTIMSS Observations (Percentage of IQCM Responses)

Question	Very Well (%)	Fairly Well (%)	Not well at all (%)	Not Answered or Not Applicable (%)
When the Test Administrator ended Part 1 of the testing session, how well did the students comply with the instructions to stop work (close their booklets and put their pens down)?	85	15	0	0
When the Test Administrator ended Part 2 of the testing session, how well did the students comply with the instructions to stop work (close their booklets and put their pens down)?	87	11	2	0

Question	Extremely (%)	Moderately (%)	Somewhat (%)	Hardly (%)	Not Answered or Not Applicable (%)
To what extent would you describe the students as orderly and cooperative?	73	22	4	0	1

Summary Observations of the TIMSS 2019 Testing Sessions

Exhibits 7.11 and 7.12 report on the IQCMs' general observations of the fourth grade paperTIMSS and eTIMSS assessment administrations, respectively, and Exhibits 7.13 and 7.14 report on the IQCMs' general observations of the eighth grade paperTIMSS and eTIMSS administrations, respectively. Overall, IQCMs reported that the quality of testing sessions was good, very good, or excellent (98% for fourth grade paperTIMSS, 96% for fourth grade eTIMSS, 98% for eighth grade paperTIMSS, and 95% for eighth grade eTIMSS). As these numbers show, the IQCMs observed very few issues overall. In only 1 percent of cases for both grade levels for paperTIMSS and 2 percent for both grade levels for eTIMSS did a student refuse to take the test. In addition, more than 92 percent of the observed testing sessions took place under favorable room conditions that were suitable for students to work without distraction. The large majority of students (93% for fourth grade paperTIMSS, 95% for fourth grade eTIMSS, 96% for eighth grade paperTIMSS, and 95% for eighth grade eTIMSS) followed the direction to store away everything, including electronic devices, for the duration of test administration. The IQCMs also reported that in most of observed testing sessions (95% for fourth grade paperTIMSS, 94% for fourth grade eTIMSS, 93% for eighth grade paperTIMSS, and 92% for eighth grade eTIMSS), students were seated in an arrangement that provided adequate space for students to work and not be distracted by one another.

Specific to eTIMSS, IQCMs reported 79 percent of fourth grade and 76 percent of eighth grade testing sessions had additional personnel in the classroom, usually an IT specialist, IT teacher, class teacher or School Coordinator. Regarding the technical problems noted by the IQCMs, most instances were addressed quickly in the classroom without any loss of data.

Exhibit 7.11: General Observations of the paperTIMSS Fourth Grade Testing Sessions – 493 Observations (Percentage of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Did the student identification information on the test booklets correspond with the Student Tracking Form?	99	1	0
Were any defective test booklets detected and replaced?	2 *(BEFORE the testing began)* 2 *(AFTER the testing began)*	98 *(BEFORE the testing began)* 98 *(AFTER the testing began)*	0 *(BEFORE the testing began)* 0 *(AFTER the testing began)*
*If any defective test booklets were replaced, did the Test Administrator replace them appropriately, following instructions in the **Test Administrator Manual**?*	2	0	98 *(Not Answered)* 0 *(Not Applicable)*
Did any students refuse to take the test (do not count the students with parental permission denied)?	1	99	0
If a student refused, did the Test Administrator accurately follow the instructions for excusing the student (collect the test booklet and record the incident on the Student Tracking Form)?	1	0	99 *(Not Answered)* 0 *(Not Applicable)*

Exhibit 7.11: General Observations of the paperTIMSS Fourth Grade Testing Sessions – 493 Observations (Percentage of IQCM Responses) (continued)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Were any late students admitted to the testing room?	3 *(BEFORE testing began)* 2 *(AFTER the testing began)*	94 *(There were no late students)* 1 *(Late students were not admitted)*	0
Did any students leave the room for an "emergency" during the testing?	17	83	0
If a student left the room for an "emergency," did the Test Administrator address the situation appropriately (collect the test booklet, and if readmitted, return the test booklet)?	12	4	84 *(Not Answered)* 0 *(Not Applicable)*
Were there any students requiring special accommodations (e.g., students with visual or hearing impairment, Dyslexia)?	9	91	0
Did students store away everything, including all electronic devices, such as calculators, cell phones, portable computers, and photo or video cameras, having only a pen or a pencil and the test booklet for the duration of the test administration? (Calculators that do not connect to the Internet are permitted for the eighth grade assessment.)	93	7	0
During the testing session, did the Test Administrator walk around the room to be sure students were working on the correct section of the test and/or behaving properly?	96	4	0
In your opinion, were the conditions in the testing room suitable (lighting, temperature, noise, etc.) for the students to work without distractions?	95	4	0
Did the seating arrangement provide adequate space for students to work and not be distracted by each other?	94	6	0
Did you see any evidence of students attempting to cheat on the test (e.g., by copying from a neighbor)?	8	92	0

Question	Excellent (%)	Very Good (%)	Good (%)	Fair (%)	Poor (%)	Not Answered or Not Applicable (%)
In general, how would you describe the overall quality of the testing session?	55	33	10	2	0	0

Exhibit 7.12: General Observations of the eTIMSS Fourth Grade Testing Sessions – 471 Observations (Percentage of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Were any defective USB sticks/tablets detected and replaced?	12 *(BEFORE the testing began)* 6 *(AFTER the testing began)*	87 *(BEFORE the testing began)* 93 *(AFTER the testing began)*	1 *(BEFORE the testing began)* 1 *(AFTER the testing began)*
Did any students refuse to take the test (do not count the students with parental permission denied)?	2	97	1
If a student refused, did the Test Administrator record the incident on the Student Tracking Form?	1	0	99 *(Not Answered)* 0 *(Not Applicable)*
Were any late students admitted to the testing room?	4 *(BEFORE the testing began)* 1 *(AFTER the testing began)*	94 *(There were no late students)* 1 *(Late students were not admitted)*	0
Did any students leave the room for an “emergency” during the testing?	20	80	0
Were there any students requiring special accommodations (e.g., students with visual or hearing impairment, Dyslexia)?	16	84	0
Did students store away everything (school books/papers and all electronic devices), having only the computer/tablet and scratch paper used for the testing session?	95	5	0
In your opinion, were the conditions in the testing room suitable (lighting, temperature, noise, etc.) for the students to work without distractions?	94	5	1
Did the seating arrangement provide adequate space for students to work and not be distracted by each other?	93	7	0
Were all students in the participating class tested together in one testing session or in groups (multiple testing sessions due to the number of computers/tablets available)?	81 *(one session)*	18 *(Multiple sessions)*	1
If laptops were used, did students have an external mouse available?	39	8	4 *(Not Answered)* 49 *(Not Applicable)*
If no, did using the laptop touchpads cause any problems?	1	7	92
In addition to the Test Administrator, were there any additional personnel (e.g., School Coordinator, class teacher, an IT specialist) available during the testing session?	79	20	1
Did any technical problems occur during the testing session?	22	77	1
Did the Test Administrator submit the data from each computer/tablet students used for the eTIMSS testing session directly after the testing session?	77	22	1

Exhibit 7.12: General Observations of the eTIMSS Fourth Grade Testing Sessions – 471 Observations (Percentage of IQCM Responses) (continued)

Question	Excellent (%)	Very Good (%)	Good (%)	Fair (%)	Poor (%)	Not Answered or Not Applicable (%)
In general, how would you describe the overall quality of the testing session?	50	33	13	3	1	0

Exhibit 7.13: General Observations of the paperTIMSS Eighth Grade Testing Sessions – 322 Observations (Percentage of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Did the student identification information on the test booklets correspond with the Student Tracking Form?	99	1	0
Were any defective test booklets detected and replaced?	1 *(BEFORE the testing began)* 1 *(AFTER the testing began)*	99 *(BEFORE the testing began)* 99 *(AFTER the testing began)*	0 *(BEFORE the testing began)* 0 *(AFTER the testing began)*
*If any defective test booklets were replaced, did the Test Administrator replace them appropriately, following instructions in the **Test Administrator Manual?***	1	1	98 *(Not Answered)* 0 *(Not Applicable)*
Did any students refuse to take the test (do not count the students with parental permission denied)?	3	97	0
If a student refused, did the Test Administrator accurately follow the instructions for excusing the student (collect the test booklet and record the incident on the Student Tracking Form)?	2	0	98 *(Not Answered)* 0 *(Not Applicable)*
Were any late students admitted to the testing room?	6 *(BEFORE the testing began)* 4 *(AFTER the testing began)*	86 *(There were no late students)* 4 *(Late students were not admitted)*	0
Did any students leave the room for an “emergency” during the testing?	13	87	0
If a student left the room for an “emergency,” did the Test Administrator address the situation appropriately (collect the test booklet, and if readmitted, return the test booklet)?	8	5	87 *(Not Answered)* 0 *(Not Applicable)*
Were there any students requiring special accommodations (e.g., students with visual or hearing impairment, Dyslexia)?	7	93	0
Did students store away everything, including all electronic devices, such as calculators, cell phones, portable computers, and photo or video cameras, having only a pen or a pencil and the test booklet for the duration of the test administration? (Calculators that do not connect to the Internet are permitted for the eighth grade assessment.)	96	4	0

Exhibit 7.13: General Observations of the paperTIMSS Eighth Grade Testing Sessions – 322 Observations (Percentage of IQCM Responses) (continued)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
During the testing session, did the Test Administrator walk around the room to be sure students were working on the correct section of the test and/or behaving properly?	96	4	0
In your opinion, were the conditions in the testing room suitable (lighting, temperature, noise, etc.) for the students to work without distractions?	93	7	0
Did the seating arrangement provide adequate space for students to work and not be distracted by each other?	96	4	0
Did you see any evidence of students attempting to cheat on the test (e.g., by copying from a neighbor)?	4	96	0

Question	Excellent (%)	Very Good (%)	Good (%)	Fair (%)	Poor (%)	Not Answered or Not Applicable (%)
In general, how would you describe the overall quality of the testing session?	59	28	11	2	0	0

Exhibit 7.14: General Observations of the eTIMSS Eighth Grade Testing Sessions – 383 Observations (Percentage of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Were any defective USB sticks/tablets detected and replaced?	12 *(BEFORE the testing began)* 6 *(AFTER the testing began)*	87 *(BEFORE the testing began)* 93 *(AFTER the testing began)*	1 *(BEFORE the testing began)* 1 *(AFTER the testing began)*
Did any students refuse to take the test (do not count the students with parental permission denied)?	2	97	1
If a student refused, did the Test Administrator record the incident on the Student Tracking Form?	1	0	1 *(Not Answered)* 98 *(Not Applicable)*
Were any late students admitted to the testing room?	8 *(BEFORE testing began)* 3 *(AFTER testing began)*	86 *(There were no late students)* 2 *(Late students were not admitted)*	1
Did any students leave the room for an “emergency” during the testing?	13	86	1
Were there any students requiring special accommodations (e.g., students with visual or hearing impairment, Dyslexia)?	9	90	1

Exhibit 7.14: General Observations of the eTIMSS Eighth Grade Testing Sessions – 383 Observations (Percentage of IQCM Responses) (continued)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Did students store away everything (school books/papers and all electronic devices), having only the computer/tablet and scratch paper used for the testing session?	95	4	1
In your opinion, were the conditions in the testing room suitable (lighting, temperature, noise, etc.) for the students to work without distractions?	92	7	1
Did the seating arrangement provide adequate space for students to work and not be distracted by each other?	91	8	1
Were all students in the participating class tested together in one testing session or in groups (multiple testing sessions due to the number of computers/tablets available)?	74	25 *(multiple sessions)*	1
If laptops were used, did students have an external mouse available?	35	10	2 *(Not Answered)* 53 *(Not Applicable)*
If no, did using the laptop touchpads cause any problems?	1	9	90
Did any technical problems occur during the testing session?	27	72	1
In addition to the Test Administrator, were there any additional personnel (e.g., School Coordinator, class teacher, an IT specialist) available during the testing session?	76	23	1

Question	Excellent (%)	Very Good (%)	Good (%)	Fair (%)	Poor (%)	Not Answered or Not Applicable (%)
In general, how would you describe the overall quality of the testing session?	50	32	13	3	10	1

Student Questionnaire Administration

All Student Questionnaires were administered on paper to all students. Exhibits 7.15 and 7.16 summarize the IQCMs' observations of the Student Questionnaire administration for fourth grade and eighth grade, respectively. IQCMs reported that in the majority of the testing sessions, the Student Questionnaires were distributed according to the Student Tracking Forms and questionnaire labels. In some cases, Test Administrators did not follow the Student Questionnaire administration script exactly. In the cases where the Test Administrator deviated from the script, the modifications were reported to be "minor" for the most part. In 28 percent of the observed testing sessions for fourth grade, Test Administrators read

Student Questionnaire questions aloud, and in 64 percent of the fourth grade sessions students answered these questions independently. It should be noted that some schools chose to administer the questionnaire on a different date than the assessment, and in these cases, IQCMs were not required to observe student questionnaire administration.

Exhibit 7.15: Fourth Grade Student Questionnaire Administration – 964 Observations (Percentage of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Was there a break between the end of the achievement testing session and the distribution and administration of the Student Questionnaires?	79	16	5
Did the Test Administrator distribute the Student Questionnaires according to the *Student Tracking Form* and questionnaire labels?	90	2	7 *(Not Answered)* 1 *(Not Applicable)*
Did the Test Administrator follow the questionnaire administration script in the Test Administrator Manual?	69	19 *(Minor changes)* 3 *(Major changes)*	6 *(Not Answered)* 3 *(Not Applicable)*
If the Test Administrator made changes to the script, how would you describe them?			
Additions	13	7	77 *(Not Answered)* 3 *(Not Applicable)*
Revisions	9	11	77 *(Not Answered)* 3 *(Not Applicable)*
Deletions	9	10	78 *(Not Answered)* 3 *(Not Applicable)*
Did the Test Administrator read the questions aloud to the students?	28	64	6 *(Not Answered)* 2 *(Not Applicable)*
After the Student Questionnaire administration, did the Test Administrator distribute the Early Learning Surveys (Home Questionnaires)?	29	62	7 *(Not Answered)* 2 *(Not Applicable)*
If the Early Learning Surveys were distributed at this time, did the Test Administrator distribute them according to the Student Tracking Form and survey labels?	27	1	70 *(Not Answered)* 2 *(Not Applicable)*

Exhibit 7.16: Eighth Grade Student Questionnaire Administration – 705 Observations (Percentage of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Was there a break between the end of the achievement testing session and the distribution and administration of the Student Questionnaires?	77	20	3
Did the Test Administrator distribute the Student Questionnaires according to the *Student Tracking Form* and questionnaire labels?	92	2	5 *(Not Answered)* 1 *(Not Applicable)*
Did the Test Administrator follow the questionnaire administration script in the Test Administrator Manual?	74	*14 (Minor changes)* 5 *(Major changes)*	3 *(Not Answered)* 4 *(Not Applicable)*
If the Test Administrator made changes to the script, how would you describe them?			
Additions	7	13	76 *(Not Answered)* 4 *(Not Applicable)*
Revisions	7	13	76 *(Not Answered)* 4 *(Not Applicable)*
Deletions	10	11	75 *(Not Answered)* 4 *(Not Applicable)*

Interview with the School Coordinator

Section D was the final component of the Classroom Observation Record and involved the IQCM conducting an interview with the School Coordinator. The interview addressed issues such as the following:

- Shipment of assessment materials
- Arrangements for test administration
- Responsiveness of the national center to queries
- Necessity for make-up sessions
- Information on the target grade classes in the school

Exhibits 7.17, 7.18, 7.19 and 7.20 show the overall ratings by the IQCMs for fourth grade paperTIMSS, fourth grade eTIMSS, eighth grade paperTIMSS, and eighth grade eTIMSS sessions, respectively. Almost all the School Coordinators reported that the TIMSS administration in their school went "very well" or "satisfactorily" overall. In addition, the School Coordinators noted that the School Coordinator Manual worked well for them and most other school staff members had positive attitudes toward TIMSS testing. The larger percentage in the "Needs Improvement" category for eighth grade paperTIMSS was mainly due to one country that combined School Coordinator and Test Administrator responsibilities. The remaining comments noted that the manual was either too detailed or not detailed enough.

Exhibit 7.17: Interview with the School Coordinator, Overview – paperTIMSS Fourth Grade – 493 Records (Percentage of School Coordinator Responses)

Question	Very well, no problems (%)	Satisfactorily, few problems (%)	Unsatisfactorily, many problems (%)	Not Answered or Not Applicable (%)
Overall, how would you say the testing went?	90	10	0	0

Question	Positive (%)	Neutral (%)	Negative (%)	Not Answered or Not Applicable (%)
Overall, how would you rate the attitude of the other school staff members towards TIMSS?	80	18	1	1

Question	Worked well (%)	Needs improvement (%)	Not Answered or Not Applicable (%)
Overall, do you feel the *School Coordinator Manual* worked well for you or does it need improvement?	94	5	1

Exhibit 7.18: Interview with the School Coordinator, Overview – eTIMSS Fourth Grade – 471 Records (Percentage of School Coordinator Responses)

Question	Very well, no problems (%)	Satisfactorily, few problems (%)	Unsatisfactorily, many problems (%)	Not Answered or Not Applicable (%)
Overall, how would you say the testing went?	79	18	2	1

Question	Positive (%)	Neutral (%)	Negative (%)	Not Answered or Not Applicable (%)
Overall, how would you rate the attitude of the other school staff members towards TIMSS?	67	28	3	2

Question	Worked well (%)	Needs improvement (%)	Not Answered or Not Applicable (%)
Overall, do you feel the *School Coordinator Manual* worked well for you or does it need improvement?	89	5	6

Exhibit 7.19: Interview with the School Coordinator, Overview – paperTIMSS Eighth Grade – 322 Records (Percentage of School Coordinator Responses)

Question	Very well, no problems (%)	Satisfactorily, few problems (%)	Unsatisfactorily, many problems (%)	Not Answered or Not Applicable (%)
Overall, how would you say the testing went?	90	9	0	1
Question	**Positive (%)**	**Neutral (%)**	**Negative (%)**	**Not Answered or Not Applicable (%)**
Overall, how would you rate the attitude of the other school staff members towards TIMSS?	85	15	0	0
Question	**Worked well (%)**	**Needs improvement (%)**	**Not Answered or Not Applicable (%)**	
Overall, do you feel the *School Coordinator Manual* worked well for you or does it need improvement?	90	10	0	

Exhibit 7.20: Interview with the School Coordinator, Overview – eTIMSS Eighth Grade – 383 Records (Percentage of School Coordinator Responses)

Question	Very well, no problems (%)	Satisfactorily, few problems (%)	Unsatisfactorily, many problems (%)	Not Answered or Not Applicable (%)
Overall, how would you say the testing went?	82	16	1	1
Question	**Positive (%)**	**Neutral (%)**	**Negative (%)**	**Not Answered or Not Applicable (%)**
Overall, how would you rate the attitude of the other school staff members towards TIMSS?	69	28	1	2
Question	**Worked well (%)**	**Needs improvement (%)**	**Not Answered or Not Applicable (%)**	
Overall, do you feel the *School Coordinator Manual* worked well for you or does it need improvement?	91	7	2	

Exhibits 7.21, 7.22, 7.23, and 7.24 present the details of the School Coordinator interviews for fourth grade paperTIMSS, fourth grade eTIMSS, eighth grade paperTIMSS, and eighth grade eTIMSS, respectively. There were only a small number of cases where components were missing from the shipments of test materials. In some cases where the School Coordinator reported not receiving all of the TIMSS materials, test materials were brought to the school on the testing day by an external Test Administrator. The School Coordinators also reported that in over 90 percent of the schools observed for TIMSS 2019, the national centers were responsive to the school's questions and concerns.

Exhibit 7.21: Interview with the School Coordinator, Details – paperTIMSS Fourth Grade – 493 Records (Percentage of School Coordinator Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Prior to the testing day, did you have time to check the shipment of materials from the national center?	84	15	1
Did you receive the correct shipment of the materials as listed in your *School Coordinator Manual* and according to the tracking forms?	93	7	0
If no, did the national center provide the missing materials in time for the testing?	4	2	94 *(Not Answered)* 0 *(Not Applicable)*
Was the national center responsive to your questions or concerns?	94	4	2
Was the Teacher Questionnaire(s) administered online?	19	80	1
If the Teacher Questionnaire(s) was administered online, did the teacher(s) encounter any problems?	2	15	83 *(Not Answered)* 0 *(Not Applicable)*
Was the School Questionnaire administered online?	19	80	1
If the School Questionnaire was administered online, did the person completing it encounter any problems?	1	17	82 *(Not Answered)* 0 *(Not Applicable)*
Was the Early Learning Survey administered online?	6	88	6
If the Early Learning Survey was administered online, do you know of any problems that parents/guardians encountered?	1	4	95 *(Not Answered)* 0 *(Not Applicable)*
Do you anticipate that a makeup session will be required at your school?	8	92	0
If yes, do you intend to conduct one?	8	1	2 *(Not Answered)* 89 *(Not Applicable)*
Did the students receive any special instructions, motivational talk, or incentives to prepare them for the assessment?	61	39	0
Did you provide the list of classes in the tested grade to the national center?	95	5	0
If there was another international assessment, would you be willing to serve as a School Coordinator?	95	5	0

Exhibit 7.22: Interview with the School Coordinator, Details – eTIMSS Fourth Grade – 471 Records (Percentage of School Coordinator Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Prior to the testing day, did you have time to check the shipment of materials from the national center?	90	9	1
Did you receive the correct shipment of the materials as listed in your *School Coordinator Manual* and according to the tracking forms?	96	2	2
If no, did the national center provide the missing materials in time for the testing?	1	1	98 *(Not Answered)* 0 *(Not Applicable)*
Was the national center responsive to your questions or concerns?	93	2	5
Was the Teacher Questionnaire(s) administered online?	71	25	4
If the Teacher Questionnaire(s) was administered online, did the teacher(s) encounter any problems?	4	61	35 *(Not Answered)* 0 *(Not Applicable)*
Was the School Questionnaire administered online?	71	25	4
If the School Questionnaire was administered online, did the person completing it encounter any problems?	4	60	36 *(Not Answered)* 0 *(Not Applicable)*
Was the Early Learning Survey administered online?	35	56	9
If the Early Learning Survey was administered online, do you know of any problems that parents/guardians encountered?	4	24	72 *(Not Answered)* 0 *(Not Applicable)*
Do you anticipate that a makeup session will be required at your school?	13	86	1
If yes, do you intend to conduct one?	11	1	3 *(Not Answered)* 85 *(Not Applicable)*
Did the students receive any special instructions, motivational talk, or incentives to prepare them for the assessment?	68	31	1
Did you provide the list of classes in the tested grade to the national center?	85	13	1
If there was another international assessment, would you be willing to serve as a School Coordinator?	87	11	2

Exhibit 7.23: Interview with the School Coordinator, Details – paperTIMSS Eighth Grade – 322 Records (Percentage of School Coordinator Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Prior to the testing day, did you have time to check the shipment of materials from the national center?	76	23	1
Did you receive the correct shipment of the materials as listed in your *School Coordinator Manual* and according to the tracking forms?	91	8	1
If no, did the national center provide the missing materials in time for the testing?	1	7	93 *(Not Answered)* 0 *(Not Applicable)*
Was the national center responsive to your questions or concerns?	97	2	1
Was the Teacher Questionnaire(s) administered online?	21	77	2
If the Teacher Questionnaire(s) was administered online, did the teacher(s) encounter any problems?	3	17	80 *(Not Answered)* 0 *(Not Applicable)*
Was the School Questionnaire administered online?	19	78	3
If the School Questionnaire was administered online, did the person completing it encounter any problems?	1	18	81 *(Not Answered)* 0 *(Not Applicable)*
Do you anticipate that a makeup session will be required at your school?	12	88	3
If yes, do you intend to conduct one?	8	3	87 *(Not Answered)* 0 *(Not Applicable)*
Did the students receive any special instructions, motivational talk, or incentives to prepare them for the assessment?	71	29	0
Did you provide the list of classes in the tested grade to the national center?	86	14	0
If there was another international assessment, would you be willing to serve as a School Coordinator?	94	6	0

Exhibit 7.24: Interview with the School Coordinator, Details – eTIMSS Eighth Grade – 383 Records (Percentage of School Coordinator Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Prior to the testing day, did you have time to check the shipment of materials from the national center?	83	11	6
Did you receive the correct shipment of the materials as listed in your *School Coordinator Manual* and according to the tracking forms?	89	8	3
If no, did the national center provide the missing materials in time for the testing?	6	2	1 *(Not Answered)* 91 *(Not Applicable)*
Was the national center responsive to your questions or concerns?	94	2	4

Exhibit 7.24: Interview with the School Coordinator, Details – eTIMSS Eighth Grade – 383 Records (Percentage of School Coordinator Responses) (continued)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Was the Teacher Questionnaire(s) administered online?	81	15	4
If the Teacher Questionnaire(s) was administered online, did the teacher(s) encounter any problems?	7	69	1 *(Not Answered)* 23 *(Not Applicable)*
Was the School Questionnaire administered online?	81	15	3
If the School Questionnaire was administered online, did the person completing it encounter any problems?	4	72	1 *(Not Answered)* 23 *(Not Applicable)*
Do you anticipate that a makeup session will be required at your school?	12	87	1
If yes, do you intend to conduct one?	10	2	84 *(Not Answered)* 4 *(Not Applicable)*
Did the students receive any special instructions, motivational talk, or incentives to prepare them for the assessment?	74	25	2
Did you provide the list of classes in the tested grade to the national center?	77	22	1
If there was another international assessment, would you be willing to serve as a School Coordinator?	87	12	1

There were large but expected differences between schools that administered paperTIMSS and eTIMSS regarding the administration of online Teacher Questionnaires, School Questionnaires, and Early Learning Surveys. The School Questionnaire, for example, was administered online by 19 percent of schools for both the fourth grade and eighth grade in paperTIMSS countries. In comparison, the percentage of eTIMSS countries who administered this questionnaire online was 71 percent for fourth grade and 82 percent for eighth grade classes. Most of the issues reported regarding the use of online questionnaires concerned login information that was received close to the testing day.

In a large number of the visited schools, School Coordinators indicated that students were given special instructions, motivational talks, or incentives by a school official or the classroom teacher prior to testing. This ranged from 61 percent (fourth grade paperTIMSS) to 74 percent (eighth grade eTIMSS).

From 8 to 13 percent of School Coordinators anticipated needing a makeup session and most intended to conduct one.

Because the sampling of classes requires a complete list of all classes in the school at the target grade, IQCMs were also asked to verify that all classes were included in the sampling process. School Coordinators were asked how many classes of the tested grade are in the school, how many were selected to participate, and whether he/she provided the list of classes to the national center. More than 77 percent of School Coordinators confirmed that they sent a complete list of classes to the national center. Most of the remaining School Coordinators reported that centralized databases from Ministries of Education were used instead of class lists.

As a reflection of the successful planning and implementation of TIMSS 2019, 95 percent of fourth grade paperTIMSS respondents, 87 percent of fourth grade eTIMSS respondents, 94 percent of eighth grade paperTIMSS respondents and 87 percent of eighth grade eTIMSS respondents said that they would be willing to serve as a School Coordinator in future international assessments.

CHAPTER 8

Creating the TIMSS 2019 International Database

Kamil Kowolik
Mark Cockle
Milena Taneva

Preparing the TIMSS 2019 International Database and ensuring its integrity was a complex endeavor requiring extensive collaboration among IEA Hamburg, the TIMSS & PIRLS International Study Center, Statistics Canada, and the national centers of participating countries. Once the countries had created their data files and submitted them to IEA Hamburg, an exhaustive process of checking and editing known as "data cleaning" began. Data cleaning is the process of checking data for inconsistencies and formatting the data to create a standardized output.

For each TIMSS assessment, the overriding concerns of the data cleaning process are to ensure the following:

- All information in the database conformed to the internationally defined data structure
- The content of all codebooks and documentation appropriately reflected national adaptations to questionnaires
- All variables used for international comparisons were in fact comparable across countries (after harmonization, where necessary)
- All institutions involved in this process applied quality control measures throughout in order to assure the quality and accuracy of the TIMSS 2019 data.

For TIMSS 2019, IEA Hamburg was responsible for checking the data files from each country, applying standardized data cleaning rules to verify the accuracy and consistency of the data, and documenting any deviations from the international file structure. In addition, IEA Hamburg was responsible for processing and cleaning the data collected by eAssessment Player that delivers the assessment to students, importing student achievement response data for human-scoring into

IEA's Scoring System, and implementing machine scoring rules for achievement items according to specifications from the TIMSS & PIRLS International Study Center.

For countries participating in eTIMSS, student achievement data files were created by IEA Hamburg from the raw data collected by the eAssessment Player as well as the achievement item scores assigned through the Scoring System. For paperTIMSS instruments (including "bridge" booklets forming the basis for a link between paperTIMSS and eTIMSS) and paper-based context questionnaires, data files were created at each country's national center and reviewed prior to submission to IEA Hamburg. The National Research Coordinators (NRCs) from each participating country collaborated with IEA Hamburg to resolve any queries which emerged during the data cleaning process, and the NRCs checked interim versions of the national/benchmarking participant database(s) produced by IEA Hamburg. The TIMSS & PIRLS International Study Center provided the NRCs with univariate data almanacs containing summary item statistics on each variable so that the national centers could evaluate their data from an international perspective (see Chapter 10).

The TIMSS & PIRLS International Study Center also conducted all operational psychometric analyses of the achievement and context questionnaire data, as documented in Chapter 12 (achievement scaling) and Chapter 16 (context scaling), and produced achievement scores (plausible values), and context questionnaire scores, as well as other derived variables based on the context data. Using the Within-School Sampling Software (WinW3S)[1] database and response data provided by IEA Hamburg, Statistics Canada in collaboration with IEA Hamburg calculated the sampling weights, population coverage, and school and student participation rates—as documented in Chapter 3 and Chapter 9.

Data Sources

All data collected as part of TIMSS 2019 arrived at IEA Hamburg for processing and cleaning before going to the TIMSS & PIRLS International Study Center for verification and analysis and to Statistics Canada for calculating sampling weights and outcomes. This included data collected from: 1) paperTIMSS instruments, including achievement booklets and all context questionnaires; 2) IEA's Online SurveySystem which countries could use to administer home, teacher, and school questionnaires; and 3) the eAssessment Player that delivered the TIMSS assessment and a short questionnaire to students.

Data Entry and Verification of Paper Instruments

Each national center was responsible for entering the responses collected in paperTIMSS achievement booklets and paper-based context questionnaires into data files using the IEA Data Management Expert (DME) software. The DME is a software system developed by IEA Hamburg that facilitates data entry

1 WinW3S is a software developed by IEA Hamburg that stores participation information at school, teacher, class, and student levels in a relational database while maintaining a hierarchical ID system. The software allows users to perform all necessary within-school sampling according to the TIMSS standards, and also provides some data validation in and across these levels.

and includes validation checks to identify inconsistencies. As a general principle, national centers were instructed to enter data for any questionnaire that contained at least one valid response, discarding unused or empty instruments. This applied to countries that administered paperTIMSS as well as eTIMSS countries, as these countries administered at least some questionnaires on paper and also administered paper "bridge" booklets of achievement items to a subsample of students.

National centers entered responses from the paper instruments into data files using a predefined international codebook. The codebook defines the structure of the data to be entered and contains information about the variable names, lengths, labels, and missing codes, as well as variable ranges for continuous measures or counts and valid values for nominal or ordinal questions.

As documented in Chapter 5, countries participating in TIMSS are expected to make national adaptations to certain questions in the international questionnaires (e.g., the questions about parents' education must be adapted to the national context). Countries making such adaptations were required to adapt the codebook structure to reflect the adaptations made to the national questionnaire versions before beginning the data entry process.

To ensure consistency across participating countries, the basic rule for data entry in the DME required national staff to enter data "as is" without any interpretation, correction, truncation, imputation, or cleaning.

The guiding principles for data entry included the following:

- Responses to closed response items were coded as "1" if the first option was used, "2" if the second option was marked, and so on
- Responses to open response questions, for example number of students in the TIMSS class, were entered "as is" even if the value was outside the originally expected range
- Responses to filter questions and filter-dependent questions were entered exactly as filled in by the respondent, even if the information provided is logically inconsistent
- Non-response, ambiguous responses, responses given outside of the expected format, or conflicting responses (e.g., selection of two options in a multiple-choice question) were coded as "omitted or invalid."

As each respondent ID number was entered it was checked by the DME software for alignment with a five-digit checksum generated by WinW3S. A mistype in either the ID or the checksum resulted in an error message prompting the person entering the data to check the entry. The data-verification module of DME also checked for a range of other issues such as inconsistencies in identification codes and out-of-range or otherwise invalid codes. When such issues were flagged by the software, the individuals entering the data were prompted to resolve the inconsistency or confirm that an issue existed before resuming data entry.

Double-Data Entry

To check data entry reliability in participating countries, national centers were required to enter a 5 percent sample of each survey instrument (achievement booklet or questionnaire) twice by two different data entry persons (punchers) operating independently. IEA Hamburg recommended that countries begin the double-data entry process as early as possible during the data capture period in order to identify possible systematic misunderstandings or mishandlings of data-entry rules and to initiate appropriate remedial actions—for example, retraining national center staff. Those entering the data were required to resolve discrepancies between the first and second data entries by consulting the original questionnaire and applying the international rules in a uniform way.

Although it was desirable that each and every discrepancy be resolved before submission of the complete dataset, the acceptable level of disagreement between the originally entered and double-entered data was established at 1 percent or less for questionnaire data and at 0.1 percent or less level for achievement data. Values above this level required resolution of the discrepancy and re-entry of data.

The level of disagreement between the originally entered and double-entered data was evaluated by IEA Hamburg, and it was found that in general the margin of error observed for processed data was well below the required threshold.

Data from Online Questionnaire Administration

As documented in Chapter 6: Survey Operations Procedures for TIMSS 2019, national centers had the option of administering the school, teacher, and home questionnaires online through IEA's Online SurveySystem instead of or in addition to using paper-based questionnaires. In addition, National Research Coordinators from participating countries completed the TIMSS 2019 Curriculum Questionnaire through this system.

To ensure confidentiality, national centers provided every respondent with a letter containing individual login information along with information on how to access the online questionnaire. This login information corresponded to the ID and checksum provided from WinW3S, meaning that the identity validation step occurring at the national centers for paper-based questionnaires occurred when the respondents' logged-in to the survey.

Online administration of questionnaires had a number of advantages. Because responses were collected in digital format and stored directly on the IEA Hamburg server, there was no need for data entry, reducing the workload for national centers. Also, the online system does not allow for inconsistent response patterns, meaning that the data collected had fewer inconsistencies when compared with data collected through the paper-based questionnaires. For example, if the directions ask the respondent to "Check one circle for each line," the system does not allow the respondent to check more than one response category on each line.

The TIMSS 2019 online questionnaires also include skip logic, which minimized response burden and improved data consistency. The TIMSS questionnaires have a number of questions that filter out respondents—meaning the subsequent questions are not applicable given the response to the filter question. For example, Question 10A of the eighth grade school questionnaire reads “Does your school have a school library? If no, go to Question 11.” If a respondent chooses “No,” the online survey skips directly to Question 11, omitting Questions 10B. Not only does the skip logic save the respondents’ time, it also results in fewer inconsistencies in the data received by IEA Hamburg and instead produces planned missingness of the skipped responses which are coded in the final database as “not applicable.”

Data Verification at the National Centers

Before sending the data to IEA Hamburg for further processing, national centers carried out mandatory validation and verification steps on all entered data and undertook corrections as necessary.

While the questionnaire data were being entered, the data manager or other staff at each national center used the information from the Teacher Tracking Forms to verify the completeness of the materials. Student participation information (e.g., whether a student participated in the assessment or was absent) was entered via WinW3S.

The validation process was supported by an option in WinW3S to generate an inconsistency report. This report listed all of the types of discrepancies between variables recorded during the within-school sampling and test administration process and made it possible to cross-check these data against data entered in the DME, the database for online respondents, and the uploaded student data on the central international server.

Data managers were requested to resolve such issues before final data submission to IEA Hamburg. If inconsistencies remained or the national center could not solve them, IEA Hamburg asked the center to provide documentation on these problems.

Upon submitting the validated data to IEA Hamburg, NRCs also provided extensive documentation including hard copies or electronic scans of all original Student and Teacher Tracking Forms, Student Listing Forms, and when applicable, a report on procedural activities collected as part of the online Survey Activities Questionnaire (see Chapter 6).

Data from eTIMSS Administration

As described in Chapter 4, the eTIMSS assessment was designed to run on PCs and tablets using USB or local server delivery. For both delivery methods, the student response data were stored in a SQLite database, the contents of which could be uploaded to the IEA Hamburg server immediately following the assessment, or later off-site. Following data upload, IEA Hamburg performed some pre-processing, verification, and cleaning steps and then student responses to constructed response items were sent to the IEA Online Scoring System (IEA CodingExpert software), which almost immediately made student responses available to be allocated to scorers. Scoring took place directly on the IEA Hamburg server—allowing IEA Hamburg to monitor, in real time, the progress of scoring within countries.

Also available online to national centers was an upload monitor listing all the student records that had been uploaded to the IEA Hamburg server. In the rare cases that duplicate IDs were detected, the IDs were flagged and national centers indicated which record to keep. The data monitor also allows a list of IDs to be downloaded so that they can be used to update data availability status in WinW3S.

Description of eAssessment Data

The assessment Player recorded student item responses as well as other actions taken by the student and the data were stored in a SQLite database. Student actions were broken down into timestamped events that recorded process data such as navigation behavior and tool use, but also messages to the student that were created by the system (e.g. time remaining towards the end of the test). The events for process data and student responses were stored separately, with the events in an "event table" and item responses in a "response table." There were also auxiliary tables containing the student ID together with the language in which the assessment was administered and information about whether the data had already been uploaded to the IEA server. Other tables were used for error handling.

Each event or item response was stored with both general attributes and attributes specific to that response or event. The following general attributes were recorded:

- Two timestamp parts: The first recorded events and item responses in standard Unix time format and gave the elapsed time in seconds since January 1, 1970. Since a more precise time information was needed for event data, the second timestamp added the milliseconds.
- A sequential number recording the correct sequence of actions: This number reflected the exact order of events and responses and had to coincide with the sequence obtained using the timestamp information.
- A screen ID number: This number indicated the specific screen (or item) on which the response was saved, or the event occurred.
- A page identifier: Due to the rotation of item blocks within booklets, an item could be displayed in different positions in the assessment. Therefore, it was necessary to also include a "page number" as a general attribute.
- An item ID number: For recording responses, the item identifier referred to the particular item or item input (e.g., keyboard field) on the screen. This number corresponded to a given "raw variable name" specified by the TIMSS & PIRLS International Study Center.

- An event-type ID number: For recording events, using ID numbers instead of names helped to minimize data traffic during the assessment administration. A separate reference look-up table held the actual event names that corresponded to the event-type ID numbers.
- A response ID number: For recording responses, this identifier indicated if a response was changed later during the response process. It showed the sequential number (ID) under which the subsequent answer was saved. The final answer the student gave to an item was marked with a "NULL" value for this field.

Attributes specific to each event were stored as JavaScript Object Notation (JSON) objects. JSON objects in general hold for each attribute the name of the attribute (property) and the value of the property. Exhibit 8.1 shows an example extract from the event table for the "UI:IsLoaded" event type. This event indicates that the appropriate test form was loaded with the first item presented to the student. The event-specific attribute is the "index" which is set to zero for the first page of the test, stores as the JSON object {"index":0}.

Exhibit 8.1: Extract from the Event Table for Event Type "UI:IsLoaded"

Event-Type Id	Screen ID	Page Identifier	Information
26	13617	0	{"index":0}

In the response table, each response was stored in a separate record. The response table held the entire response history of each item the student worked on. All item responses were stored as one or more records with string of characters indicating the student response. This could be a single number, but also an extended string containing information about drawn lines or the dragging and dropping of objects. In addition, the student response table contained typed student responses that were later transferred to the Scoring System for human scoring, along with screenshot images of responses from the line-drawing tool. Responses that did not need human scoring were machine scored. For these responses, a set of detailed scoring rules provided by the TIMSS & PIRLS International Study Center were incorporated in a scoring algorithm and applied to each response to determine the appropriate score.

Pre-Processing and Scoring eTIMSS Data

Some pre-processing steps were required to prepare eAssessment data in a suitable format for scoring and further processing. Data uploads from the eTIMSS Players were processed at IEA Hamburg by several data servers that received and time stamped and then extracted the raw data from the uploaded SQLite databases into the "central" SQL database for all countries. This new structure contained a separate database for each country and grade, including all data from the original SQLite databases with the addition of identifiers relating to the import of data and additional fields for scoring purposes.

For eTIMSS countries, the new mode of administration allowed for a substantial portion of the digital items to be machine scored, particularly in mathematics. For eTIMSS items suitable for machine-scoring, the TIMSS & PIRLS International Study Center used scoring guides as the basis for developing machine scoring specifications that could be accurately applied without human judgment of student responses. Developing the machine scoring specifications involved testing each item in the eTIMSS Player, reviewing the output, and writing rules in terms of the output to classify all possible responses to a code in the item's scoring guide. The scoring unit at IEA Hamburg reviewed all specifications and provided feedback on an item-by-item basis, resulting in several rounds of revision until the rules for all items were clarified. The scoring unit at IEA Hamburg then applied the scoring rules for all machine-scored items and the data analysis team at the TIMSS & PIRLS International Study Center independently replicated the results to validate the scoring.

The IEA Scoring System (CodingExpert software) was used by NRCs and their scoring staff to score the eTIMSS constructed response items that were not suitable for machine scoring. Although scoring supervisors controlled the distribution of responses to scorers within countries, the responses themselves became available in the system soon after upload (with some delay due to the asynchronous handling of the import to the central database and thence to the scoring system). To avoid unnecessary scoring, therefore, it was essential that any duplicates in the central database were dealt with before import to the scoring system. In addition to measures to prevent a database from being uploaded a second time from the client side, checks were made to the results database creation date and content to ensure any possible duplicates were flagged before import. There were, however, some kinds of duplicate records that could be legitimate. Two databases with the same student ID but with different creation times could have originated in several scenarios. For example, this could be simply a case of the test administrator mistakenly using the same ID twice for two different students, or an interruption in the assessment may have led to part 1 being conducted from one USB stick and part 2 from a second. Such cases needed to be reconciled by IEA Hamburg's data processing procedures. When scoring was completed, the student response data were transferred to tables prepared for import into the data processing system (DPE) employed at IEA Hamburg for all large-scale international assessments.

Cleaning the International and National Databases

To ensure the integrity of the international database, a uniform data cleaning process was followed, involving regular consultation between IEA Hamburg and the NRCs. After each country had submitted its data, codebooks, and documentation, IEA Hamburg, in collaboration with the NRCs, conducted a four-step cleaning procedure upon the submitted data and documentation:

- A structural check
- A check of the identification (ID) variables
- Linkage cleaning
- Background cleaning

The data cleaning process included numerous iterations of the four-step cleaning procedure and were completed on each national data set in close collaboration with national centers. This repeated multi-step cleaning ensured that all data were properly cleaned and that any new errors that could have been introduced during the data cleaning were rectified. The cleaning process was repeated as many times as necessary until all data were made consistent and comparable. Any inconsistencies detected during the cleaning process were resolved in collaboration with national centers, and all corrections made during the cleaning process were documented in a cleaning report, produced for each country.

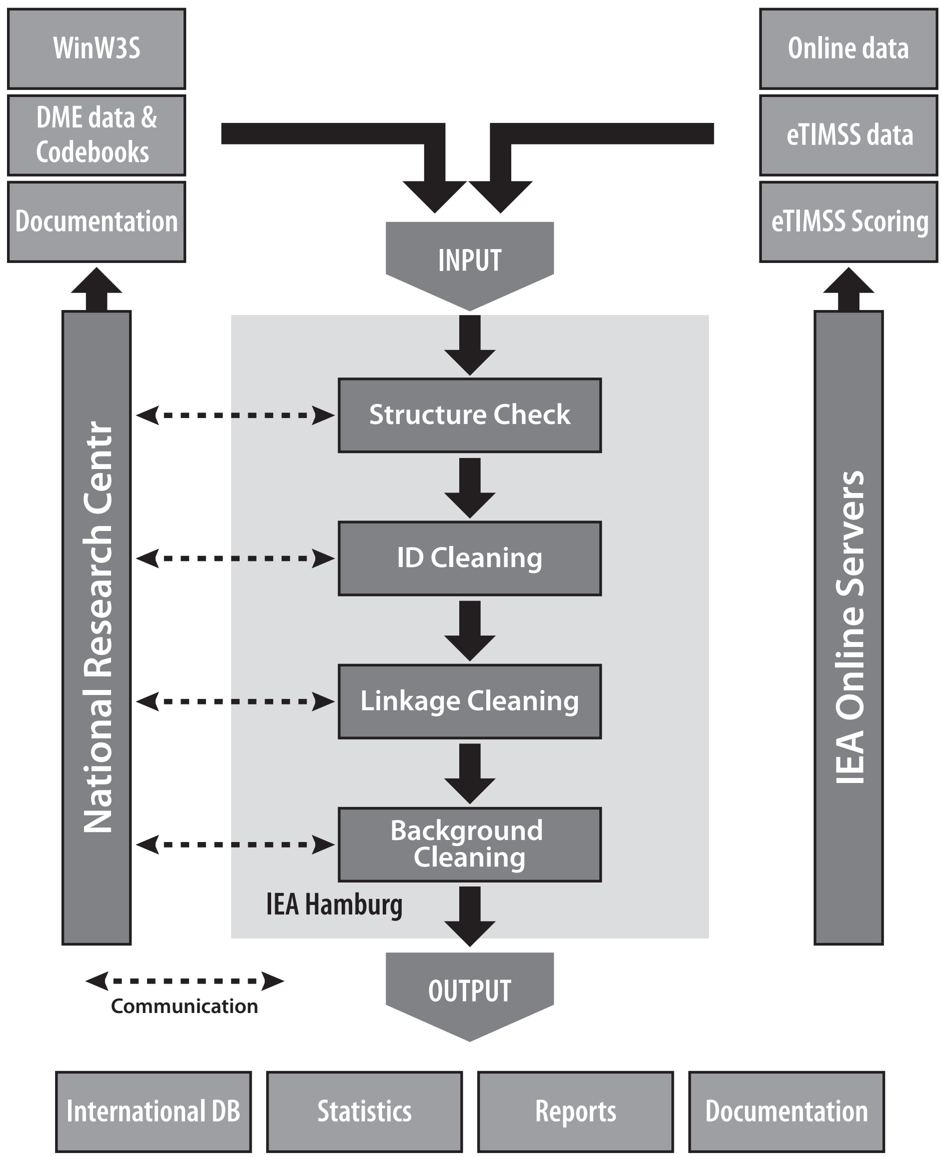
After the final cleaning iteration, each country's data were sent to Statistics Canada for the calculation of sampling weights, and then the data, including sampling weights, were sent to the TIMSS & PIRLS International Study Center so that the psychometric analyses (as described in Chapter 11 and Chapter 12) could be conducted. The NRCs were provided with interim data products to review at different points in the process.

Preparing National Data Files for Analysis

The main objectives of the data cleaning process were to ensure that the data adhered to international formats, that school, teacher, and student information could be linked across different survey files, and that the data reflected the information collected within each country in an accurate and consistent manner.

As illustrated in Exhibit 8.2, the program-based data cleaning consisted of a set of activities explained in the following subsections. IEA Hamburg carried out all of these activities in close communication with the national centers as well as with the TIMSS & PIRLS International Study Center for achievement data.

Exhibit 8.2: Overview of Data Processing at IEA Hamburg



Checking Documentation, Import, and Structure

For each country, data cleaning began with a review of data file structures and its data documentation, including a review of National Adaptation Forms, Student Tracking Forms, Teacher Tracking Forms, Student-Teacher Linkage Forms, and the Survey Activities Questionnaire.

After the review, IEA Hamburg first merged the tracking information and sampling information captured in the WinW3S database with the student-level database containing the corresponding student data from eTIMSS or paperTIMSS achievement assessments. During this step, IEA Hamburg staff also merged the data from the school and teacher questionnaires for both the online and paper modes of administration. At this stage, data from the different sources was transformed and imported into one SQL database so that this information would be available during all further data-processing stages.

The first checks identified differences between the international and the national file structures. Some countries made adaptations (such as adding national variables or omitting or modifying international variables) to their questionnaires. The extent and nature of these changes differed across countries: some countries administered the questionnaires without any modifications (apart from translations and necessary adaptations relating to cultural or language-specific terms), whereas other countries inserted response categories within existing international variables or added national variables.

To keep track of adaptations, staff at the TIMSS & PIRLS International Study Center asked the national centers to complete National Adaptation Forms. In their adaptations, countries sometimes modified the structure and values of the international codebooks, and if IEA Hamburg had to recode variables in the national data files to ensure that the resulting data remained comparable across countries. The national adaptation process is described in Chapter 5 and details about country-specific adaptations to the international instruments can be found in Supplement 2 of the *TIMSS 2019 User Guide for the International Database*.

IEA Hamburg then discarded variables created purely for verification purposes during data entry and made provision for adding new variables necessary for analysis and reporting, including reporting variables, derived variables, sampling weights, and scale scores.

Once IEA staff had ensured that each data file matched the international format, they applied a series of standard data cleaning rules for further processing. Processing during this step employed software developed by IEA Hamburg that identifies and corrects inconsistencies in the data. Each potential problem flagged at this stage was identified by a unique problem number, and then described and recorded in a database. The action taken by the cleaning program or IEA Hamburg staff with respect to each problem was also recorded.

IEA Hamburg referred problems that could not be rectified automatically to the responsible NRC so that national center staff could check the original data-collection instruments and tracking forms to trace the source of these errors. Wherever possible, staff at IEA Hamburg suggested a remedy and asked the national centers to either accept it or propose an alternative. If a national center could not solve the issue through verification of the instruments or forms, IEA Hamburg applied a general cleaning rule to the files to rectify the error. When all automatic updates had been applied, IEA Hamburg staff used SQL recoding scripts to directly apply any remaining corrections to the data files.

Checking Identification Variables

Each record in a data file needs to have a unique identification number. The existence of records with duplicate ID numbers in a file implies an error of some kind. Some countries administered the school, teacher, and home questionnaire (fourth grade only) online in addition to the paper mode. Therefore, by mistake a respondent could have completed both the paper and the online versions of the questionnaire. Similarly, it was possible for an eTIMSS login to be used (and uploaded) twice. If two records in a TIMSS 2019 database shared the same ID number and contained exactly the same data, IEA Hamburg deleted one of the records and kept the other one in the database. In the rare case that both records contained different data and IEA staff found it impossible to identify which record contained the more reliable or complete version of the data, national centers were asked which record to keep.

Although the ID cleaning covered all data from all instruments, it focused mainly on the student data file. In addition to checking the unique student ID number, it was crucial to check variables pertaining to student participation and exclusion status, as well as students' birth dates and dates of testing in order to calculate student age at the time of testing. The Student Tracking Forms provided an important tool for resolving anomalies in the database.

As mentioned previously, IEA Hamburg conducted all cleaning procedures in close cooperation with the national centers. After national center staff had cleaned the identification variables, they passed the clean databases with information about student participation and exclusion on to Statistics Canada, which used this information to calculate students' participation rates, exclusion rates, and student sampling weights.

Checking Linkages

As data on students, parents, teachers, and schools appeared in a number of different data files, a process of linkage cleaning was implemented to ensure that the data files would correctly link together. The linking of the data files followed a hierarchical system of identification codes that included school, class, and student components. These codes linked the students with their class and/or school membership. Further information on linkage codes can be found in Chapter 6: Survey Operations Procedures for TIMSS 2019.

Linkage cleaning consisted of a number of checks to verify that student entries matched across achievement files, student context questionnaire data files, scoring reliability files, and home background files. In addition, at this stage, checks were conducted to ensure that teacher and student records linked correctly to the appropriate schools. The Student Tracking Forms, Teacher Tracking Forms, and Student-Teacher Linkage Forms were crucial in resolving any anomalies. IEA Hamburg also liaised with NRCs about any problematic cases, and the national centers were provided with standardized reports listing all inconsistencies identified within the data.

Resolving Inconsistencies in Context Questionnaire Data

The amount of inconsistent and implausible responses in questionnaire data files varied considerably across countries. IEA Hamburg determined the treatment of inconsistent responses on a question-by-question basis, using all available documentation to make an informed decision. IEA Hamburg staff also checked all questionnaire data for consistency across the responses given. For example, Question 1 in the school questionnaire asked for the total school enrollment in all grades, and Question 2 asked for the enrollment in the target grade only. Logically, the number given as a response to Question 2 could not exceed the number provided by school principals in Question 1. Similarly, it is not possible that the number of years a teacher has been teaching altogether (Question 1 in the teacher questionnaires) exceeds the minimum possible age of a beginning teacher in all participating countries (Question 3 in the teacher questionnaires). IEA Hamburg flagged inconsistencies of this kind and then asked the national centers to review these issues. IEA staff recoded those cases that could not be corrected as "invalid."

Filter questions, which appeared in some questionnaires, directed respondents to a particular set of questions that only applies to a subset of respondents. IEA Hamburg applied the following cleaning rule to these filter questions and the dependent questions that followed, for instance: If a respondent answered "No" to Question 10A in the school questionnaire "Does your school have a school library?" IEA Hamburg recoded any responses to the dependent question 10B as "logically not applicable." Also, following the same example, if the filter question was omitted but at least one valid response was found in the dependent questions then IEA Hamburg recoded the filter question to "Yes." This of course is only possible for dichotomous filter questions (e.g., with response options such "Yes/No").

IEA Hamburg also applied what are known as split variable checks to questions where the answer was coded into several variables. For example, Question 5 in the student questionnaire asked students: "Do you have any of these things at your home?" Student responses were captured in a set of nine variables, each one coded as "Yes" if the corresponding "Yes" option was filled in and "No" if the "No" option was filled in. Occasionally, students checked the "Yes" boxes but left the "No" boxes unchecked. Because, in these cases, it was clear that the unchecked boxes meant "No," these responses were recoded accordingly.

In addition, student reports to items on gender and age in the student questionnaire were checked against the tracking information provided by the School Coordinator or Test Administrator during the within-school sampling and test/questionnaire administration process. When information on gender or birth year and month was missing in the student questionnaire but the student participated, this information, when available, was copied over from the tracking data to the questionnaire. If discrepancies were found between existing tracking and questionnaire gender and age data, IEA Hamburg queried the case with the national center, and the national center investigated which source of information was correct. If unresolved, tracking data was used rather than questionnaire data (or vice versa for some items at the eighth grade).

Handling of Missing Data

Two types of entries were possible during the TIMSS 2019 data capture: valid data values and missing data values. Missing data can be assigned a value of omitted/invalid, or not administered during data capture. IEA Hamburg applied additional missing codes to the data to facilitate further analyses. This process led to four distinct types of missing data in the international database:

- **Omitted or invalid:** The respondent had a chance to answer the question but did not do so, leaving the corresponding item or question blank. This code was also used if the response was uninterpretable or out-of-range.
- **Not administered:** This signified that the item or question was not administered to the respondent, which meant that the respondent could not read and answer the question. The not administered missing code was used for those student test items that were not in the set of assessment blocks administered to a student either deliberately (due to the rotation of assessment blocks) or, in rare cases, due to technical failure or incorrect translations. This missing code was also used for those records that were included in the international database but did not contain a single response to one of the assigned questionnaires. For example, this situation applied to home questionnaire data for students who participated in the student test but the parent/guardian did not answer the home questionnaire. In addition, the not administered code was used for individual questionnaire items that a national center decided not to include in the country-specific version of the questionnaire.
- **Logically not applicable:** The respondent answered a preceding filter question in a way that made the following dependent questions not relevant to him or her.
- **Not reached:** This applied only to the individual items of the student achievement test and indicated those items that students did not attempt due to a lack of time. "Not reached" codes were derived as follows: First, the last answer given by a student in a session is identified. This could be either a valid or invalid response to an item. The first omitted response after this last answer is coded as "omitted," but all following responses to these items in the session are then coded as "not reached." For example, the response pattern "1942999999" (where "9" represents "omitted") is recoded to "19429RRRRR" (where "R" represents "not reached").

Data Cleaning Quality Control

Because TIMSS 2019 was a large and highly complex study with very high standards for data quality, maintaining these standards required an extensive set of interrelated data checking and data cleaning procedures. To ensure that all procedures were conducted in the correct sequence, that no special requirements were overlooked, and that the cleaning process was implemented independently of the persons in charge, the data quality control process included the following steps:

- Thorough testing of all data cleaning programs: Before applying the programs to real datasets, IEA Hamburg applied them to simulation datasets containing all possible problems and inconsistencies
- Registering all incoming data and documents in a dedicated database: IEA Hamburg recorded the date of arrival as well as specific issues requiring attention
- Carrying out data cleaning according to strict rules: Deviations from the cleaning sequence were not possible, and the scope for involuntary changes to the cleaning procedures was minimal
- Documenting all systematic data recoding that applied to all countries: IEA Hamburg recorded all changes to data in the comprehensive cleaning documentation provided to national centers
- Logging every "manual" correction to a country's data files in a recoding script: Logging these changes, which occurred only occasionally, allowed IEA Hamburg staff to undo changes or to redo the whole manual-cleaning process at any later stage of the data cleaning process
- Repeating, on completion of data cleaning for a country, all cleaning steps from the beginning: This step allowed IEA Hamburg to detect any problems that might have been inadvertently introduced during the data cleaning process
- Working closely with national centers at various steps of the cleaning process: IEA Hamburg provided national centers with the processed data files and accompanying documentation so that center staff could thoroughly review and correct any identified inconsistencies.

IEA Hamburg compared national adaptations recorded in the documentation for the national datasets with the structure of the submitted national data files. IEA Hamburg staff then recorded any identified deviations from the international data structure in the national adaptation database and for the supplementary materials provided with the *TIMSS 2019 User Guide for the International Database*. Whenever possible, IEA Hamburg recoded national deviations to ensure consistency with the international data structure.

Interim Data Products

Before the TIMSS 2019 International Databases were finalized, three major interim versions of the data files were sent to each country—each country receiving only its own data. In addition, countries that administered eTIMSS received files with student raw responses. These raw response files are the trace of what students answered and are in this sense comparable to the completed paper booklets that paperTIMSS countries would have available for checking. The first version of the databases was sent as soon as the data could be considered "clean" as regards identification codes and linkage issues. Documentation, with a list of the cleaning checks and corrections made in the data, was included in the first sendout to enable the National Research Coordinators to review the cleaning process before the 7th NRC meeting Agadir, Morocco in December 2019. A second version of the data was sent to the countries when all national adaptations and the feedback resulting from the review of the first version were implemented at the end of February 2020. National Research coordinators were asked to confirm that the data is ready for the operational psychometric analysis used for achievement scaling. A third version of the data files was sent to countries when the weights and international achievement scores were available and had been merged with the data files. This version, sent to the countries in advance of the 8th NRC Meeting in June 2020 contained only those records that were used in the analysis and reports to be released in December 2020 and satisfied the sampling standards, allowed the NRCs to replicate the results presented in the international reports.

Interim data products were accompanied by detailed data processing and national adaptation documentation, codebooks, and summary statistics. The summary statistics were created by the TIMSS & PIRLS International Study Center and included weighted univariate statistics for all questionnaire variables for each country. For categorical variables, representing the majority of variables, the percentages of respondents choosing each of the response options were displayed. For continuous numeric variables, various descriptive statistics were reported, including the minimum, maximum, mean, median, mode, and percentiles. For both types of variables, the percentages of missing data were reported. Additionally, for the achievement items, the TIMSS & PIRLS International Study Center provided item analysis and reliability statistics listing information regarding the number of valid cases, percentages, percentage correct, Rasch item difficulty, scoring reliability, and so forth. These statistics were used for a more in-depth review of the data at the international and national levels in terms of plausibility, unexpected response patterns, etc. More information on item almanacs and reviewing item statistics is available in Chapter 10.

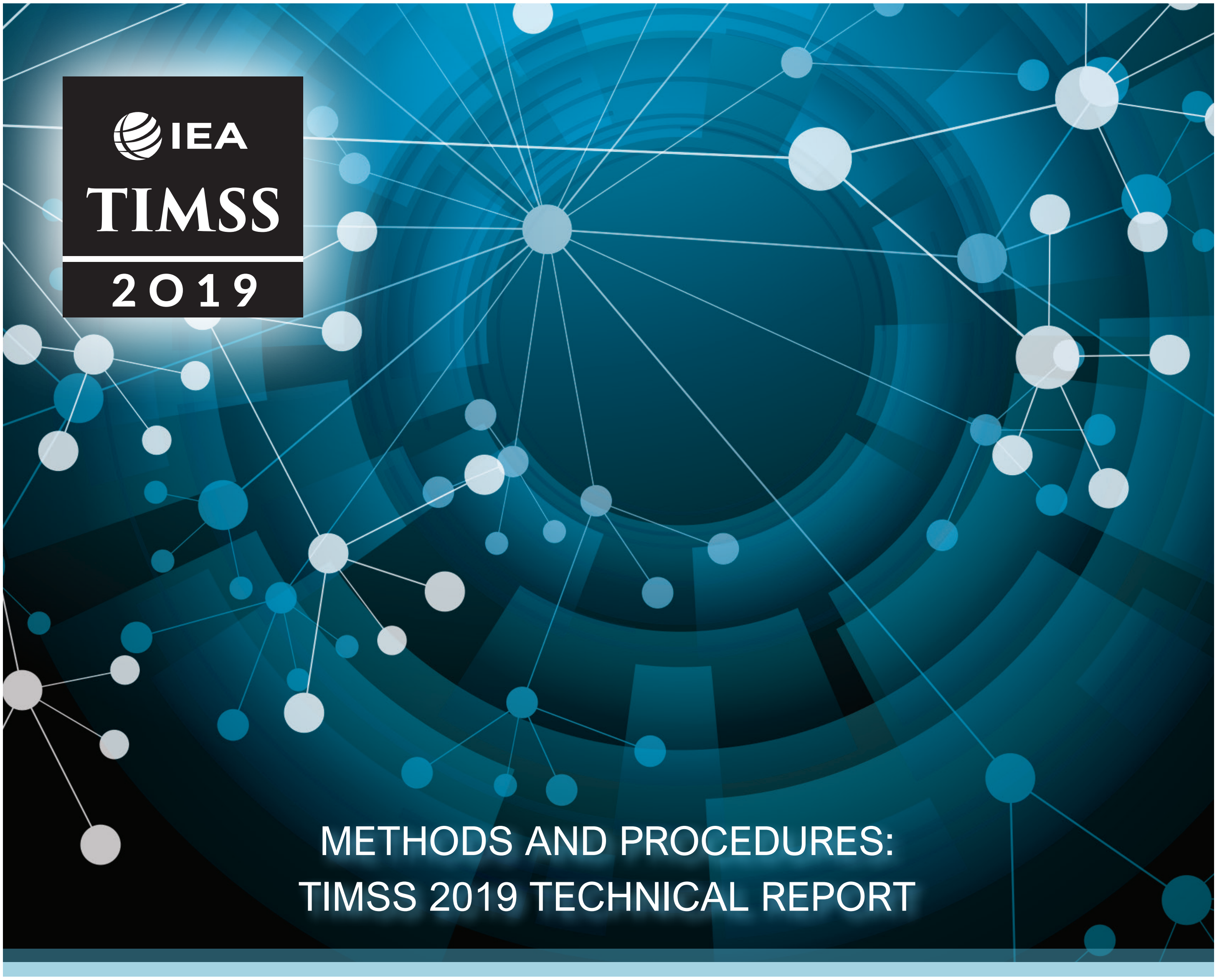
Final Product—the TIMSS 2019 International Databases

The data cleaning effort implemented at IEA Hamburg ensured that the TIMSS 2019 international databases contained high-quality data. More specifically, the process ensured that:

- Information coded in each variable was internationally comparable
- National adaptations were reflected appropriately in all variables

- All entries in the database could be successfully linked within and across levels
- Sampling weights and student achievement scores were available for international comparisons.

Supplements to the *TIMSS 2019 User Guide for the International Database* document all national adaptations made to questionnaires by individual countries and how they were handled in the data. The description of country-specific items also can be found in this supplement, as well as recoding requirements by the TIMSS & PIRLS International Study Center.



ANALYSIS AND REPORTING

IEA
TIMSS & PIRLS
International Study Center
Lynch School of Education
BOSTON COLLEGE

CHAPTER 9

Sample Implementation in TIMSS 2019

Sylvie LaRoche
Pierre Foy

Overview

Rigorous sampling of schools and students was a key component of the TIMSS 2019 project. Implementing the sampling plan was the responsibility of the National Research Coordinator (NRC) in each participating country. NRCs were supported in this endeavor by the TIMSS 2019 sampling consultants, Statistics Canada, and the Sampling Unit of IEA Hamburg. Sampling consultants conducted the school sampling for most countries and trained NRCs in using the Windows® Within-School Sampling Software (WinW3S) provided by IEA Hamburg to implement within-school sampling. As an essential part of their sampling activities, NRCs were responsible for providing detailed documentation describing their national sampling plans (sampling data, school sampling frames, and school sample selections). The documentation for each TIMSS participant was reviewed and completed by the sampling consultants, including detailed information on coverage and exclusion levels, stratification variables, sampling participation rates, and variance estimates. The TIMSS & PIRLS International Study Center and the TIMSS 2019 Sampling Referee, Dr. Keith Rust of Westat, Inc., used this information to evaluate the quality of the samples.

TIMSS 2019 marked the beginning of the TIMSS transition to computer based assessment, with countries having the option of administering the new computer-based version of the 2019 assessment, known as eTIMSS, or the paper-and-pencil version as in previous assessment cycles (paperTIMSS). In order to control for mode effects while linking the two versions to the TIMSS achievement scales and to safeguard the measurement of trends from previous assessments, eTIMSS countries also provided a separate sample of bridge data (see Chapter 3 of this volume).

This chapter gives a summary of the major characteristics of the national samples for TIMSS 2019, followed by a summary of the major characteristics of the bridge samples for trend countries that participated in eTIMSS. More detailed descriptions of the sample design for each country, including details of population coverage and exclusions, stratification variables, and schools' sampling allocations, are provided in Appendix 9A: Characteristics of National Samples.

Target Population

As described in Chapter 3 (Sample Design), the international target populations for the TIMSS 2019 fourth and eighth grade assessments were defined as the grades that represented 4 and 8 years of formal schooling, respectively, counting from the first year of primary or elementary schooling. Countries could assess either one or both student populations. In addition, at the fourth grade for the TIMSS 2019 cycle, countries could administer a less difficult mathematics assessment, consisting of one third of the items from the regular assessment and two-thirds less difficult items, along with the regular fourth grade science assessment.

Albania, Bosnia and Herzegovina, Kosovo, Kuwait, Morocco, North Macedonia, Montenegro, Pakistan, Philippines, and Saudi Arabia chose to administer the less difficult mathematics assessment at the fourth grade while South Africa administered the less difficult mathematics assessment at the fifth grade.

Exhibits 9.1 and 9.2 present the grades identified as the target grades for sampling by each country, and include the number of years of formal schooling that the grades represent and the average age of students in the target grades at the time of testing.

For most countries, the target grades did indeed turn out to be the grades with 4 and 8 years of schooling, i.e., fourth and eighth grades, respectively. However, in England and New Zealand, children begin primary school at an early age.[1] Therefore, these countries administered the TIMSS fourth grade assessment in the fifth year of schooling. The TIMSS eighth grade assessment for England and New Zealand was administered in the ninth year of schooling. Norway chose to assess its fifth and ninth grades to obtain better comparisons with Sweden and Finland.

To provide a better match with the demands of the assessments, South Africa and Turkey availed themselves of the option to assess students at a higher grade. South Africa administered the TIMSS fourth grade with less difficult mathematics assessment at the fifth grade and Turkey administered the TIMSS fourth grade assessment at the fifth grade. South Africa administered the eighth grade assessment at the ninth grade, as did its benchmarking provinces of Gauteng and Western Cape.

Exhibit 9.1: National Grade Definition – TIMSS 2019 – Fourth Grade

Country	Country's Name for Grade Tested	Years of Formal Schooling	Average Age at Time of Testing
Albania	Grade 4	4	10.0
Armenia	Grade 4	4	9.9
Australia	Year 4	4	10.1
Austria	Grade 4	4	10.4

1 Given the cognitive demands of the assessments, TIMSS wants to avoid assessing very young students. Thus, TIMSS recommends assessing the next higher grade (i.e., fifth grade for fourth grade TIMSS and ninth grade for eighth grade TIMSS) if, for fourth grade students, the average age at the time of testing would be less than 9.5 years and, for eighth grade students, less than 13.5 years.

Exhibit 9.1: National Grade Definition – TIMSS 2019 – Fourth Grade (continued)

Country	Country's Name for Grade Tested	Years of Formal Schooling	Average Age at Time of Testing
Azerbaijan	Grade 4	4	10.3
Bahrain	Grade 4	4	9.8
Belgium (Flemish)	Grade 4	4	10.0
Bosnia and Herzegovina	Grade 4	4	10.1
Bulgaria	Grade 4	4	10.7
Canada	Grade 4	4	9.9
Chile	Basic 4	4	10.1
Chinese Taipei	Grade 4	4	10.2
Croatia	Grade 4	4	10.5
Cyprus	Grade 4	4	9.8
Czech Republic	Grade 4	4	10.4
Denmark	Grade 4	4	10.9
England	Year 5	5	10.2
Finland	Grade 4	4	10.8
France	CM1	4	9.9
Georgia	Grade 4	4	10.1
Germany	Grade 4	4	10.4
Hong Kong SAR	Primary 4	4	10.1
Hungary	Grade 4	4	10.5
Iran, Islamic Rep. of	Grade 4	4	10.2
Ireland	Fourth Class	4	10.4
Italy	Primary Grade 4	4	9.6
Japan	Grade 4	4	10.4
Kazakhstan	Grade 4	4	10.4
Korea, Rep. of	Elementary School Grade 4	4	10.5
Kosovo	Grade 4	4	9.9
Kuwait	Grade 4	4	9.7
Latvia	Grade 4	4	10.8
Lithuania	Grade 4	4	10.7
Malta	Year 5	4	9.8
Montenegro	Grade 4	4	9.8
Morocco	Grade 4	4	10.1
Netherlands	Group 6	4	10.1
New Zealand	Year 5	4.5 - 5.5	10.0
North Macedonia	Grade 4	4	9.8

Exhibit 9.1: National Grade Definition – TIMSS 2019 – Fourth Grade (continued)

Country	Country's Name for Grade Tested	Years of Formal Schooling	Average Age at Time of Testing
Northern Ireland	Year 6	4	10.4
Norway (5)	Grade 5	5	10.7
Oman	Grade 4	4	9.7
Pakistan	Grade 4	4	10.6
Philippines	Grade 4	4	10.1
Poland	Primary 4	4	10.3
Portugal	Grade 4	4	10.0
Qatar	Grade 4	4	9.9
Russian Federation	Grade 4	4	10.8
Saudi Arabia	Grade 4	4	9.9
Serbia	Grade 4	4	10.6
Singapore	Primary 4	4	10.4
Slovak Republic	Grade 4	4	10.4
South Africa (5)	Grade 5	5	11.5
Spain	Grade 4	4	9.9
Sweden	Grade 4	4	10.8
Turkey (5)	Grade 4	5	10.6
United Arab Emirates	Grade 4	4	9.7
United States	Grade 4	4	10.2
Benchmarking Participants			
Ontario, Canada	Grade 4	4	9.8
Quebec, Canada	Grade 4	4	10.1
Moscow City, Russian Fed.	Grade 4	4	10.8
Madrid, Spain	Grade 4	4	9.9
Abu Dhabi, UAE	Grade 4	4	9.7
Dubai, UAE	Grade 4	4	9.9

Exhibit 9.2: National Grade Definition – TIMSS 2019 – Eighth Grade

Country	Country's Name for Grade Tested	Years of Formal Schooling	Average Age at Time of Testing
Australia	Year 8	8	14.1
Bahrain	Intermediate 2	8	13.8
Chile	Grade 8	8	14.2
Chinese Taipei	Grade 8	8	14.3
Cyprus	Grade 8	8	13.8
Egypt	Grade 8	8	13.9
England	Year 9	9	14.0
Finland	Grade 8	8	14.8
France	Quatrième	8	13.9
Georgia	Grade 8	8	13.8
Hong Kong SAR	Secondary 2	8	14.1
Hungary	Grade 8	8	14.6
Iran, Islamic Rep. of	Grade 8	8	14.1
Ireland	Second Year	8	14.4
Israel	Grade 8	8	14.0
Italy	Lower Secondary Grade 3	8	13.7
Japan	Lower Secondary Grade 2	8	14.4
Jordan	Grade 8	8	13.9
Kazakhstan	Grade 8	8	14.3
Korea, Rep. of	Middle School Grade 2	8	14.5
Kuwait	Grade 8	8	13.8
Lebanon	Grade 8	8	14.0
Lithuania	Grade 8	8	14.7
Malaysia	Form 2	8	14.3
Morocco	Middle School Year 2	8	14.5
New Zealand	Year 9	8.5 - 9.5	13.9
Norway (9)	Grade 9	9	14.7
Oman	Grade 8	8	13.9
Portugal	Grade 8	8	14.0
Qatar	Grade 8	8	14.0
Romania	Grade 8	8	14.8
Russian Federation	Grade 8	8	14.8
Saudi Arabia	Grade 8	8	13.9
Singapore	Secondary 2	8	14.3

Exhibit 9.2: National Grade Definition – TIMSS 2019 – Eighth Grade (continued)

Country	Country's Name for Grade Tested	Years of Formal Schooling	Average Age at Time of Testing
South Africa (9)	Grade 9	9	15.5
Sweden	Grade 8	8	14.8
Turkey	Secondary 4	8	13.9
United Arab Emirates	Grade 8	8	13.7
United States	Grade 8	8	14.2
Benchmarking Participants			
Ontario, Canada	Grade 8	8	13.8
Quebec, Canada	Secondary 2	8	14.2
Moscow City, Russian Fed.	Grade 8	8	14.8
Gauteng, RSA (9)	Grade 9	9	15.3
Western Cape, RSA (9)	Grade 9	9	15.5
Abu Dhabi, UAE	Grade 8	8	13.7
Dubai, UAE	Grade 8	8	13.9

National Coverage and Exclusions of the TIMSS 2019 National Samples

Exhibits 9.3 and 9.4 summarize population coverage and exclusions for the TIMSS 2019 target populations.

Coverage

National coverage of the international target population was generally comprehensive, with some exceptions. At the fourth grade, these exceptions included Canada (assessed students only from the provinces of Alberta, Manitoba, Newfoundland, Ontario and Quebec) and Georgia (assessed only students taught in Georgian). These participants chose a national target population that was less than the international target population. At the eighth grade, all countries except Georgia (assessed only students taught in Georgian) sampled from 100 percent of their international desired population. For the exceptions where coverage was below 100 percent, the results were footnoted in the TIMSS 2019 international reports.

School-Level and Student-Level Exclusions

Within the national target population, it was possible to exclude certain types of schools and students. For the most part, school-level exclusions consisted of schools for students with disabilities and very small or remote schools. Occasionally, schools were excluded for other reasons, as documented in Appendix 9A: Characteristics of National Samples.

Student-level, or within-school, exclusions generally consisted of students with disabilities or students who could not be assessed in the language of the test. For most participants, the overall percentage of excluded students (combining school and within-school levels) was 5 percent or less after rounding. However, at the fourth grade, Canada, England, Kazakhstan, Kosovo, Latvia, Lithuania, New Zealand, Pakistan, Philippines, Portugal, Russian Federation, Saudi Arabia, Serbia, Slovak Republic, Turkey (5), United States, and the benchmarking participants Ontario and Dubai had exclusions accounting for between 5 and 10 percent of the desired population after rounding, and Singapore had exclusions exceeding 10 percent. At the eighth grade, Egypt, Kazakhstan, Russian Federation, Saudi Arabia, Singapore, Sweden, and the benchmarking participant Dubai had exclusions accounting for between 5 and 10 percent of the national target population after rounding. Israel had exclusions exceeding 10 percent.

Results for participants with an exclusion rate of more than 5 percent after rounding were annotated in the international reports.

Exhibit 9.3: Coverage of TIMSS 2019 – Fourth Grade Target Population

Country	International Target Population		Exclusions from National Target Population		
	Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
Albania	100%		2.6%	1.6%	4.2%
Armenia	100%		0.9%	0.4%	1.2%
Australia	100%		1.9%	2.9%	4.8%
Austria	100%		0.9%	4.5%	5.4%
Azerbaijan	100%		2.3%	0.3%	2.6%
Bahrain	100%		0.4%	0.4%	0.8%
Belgium (Flemish)	100%		0.8%	2.2%	3.0%
Bosnia and Herzegovina	100%		0.6%	1.4%	2.0%
Bulgaria	100%		0.8%	2.6%	3.4%
[1 2] Canada	79%	Students from the provinces of Alberta, Manitoba, Newfoundland, Ontario, and Quebec	3.1%	3.9%	7.0%

Exhibit 9.3: Coverage of TIMSS 2019 – Fourth Grade Target Population (continued)

Country	International Target Population		Exclusions from National Target Population		
	Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
Chile	100%		1.2%	2.6%	3.8%
Chinese Taipei	100%		0.3%	1.6%	2.0%
Croatia	100%		1.1%	3.1%	4.2%
Cyprus	100%		1.1%	3.4%	4.6%
Czech Republic	100%		2.5%	2.2%	4.7%
Denmark	100%		1.6%	1.5%	3.1%
[2] England	100%		2.2%	3.6%	5.8%
Finland	100%		1.8%	1.5%	3.3%
France	100%		2.5%	1.9%	4.4%
[1] Georgia	92%	Students taught in Georgian	2.8%	1.8%	4.7%
Germany	100%		1.7%	2.2%	3.9%
Hong Kong SAR	100%		1.1%	2.4%	3.5%
Hungary	100%		2.1%	2.0%	4.1%
Iran, Islamic Rep. of	100%		3.0%	1.2%	4.2%
Ireland	100%		1.9%	1.1%	3.0%
Italy	100%		0.9%	4.1%	4.9%
Japan	100%		0.6%	1.5%	2.2%
[2] Kazakhstan	100%		2.7%	3.0%	5.8%
Korea, Rep. of	100%		0.9%	1.5%	2.3%
[2] Kosovo	100%		5.3%	3.3%	8.6%
Kuwait	100%		1.0%	0.7%	1.7%
[2] Latvia	100%		3.9%	3.0%	6.9%
[2] Lithuania	100%		2.6%	4.1%	6.7%
Malta	100%		1.4%	3.1%	4.5%
Montenegro	100%		1.3%	3.3%	4.6%
Morocco	100%		1.8%	0.0%	1.8%
Netherlands	100%		2.6%	0.9%	3.5%
[2] New Zealand	100%		2.6%	4.2%	6.9%
North Macedonia	100%		1.2%	2.5%	3.8%
Northern Ireland	100%		2.2%	0.6%	2.8%
Norway (5)	100%		1.4%	3.3%	4.7%
Oman	100%		1.4%	0.8%	2.2%
[2] Pakistan	100%		7.5%	0.0%	7.5%

Exhibit 9.3: Coverage of TIMSS 2019 – Fourth Grade Target Population (continued)

Country	International Target Population		Exclusions from National Target Population		
	Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
[2] Philippines	100%		6.1%	1.6%	7.7%
Poland	100%		1.1%	2.0%	3.1%
[2] Portugal	100%		0.9%	6.9%	7.8%
Qatar	100%		1.2%	1.0%	2.2%
[2] Russian Federation	100%		2.4%	3.9%	6.3%
[2] Saudi Arabia	100%		10.1%	0.4%	10.5%
[2] Serbia	100%		4.0%	4.2%	8.2%
[3] Singapore	100%		12.5%	0.4%	12.8%
[2] Slovak Republic	100%		3.6%	1.9%	5.5%
South Africa (5)	100%		1.1%	0.0%	1.1%
Spain	100%		1.6%	3.8%	5.4%
Sweden	100%		1.6%	3.8%	5.4%
[2] Turkey (5)	100%		1.0%	5.9%	7.0%
United Arab Emirates	100%		1.1%	2.0%	3.2%
[2] United States	100%		0.0%	7.2%	7.2%
Benchmarking Participants					
[2] Ontario, Canada	100%		2.3%	4.7%	7.0%
Quebec, Canada	100%		3.3%	1.2%	4.4%
Moscow City, Russian Fed.	100%		0.7%	1.4%	2.1%
Madrid, Spain	100%		0.5%	3.1%	3.6%
Abu Dhabi, UAE	100%		1.1%	2.5%	3.6%
[2] Dubai, UAE	100%		2.6%	3.0%	5.6%

1 National Target Population does not include all of the International Target Population.
2 National Defined Population covers 90% to 95% of National Target Population.
3 National Defined Population covers less than 90% of National Target Population (but at least 77%).

Exhibit 9.4: Coverage of TIMSS 2019 – Eighth Grade Target Population

Country	International Target Population		Exclusions from National Target Population		
	Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
Australia	100%		1.7%	2.1%	3.8%
Bahrain	100%		0.3%	0.2%	0.6%
Chile	100%		0.3%	1.9%	2.2%
Chinese Taipei	100%		0.1%	1.3%	1.5%
Cyprus	100%		0.5%	2.3%	2.8%
[2] Egypt	100%		7.6%	1.5%	9.1%
England	100%		2.9%	2.0%	4.8%
Finland	100%		1.5%	1.5%	3.1%
France	100%		2.8%	1.0%	3.8%
[1] Georgia	91%	Students taught in Georgian	2.2%	2.1%	4.3%
Hong Kong SAR	100%		1.2%	2.1%	3.3%
Hungary	100%		2.5%	1.9%	4.4%
Iran, Islamic Rep. of	100%		0.4%	0.5%	0.9%
Ireland	100%		0.0%	0.9%	1.0%
[3] Israel	100%		19.5%	3.8%	23.2%
Italy	100%		0.8%	3.6%	4.3%
Japan	100%		0.9%	1.0%	1.8%
Jordan	100%		0.0%	0.2%	0.2%
[2] Kazakhstan	100%		2.9%	2.9%	5.8%
Korea, Rep. of	100%		0.7%	0.9%	1.6%
Kuwait	100%		1.0%	1.0%	2.0%
Lebanon	100%		1.2%	0.0%	1.2%
Lithuania	100%		3.2%	2.0%	5.3%
Malaysia	100%		1.9%	1.3%	3.2%
Morocco	100%		0.0%	0.0%	0.0%
New Zealand	100%		1.5%	2.7%	4.2%
Norway (9)	100%		1.4%	2.5%	4.0%
Oman	100%		0.5%	1.6%	2.2%
Portugal	100%		1.0%	4.5%	5.5%
Qatar	100%		1.3%	0.9%	2.2%
Romania	100%		2.7%	0.5%	3.2%
[2] Russian Federation	100%		2.8%	2.9%	5.7%
[2] Saudi Arabia	100%		9.1%	0.9%	10.0%

Exhibit 9.4: Coverage of TIMSS 2019 – Eighth Grade Target Population (continued)

Country	International Target Population		Exclusions from National Target Population		
	Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
2 Singapore	100%		10.1%	0.2%	10.3%
South Africa (9)	100%		1.0%	0.1%	1.1%
2 Sweden	100%		1.7%	4.6%	6.3%
Turkey	100%		1.1%	2.4%	3.4%
United Arab Emirates	100%		1.1%	1.3%	2.4%
United States	100%		0.0%	3.9%	3.9%
Benchmarking Participants					
Ontario, Canada	100%		2.1%	3.4%	5.5%
Quebec, Canada	100%		3.3%	0.9%	4.2%
Moscow City, Russian Fed.	100%		0.7%	0.8%	1.5%
Gauteng, RSA (9)	100%		1.8%	0.2%	2.1%
Western Cape, RSA (9)	100%		0.8%	0.3%	1.1%
Abu Dhabi, UAE	100%		0.9%	0.8%	1.7%
2 Dubai, UAE	100%		3.0%	2.5%	5.5%

1 National Target Population does not include all of the International Target Population.
2 National Defined Population covers 90% to 95% of National Target Population.
3 National Defined Population covers less than 90% of National Target Population (but at least 77%).

Target Population Size of the TIMSS 2019 National Samples

Exhibits 9.5 and 9.6 show the number of schools and students in each participant's target population[2] and sample, as well as an estimate of the student population size based on the sample data. The target population figures are derived from the sampling frame used to select the TIMSS 2019 samples, while the sample figures are based on the number of sampled schools and students that participated in the assessments. The student population sizes estimated from the sample were computed using sampling weights, which are explained in more detail in Chapter 3. The student population size based on the sampling frame did not take into account the portion of the population excluded within sampled schools and made no adjustment for changes in the population between the date when the information in the sampling frame was collected and the date of the TIMSS 2019 data collection—usually a 2-year interval. Nevertheless, a comparison of the two figures of population size can be seen as a validity check on the sampling procedure. In most cases, the population size estimated from the sample closely matched the population size from the sampling frame.

2 After school-level exclusions.

Exhibit 9.5: Population and Sample Sizes – TIMSS 2019 – Fourth Grade

Country	Population		Sample		
	Schools	Students	Schools	Students	Student Population Size Estimated from Sample
Albania	1,604	33,144	167	4,426	31,609
Armenia	1,028	34,115	150	5,399	36,754
Australia	6,628	301,426	287	5,890	311,753
Austria	3,095	81,406	193	4,464	82,158
Azerbaijan	3,689	145,451	194	5,245	150,309
Bahrain	185	19,466	185	5,762	19,169
Belgium (Flemish)	2,401	78,062	147	4,655	77,006
Bosnia and Herzegovina	587	31,373	178	5,628	29,086
Bulgaria	1,679	63,094	151	4,268	64,338
Canada	9,796	304,798	704	13,653	306,137
Chile	6,081	252,190	169	4,174	250,230
Chinese Taipei	2,476	190,975	162	3,765	188,886
Croatia	1,571	39,244	153	3,785	39,860
Cyprus	289	9,119	151	4,062	9,453
Czech Republic	3,578	114,774	152	4,692	113,904
Denmark	1,644	66,225	166	3,227	66,950
England	15,349	644,127	139	3,396	667,451
Finland	1,840	59,755	158	4,730	59,198
France	31,716	822,438	155	4,186	827,474
Georgia	1,678	42,980	154	3,787	40,185
Germany	17,584	716,091	203	3,437	725,273
Hong Kong SAR	564	60,786	139	2,968	60,761
Hungary	2,888	94,673	149	4,571	89,198
Iran, Islamic Rep. of	38,645	1,334,250	224	6,010	1,261,874
Ireland	2,833	66,818	150	4,582	70,566
Italy	6,809	556,298	162	3,741	549,275
Japan	18,463	1,052,355	147	4,196	1,057,008
Kazakhstan	5,917	289,367	168	4,791	298,341
Korea, Rep. of	5,478	472,130	151	3,893	453,918
Kosovo	620	24,767	145	4,496	24,507
Kuwait	392	53,341	164	4,437	51,932
Latvia	608	20,799	154	4,481	20,657
Lithuania	827	28,035	207	3,741	28,383
Malta	98	4,429	98	3,630	4,461

Exhibit 9.5: Population and Sample Sizes – TIMSS 2019 – Fourth Grade (continued)

Country	Population		Sample		
	Schools	Students	Schools	Students	Student Population Size Estimated from Sample
Montenegro	140	8,034	140	5,076	7,994
Morocco	19,360	672,418	264	7,723	674,824
Netherlands	6,291	178,200	112	3,355	181,849
New Zealand	1,799	63,894	160	5,019	62,839
North Macedonia	326	20,149	150	3,270	19,595
Northern Ireland	771	24,818	134	3,497	25,017
Norway (5)	1,945	62,012	150	3,951	63,745
Oman	736	62,728	228	6,814	63,698
Pakistan	164,364	3,096,192	139	3,980	2,929,483
Philippines	37,092	2,301,861	180	5,515	1,933,761
Poland	12,218	500,265	149	4,882	489,880
Portugal	1,245	99,927	181	4,300	96,042
Qatar	247	25,506	242	4,933	24,518
Russian Federation	40,575	1,414,240	200	4,022	1,602,928
Saudi Arabia	11,216	457,552	220	5,453	455,724
Serbia	2,338	65,777	165	4,380	61,627
Singapore	187	39,934	187	5,986	40,099
Slovak Republic	2,000	52,222	157	4,247	51,506
South Africa (5)	16,254	943,115	297	11,891	1,009,289
Spain	12,861	489,765	501	9,555	493,083
Sweden	3,276	114,494	145	3,965	114,323
Turkey (5)	16,205	1,239,900	180	4,028	1,195,922
United Arab Emirates	754	85,609	688	25,834	85,132
United States	72,902	4,153,454	287	8,776	4,056,773
Benchmarking Participants					
Ontario, Canada	3,683	147,295	163	3,830	147,661
Quebec, Canada	1,764	85,132	148	3,837	88,299
Moscow City, Russian Fed.	695	92,630	150	3,843	102,549
Madrid, Spain	1,343	70,232	167	3,390	72,588
Abu Dhabi, UAE	285	29,938	247	9,037	29,215
Dubai, UAE	184	22,567	199	7,265	23,893

Exhibit 9.6: Population and Sample Sizes – TIMSS 2019 – Eighth Grade

Country	Population		Sample		
	Schools	Students	Schools	Students	Student Population Size Estimated from Sample
Australia	2,471	271,871	284	9,060	282,176
Bahrain	112	17,550	112	5,725	17,204
Chile	5,767	246,120	164	4,115	238,684
Chinese Taipei	931	214,516	203	4,915	205,439
Cyprus	98	8,901	98	3,521	8,856
Egypt	11,061	1,704,928	169	7,210	1,471,594
England	3,706	584,697	136	3,365	591,308
Finland	693	57,591	154	4,874	56,237
France	6,977	814,850	150	3,874	813,845
Georgia	1,837	45,339	145	3,315	44,727
Hong Kong SAR	478	54,160	136	3,265	55,130
Hungary	2,724	87,805	154	4,569	89,223
Iran, Islamic Rep. of	23,895	1,095,026	220	5,980	1,075,783
Ireland	704	65,084	149	4,118	65,561
Israel	979	106,971	157	3,731	108,119
Italy	5,775	566,636	158	3,619	553,839
Japan	10,138	1,098,159	142	4,446	1,094,387
Jordan	2,705	147,483	235	7,176	144,949
Kazakhstan	5,701	225,638	168	4,453	238,290
Korea, Rep. of	3,006	465,626	168	3,861	444,287
Kuwait	348	41,058	171	4,574	46,254
Lebanon	1,746	68,077	204	4,730	65,930
Lithuania	706	25,394	194	3,826	25,427
Malaysia	2,565	423,150	177	7,065	412,165
Morocco	3,469	506,427	251	8,458	479,968
New Zealand	523	58,683	134	6,051	59,650
Norway (9)	1,012	60,847	157	4,575	62,287
Oman	784	54,282	228	6,751	54,066
Portugal	1,039	108,807	156	3,377	106,814
Qatar	156	19,513	152	3,884	18,715
Romania	5,697	182,020	198	4,494	183,845
Russian Federation	37,308	1,326,933	204	3,901	1,392,266
Saudi Arabia	7,248	397,795	209	5,680	390,646
Singapore	153	38,517	153	4,853	38,595

Exhibit 9.6: Population and Sample Sizes – TIMSS 2019 – Eighth Grade (continued)

Country	Population		Sample		
	Schools	Students	Schools	Students	Student Population Size Estimated from Sample
South Africa (9)	8,340	887,952	519	20,829	877,201
Sweden	1,600	108,164	150	3,996	110,810
Turkey	16,179	1,204,063	181	4,077	1,158,547
United Arab Emirates	685	68,113	623	22,334	68,388
United States	48,557	4,059,757	273	8,698	3,799,856
Benchmarking Participants					
Ontario, Canada	2,896	143,484	158	3,776	140,990
Quebec, Canada	539	80,005	124	3,178	75,411
Moscow City, Russian Fed.	704	85,856	150	3,783	92,180
Gauteng, RSA (9)	988	167,128	150	5,633	170,315
Western Cape, RSA (9)	498	75,596	149	5,351	77,855
Abu Dhabi, UAE	266	24,654	230	8,204	23,805
Dubai, UAE	153	17,560	163	5,728	18,752

Stratification

TIMSS 2019 National Research Coordinators consulted with Statistics Canada and IEA Hamburg to identify the stratification variables to be included in their sampling plans. Exhibits 9.7 and 9.8 provide the list of explicit and implicit stratification variables implemented by the countries participating at the fourth grade at the eighth grade respectively.

Exhibit 9.7: Stratification Variables – TIMSS 2019 – Fourth Grade

Country	Explicit Stratification Variables	Number of Explicit Strata	Implicit Stratification Variables
Albania	School type (2) Urbanization (2)	3	Urbanization (2)
Armenia	Region (10)	10	None
Australia	State or territory (8)	8	School type (3) Geographic location (3) Socioeconomic status (2)
Austria	Urbanization (2) Achievement (3) School size (2)	12	Region (9)

Exhibit 9.7: Stratification Variables – TIMSS 2019 – Fourth Grade (continued)

Country	Explicit Stratification Variables	Number of Explicit Strata	Implicit Stratification Variables
Azerbaijan	Language (2) Urbanization (2) City (2)	4	None
Bahrain	School type (2) Governorate (4) Gender (2)	9	None
Belgium (Flemish)	Region (6) School type (2) Socioeconomic status (4)	20	None
Bosnia and Herzegovina	Region (12) Urbanization (2)	8	Urbanization (2)
Bulgaria	School type (3) Urbanization (3)	8	Score (3)
Canada	Province (5) Language (2) School type (2 or 3) School size (2)	25	Region (6)
Chile	Grade 4 / grade 4 and 8 schools (2) School type (3) Urbanization (2)	7	National assessment score level (4)
Chinese Taipei	Urbanization (4) Region (2) School size (2)	11	None
Croatia	Region (6) School type (2) School size (2)	13	Urbanization (2)
Cyprus	School type (2) Curriculum (2) District (4)	5	Urbanization (2)
Czech Republic	Region (14)	15	None
Denmark	School type (2) School size (2)	3	None
England	School type (3) Attainment level (5)	9	Attainment level (7)
Finland	Language (2) Major region (4) Urbanization (2)	8	Regional state administrative agency (6)
France	School type (3)	3	None
Georgia	Grade 4 only / grade 4 and 8 schools (2) Region (2) Math average score (3)	6	Urbanization (2) School type (2)

Exhibit 9.7: Stratification Variables – TIMSS 2019 – Fourth Grade (continued)

Country	Explicit Stratification Variables	Number of Explicit Strata	Implicit Stratification Variables
Germany	School type (2) Socioeconomic status (3) school size (2)	8	None
Hong Kong SAR	School finance type (5)	5	None
Hungary	Grade 4 only / grade 4 and 8 schools (2) Type of community (4) National assessment score (3)	11	None
Iran, Islamic Rep. of	School type (2) Gender (3) Province or grouped provinces (7)	16	None
Ireland	School level socioeconomic status DEIS (3) Language of instruction (3) Gender (3)	8	Location (2)
Italy	Grade 4 only / grade 4 and 8 schools (2) School type (2) Region (5)	8	Region (5)
Japan	School location (4)	4	None
Kazakhstan	Grade 4 only / grade 4 and 8 schools (2) Region (4) Urbanization (2) Language (2)	18	None
Korea, Rep. of	Urbanization (3) School size (2)	8	None
Kosovo	Urbanization (2) Shifts (2)	4	None
Kuwait	Grade 4 only / grade 4 and 8 schools (2) School type (2) Region (6) Gender (2) Language (3)	15	None
Latvia	Urbanization (3) Language (2) School type (2)	7	None
Lithuania	Grade 4 / grade 4 and 8 schools (2) Languages (5)	6	Urbanization (4) School type (4)
Malta	School type (3)	3	None
Montenegro	Region (3)	3	Urbanization (2)

Exhibit 9.7: Stratification Variables – TIMSS 2019 – Fourth Grade (continued)

Country	Explicit Stratification Variables	Number of Explicit Strata	Implicit Stratification Variables
Morocco	School type (2) Region (12)	14	Urbanization (2)
Netherlands	Socioeconomic status (3)	3	None
New Zealand	None	1	School type (2) Socioeconomic status (4) Urbanization (2)
North Macedonia	Urbanization (3) Language (3)	8	None
Northern Ireland	Region (5) Deprivation group (9)	14	None
Norway (5)	Grade 5 only / grade 5 and 9 schools (2) City (2) Municipality size (3)	8	National numeracy test score (4)
Oman	Governorates (11) School type (2)	13	None
Pakistan	School type (2) Region (5)	6	Region (7) Urbanization (2) Gender (2)
Philippines	School type (2) Socioeconomic index (3) Geographic location (2) Unknown (1)	10	None
Poland	Urbanization (4)	4	None
Portugal	School type (2) Region (8) School size (2)	10	NUTS 3 region (25) NUTS 2 region (8)
Qatar	Grade 4 only / grade 4 and 8 (2) Gender (3)	4	Gender (3) School type (4)
Russian Federation	Region (43)	43	None
Saudi Arabia	School type (3) Gender (2)	6	None
Serbia	Region (3) Urbanization (2) School hierarchy (2)	7	None
Singapore	None	1	None
Slovak Republic	Language (2) National testing score (4) School size (2)	8	None
South Africa (5)	School type (2) Province (9)	10	Performance level (5) Province (5)

Exhibit 9.7: Stratification Variables – TIMSS 2019 – Fourth Grade (continued)

Country	Explicit Stratification Variables	Number of Explicit Strata	Implicit Stratification Variables
Spain	Region (9) School type (2) School funding (2) Bilingual status (2)	19	Region (12) School type (2)
Sweden	Average achievement (4) School type (2)	6	None
Turkey (5)	Grade 5 only / grade 5 and 8 School type (2) Region (13) School size (2)	25	None
United Arab Emirates	Grade 4 only / grade 4 and 8 schools (2) Emirate (3) School type (2) Main curriculum (2)	18	School size (2) Region (5) Language of test (3) Curriculum (3)
United States	Poverty level (2) School type (2) Census region of public school (4) Type of private school (2)	10	Urbanization (4) Ethnicity status (2) State (52)
Benchmarking Participants			
Ontario, Canada	Language (2) School type (3) School size (2)	6	Regional office (6)
Quebec, Canada	Language (2) School type (2) School size (2)	7	None
Moscow City, Russian Fed.	Grade 4 / grade 4 and 8 schools (2) School type (2)	3	School size (3)
Madrid, Spain	School type (3) Bilingual status (2)	5	None
Abu Dhabi, UAE	Grade 4 only / grade 4 and 8 schools (2) School type (2) Main curriculum (3)	7	School size (2) Region (3)
Dubai, UAE	Grade 4 only / grade 4 and 8 schools (2) School type (2)	4	School size (2) Language of test (3)

Exhibit 9.8: Stratification Variables – TIMSS 2019 – Eighth Grade

Country	Explicit Stratification Variables	Number of Explicit Strata	Implicit Stratification Variables
Australia	State or territory (8)	8	School type (3) Geographic location (3) Socioeconomic status (2)
Bahrain	School type (2) Governorate (4) Gender (2)	9	None
Chile	Grade 8 / grade 4 and 8 schools (2) School type (3) Urbanization (2)	6	National assessment score level (4)
Chinese Taipei	Urbanization (4) Region (2) School size (2)	9	Performance (5)
Cyprus	School type (2) Curriculum (2) District (4)	5	Urbanization (2)
Egypt	Region (3) School type (4) Gender schools (3)	12	School shift (4)
England	School type (3) Attainment level (5)	9	Attainment level (7)
Finland	Language (2) Major region (4) Urbanization (2)	8	Regional state administrative agency (6)
France	School type (3)	3	None
Georgia	Grade 4 only / grade 4 and 8 schools (2) Region (2) Math average score (3)	6	Urbanization (2) School type (2)
Hong Kong SAR	School finance type (4)	4	Other school characteristic (3)
Hungary	Grade 8 only / grade 4 and 8 schools (2) Type of community (4) National assessment score (3)	11	None
Iran, Islamic Rep. of	School type (2) Gender (3) Province or grouped provinces (7)	16	None
Ireland	School sector (3) Socioeconomic status (3) Gender (3)	13	None
Israel	School sector (3) Socioeconomic status (3) Subgroups within Arab sector (3) School size (2)	11	Gender (3) Region (3)

Exhibit 9.8: Stratification Variables – TIMSS 2019 – Eighth Grade (continued)

Country	Explicit Stratification Variables	Number of Explicit Strata	Implicit Stratification Variables
Italy	Grade 8 only / grade 4 and 8 schools (2) School type (2) Region (5)	8	Region (5)
Japan	School type (2) school location (4)	5	None
Jordan	School type (6) Achievement level (4)	24	Region (4)
Kazakhstan	Grade 8 only / grade 4 and 8 schools (2) Region (4) Urbanization (2) Language (2)	18	None
Korea, Rep. of	Urbanization (3) School gender (3) School size (2)	8	None
Kuwait	Grade 8 only / grade 4 and 8 schools (2) School type (2) Region (6) Gender (2) Language (3)	15	None
Lebanon	Regions or grouped regions (6) school type (2) school size (2)	24	None
Lithuania	Grade 8 / grade 4 and 8 schools (2) Languages (5)	6	Urbanization (4) School type (4)
Malaysia	School type (6) Score level (3) Urbanization (2)	12	None
Morocco	School type (2) Region (12)	14	Urbanization (2)
New Zealand	School type (2) Socioeconomic status (4) Urbanization (2)	9	Gender (3)
Norway (9)	Grade 8 only / grade 5 and 9 schools (2) City (2) Municipality size (3)	8	National numeracy test score (4)
Oman	Governorates (11) School type (2)	13	Gender (3)
Portugal	School type (2) Region (8)	9	NUTS 3 region (25) NUTS 2 region (5)

Exhibit 9.8: **Stratification Variables – TIMSS 2019 – Eighth Grade (continued)**

Country	Explicit Stratification Variables	Number of Explicit Strata	Implicit Stratification Variables
Qatar	Grade 4 only / grade 4 and 8 (2)	2	Gender (3) School type (4)
Romania	Urbanization (2) Region (5)	10	None
Russian Federation	Region (43)	43	None
Saudi Arabia	School type (3) Gender (2)	6	None
Singapore	None	1	None
South Africa (9)	School type (2) Province (9)	10	Performance level (5) Province (3)
Sweden	Average achievement (4) School type (2)	6	None
Turkey	Grade 8 only / grade 5 and 8 School type (2) Region (13) School size (2)	25	None
United Arab Emirates	Grade 8 only / grade 4 and 8 schools (2) Emirate (3) School type (2) Main curriculum (2)	14	School size (2) Region (5) Language of test (3) Curriculum (3)
United States	Poverty level (2) School type (2) Census region of public school (4) Type of private school (2)	10	Urbanization (4) Ethnicity status (2) State (52)
Benchmarking Participants			
Ontario, Canada	Language (2) School type (3) School size (2)	5	Regional office (6)
Quebec, Canada	Language (2) School type (2) School size (2)	5	Mathematics average score (4) Program (2)
Moscow City, Russian Fed.	Grade 8 / grade 4 and 8 schools (2) School type (2)	3	School size (3)
Gauteng, RSA (9)	School type (2)	2	Performance level (6)
Western Cape, RSA (9)	School type (2)	2	Performance level (6)
Abu Dhabi, UAE	School type (2) Main curriculum (3)	4	School size (2) Region (3)
Dubai, UAE	Grade 8 only / grade 4 and 8 schools (2) School type (2)	4	School size (2) Language of test (3)

Meeting TIMSS 2019 Standards for Sampling Participation

TIMSS 2019 participants understood that the goal for sampling participation was 100 percent for all sampled schools, classrooms, and students. Guidelines for reporting achievement data for participants securing less than full participation were modeled after IEA's previous TIMSS assessment cycles. As summarized below in Exhibit 9.9, countries were assigned to one of three categories on the basis of their sampling participation. Countries in Category 1 were considered to have met all TIMSS 2019 sampling requirements and to have acceptable participation rates. Countries in Category 2 met the participation requirements only after including replacement schools. Countries that failed to meet the participation requirements even with the use of replacement schools were assigned to Category 3. One of the main goals for quality data in TIMSS 2019 was to have as many countries as possible achieve Category 1 status.

Exhibit 9.9: Categories of Sampling Participation

Category 1	Acceptable sampling participation rate **without** the use of replacement schools. In order to be placed in this category, a country had to have: • An **unweighted** school response rate **without** replacement of at least 85% (after rounding to nearest whole percent) AND an **unweighted** student response rate (after rounding) of at least 85% OR • A **weighted** school response rate **without** replacement of at least 85% (after rounding to nearest whole percent) AND a **weighted** student response rate (after rounding) of at least 85% OR • The product of the (unrounded) weighted school response rate without replacement and the (unrounded) weighted student response rate of at least 75% (after rounding to the nearest whole percent). Countries in this category would appear in the tables and figures in international reports without annotation, and will be ordered by achievement as appropriate.
Category 2	Acceptable sampling participation rate **only when replacement schools are included**. A country would be placed in this category 2 if: • It failed to meet the requirements for Category 1 but had a weighted school response rate **without** replacement of at least 50% (after rounding to the nearest percent) AND HAD EITHER • A **weighted** school response rate **with** replacement of at least 85% (after rounding to nearest whole percent) AND a **weighted** student response rate (after rounding) of at least 85% OR • The product of the (unrounded) weighted school response rate with replacement and the (unrounded) weighted student response rate of at least 75% (after rounding to the nearest whole percent). Countries in this category would be annotated with † in the tables and figures in international reports, and ordered by achievement as appropriate.

Exhibit 9.9: Categories of Sampling Participation (continued)

Category 3	Unacceptable sampling response rate even when replacement schools are included. Countries that could provide documentation to show that they complied with PIRLS sampling procedures and requirements but did not meet the requirements for Category 1 or Category 2 would be placed in Category 3. Countries in this category would be annotated with ‡ if they nearly met the requirements for Category 2. Countries would be annotated with ≡ if they failed to meet the participation requirements but had a school participation rate of at least 50% before the use of replacement schools. At last, if none of these conditions are met, countries would appear in a separate section of the achievement tables, below the other countries, in international reports. These countries would be presented in alphabetical order.

Participation Rates of the TIMSS 2019 National Samples

Exhibits 9.10 through 9.13 present the school, classroom, student, and overall weighted and unweighted participation rates for each of the participants in the TIMSS 2019 fourth and eighth grade assessments, respectively. Almost all participants had excellent participation rates and belonged in Category 1. At the fourth grade, Belgium (Flemish), Denmark, Hong Kong SAR, Northern Ireland, Norway (5), and the United States achieved the minimum acceptable participation rate only after including replacement schools, and therefore their results were annotated with a dagger (†) in the achievement exhibits of the international reports (Category 2). Despite efforts to secure full participation, Netherlands did not meet the required sampling participation rate even with the use of replacement schools and were annotated with a triple-dagger (≡) in the achievement exhibits of the international reports.

At the eighth grade, Hong Kong SAR, New Zealand, Norway (9), the United States achieved the minimum acceptable participation rates only after including replacement schools, and therefore their results were annotated with a dagger (†) in the achievement exhibits of the international reports (Category 2). Finally, the benchmarking participant of Quebec, Canada, nearly met the required sampling participation rate at the fourth and eighth grades with the use of replacement schools and were annotated with a double-dagger (‡) in the achievement exhibits of the international reports (Category 3).

Exhibit 9.10: Participation Rates (Weighted) – TIMSS 2019 – Fourth Grade

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Albania	99%	99%	100%	99%	98%	98%
Armenia	100%	100%	100%	97%	97%	97%
Australia	99%	100%	100%	94%	93%	94%
Austria	99%	99%	100%	97%	97%	97%

Exhibit 9.10: **Participation Rates (Weighted) – TIMSS 2019 – Fourth Grade (continued)**

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Azerbaijan	94%	98%	100%	95%	89%	92%
Bahrain	100%	100%	100%	98%	98%	98%
† Belgium (Flemish)	66%	95%	100%	93%	62%	89%
Bosnia and Herzegovina	100%	100%	99%	96%	95%	95%
Bulgaria	97%	100%	100%	95%	92%	95%
Canada	86%	90%	100%	95%	82%	86%
Chile	89%	99%	100%	96%	86%	95%
Chinese Taipei	95%	99%	100%	99%	94%	98%
Croatia	95%	97%	99%	91%	85%	87%
Cyprus	99%	100%	100%	97%	96%	97%
Czech Republic	99%	100%	100%	96%	95%	96%
† Denmark	70%	95%	99%	87%	61%	83%
England	86%	93%	100%	96%	82%	89%
Finland	99%	100%	100%	97%	96%	97%
France	100%	100%	100%	98%	98%	98%
Georgia	97%	99%	100%	97%	94%	96%
Germany	97%	100%	100%	97%	94%	97%
† Hong Kong SAR	67%	88%	100%	90%	60%	79%
Hungary	93%	99%	100%	97%	90%	96%
Iran, Islamic Rep. of	100%	100%	100%	99%	99%	99%
Ireland	100%	100%	100%	91%	91%	91%
Italy	96%	100%	100%	97%	92%	97%
Japan	84%	98%	100%	97%	82%	95%
Kazakhstan	100%	100%	100%	99%	99%	99%
Korea, Rep. of	99%	99%	100%	98%	97%	97%
Kosovo	100%	100%	100%	97%	97%	97%
Kuwait	97%	98%	100%	96%	93%	94%
Latvia	92%	99%	100%	94%	87%	93%
Lithuania	100%	100%	100%	94%	94%	94%
Malta	100%	100%	100%	96%	96%	96%
Montenegro	100%	100%	100%	98%	98%	98%
Morocco	100%	100%	100%	99%	99%	99%
≡ Netherlands	46%	75%	100%	97%	45%	73%
New Zealand	87%	99%	100%	94%	81%	93%

Exhibit 9.10: Participation Rates (Weighted) – TIMSS 2019 – Fourth Grade (continued)

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
North Macedonia	98%	100%	100%	95%	94%	95%
† Northern Ireland	60%	86%	100%	91%	55%	78%
† Norway (5)	70%	90%	100%	94%	66%	84%
Oman	99%	100%	100%	98%	98%	98%
Pakistan	77%	99%	100%	98%	75%	96%
Philippines	100%	100%	100%	98%	98%	98%
Poland	96%	100%	100%	93%	89%	93%
Portugal	87%	100%	99%	94%	81%	94%
Qatar	100%	100%	100%	97%	97%	97%
Russian Federation	99%	99%	100%	98%	97%	97%
Saudi Arabia	98%	99%	100%	99%	97%	98%
Serbia	97%	100%	100%	97%	95%	97%
Singapore	100%	100%	100%	97%	97%	97%
Slovak Republic	97%	99%	100%	97%	93%	96%
South Africa (5)	96%	99%	100%	98%	94%	97%
Spain	97%	99%	100%	95%	92%	95%
Sweden	100%	100%	100%	95%	95%	95%
Turkey (5)	99%	100%	100%	99%	98%	99%
United Arab Emirates	100%	100%	100%	96%	96%	96%
† United States	76%	88%	100%	96%	73%	84%
Benchmarking Participants						
Ontario, Canada	93%	95%	100%	95%	88%	90%
Quebec, Canada	82%	86%	100%	96%	79%	83%
Moscow City, Russian Fed.	99%	100%	100%	98%	97%	97%
Madrid, Spain	100%	100%	100%	96%	96%	96%
Abu Dhabi, UAE	100%	100%	100%	95%	95%	95%
Dubai, UAE	100%	100%	100%	97%	97%	97%

TIMSS guidelines for sampling participation: The minimum acceptable participation rates were 85 percent of both schools and students, or a combined rate (the product of school and student participation) of 75 percent.

Participants not meeting these guidelines were annotated as follows:

† Met guidelines for sample participation rates only after replacement schools were included

‡ Nearly satisfied guidelines for sample participation rates after replacement schools were included

≡ Did not satisfy guidelines for sample participation rates

Exhibit 9.11: **Participation Rates (Weighted) – TIMSS 2019 – Eighth Grade**

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Australia	98%	100%	100%	91%	89%	91%
Bahrain	100%	100%	100%	97%	97%	97%
Chile	90%	99%	100%	96%	86%	95%
Chinese Taipei	98%	99%	100%	98%	96%	97%
Cyprus	100%	100%	100%	96%	96%	96%
Egypt	99%	100%	100%	97%	96%	97%
England	83%	90%	100%	95%	79%	85%
Finland	100%	100%	100%	96%	95%	95%
France	100%	100%	100%	97%	97%	97%
Georgia	90%	92%	100%	97%	88%	89%
† Hong Kong SAR	70%	86%	100%	94%	66%	81%
Hungary	95%	99%	100%	97%	92%	96%
Iran, Islamic Rep. of	100%	100%	100%	98%	98%	98%
Ireland	97%	98%	100%	88%	85%	86%
Israel	95%	98%	100%	93%	88%	91%
Italy	97%	100%	100%	97%	94%	97%
Japan	83%	94%	100%	94%	77%	88%
Jordan	100%	100%	100%	98%	98%	98%
Kazakhstan	100%	100%	100%	99%	99%	99%
Korea, Rep. of	100%	100%	100%	98%	98%	98%
Kuwait	99%	99%	100%	97%	96%	96%
Lebanon	82%	93%	100%	95%	78%	88%
Lithuania	99%	99%	100%	93%	92%	92%
Malaysia	99%	100%	100%	98%	97%	98%
Morocco	100%	100%	100%	98%	98%	98%
† New Zealand	77%	89%	100%	91%	70%	81%
† Norway (9)	79%	95%	99%	89%	70%	84%
Oman	99%	100%	100%	99%	97%	99%
Portugal	95%	99%	99%	96%	90%	94%
Qatar	100%	100%	100%	97%	97%	97%
Romania	95%	100%	100%	94%	89%	94%
Russian Federation	99%	100%	100%	97%	97%	97%
Saudi Arabia	100%	100%	100%	99%	99%	99%
Singapore	100%	100%	100%	96%	96%	96%
South Africa (9)	99%	100%	100%	96%	95%	96%

Exhibit 9.11: Participation Rates (Weighted) – TIMSS 2019 – Eighth Grade (continued)

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Sweden	98%	99%	100%	92%	90%	91%
Turkey	100%	100%	100%	99%	98%	99%
United Arab Emirates	100%	100%	100%	96%	96%	96%
† United States	72%	85%	100%	94%	67%	79%
Benchmarking Participants						
Ontario, Canada	93%	93%	100%	94%	87%	88%
‡ Quebec, Canada	74%	77%	99%	95%	70%	73%
Moscow City, Russian Fed.	99%	100%	100%	97%	96%	97%
Gauteng, RSA (9)	99%	100%	100%	97%	95%	97%
Western Cape, RSA (9)	99%	100%	100%	95%	95%	95%
Abu Dhabi, UAE	100%	100%	100%	96%	96%	96%
Dubai, UAE	100%	100%	100%	96%	96%	96%

TIMSS guidelines for sampling participation: The minimum acceptable participation rates were 85 percent of both schools and students, or a combined rate (the product of school and student participation) of 75 percent.

Participants not meeting these guidelines were annotated as follows:

† Met guidelines for sample participation rates only after replacement schools were included

‡ Nearly satisfied guidelines for sample participation rates after replacement schools were included

≡ Did not satisfy guidelines for sample participation rates

Exhibit 9.12: **Participation Rates (Unweighted) – TIMSS 2019 – Fourth Grade**

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Albania	99%	99%	100%	99%	97%	97%
Armenia	100%	100%	100%	97%	97%	97%
Australia	99%	100%	100%	94%	93%	93%
Austria	99%	99%	100%	97%	96%	96%
Azerbaijan	93%	97%	99%	94%	87%	91%
Bahrain	100%	100%	100%	98%	98%	98%
Belgium (Flemish)	65%	94%	100%	94%	61%	88%
Bosnia and Herzegovina	100%	100%	99%	95%	95%	95%
Bulgaria	97%	100%	100%	95%	92%	95%
Canada	86%	91%	100%	94%	81%	85%
Chile	88%	98%	100%	95%	83%	93%
Chinese Taipei	95%	99%	100%	98%	94%	98%
Croatia	95%	97%	98%	89%	83%	85%
Cyprus	99%	100%	100%	97%	96%	97%
Czech Republic	99%	100%	100%	95%	94%	95%
Denmark	71%	95%	99%	86%	60%	81%
England	86%	93%	99%	96%	82%	88%
Finland	99%	100%	100%	96%	96%	96%
France	100%	100%	100%	97%	97%	97%
Georgia	96%	98%	100%	97%	93%	95%
Germany	98%	100%	100%	96%	94%	96%
Hong Kong SAR	69%	87%	100%	89%	61%	78%
Hungary	93%	99%	100%	96%	89%	96%
Iran, Islamic Rep. of	100%	100%	100%	99%	99%	99%
Ireland	100%	100%	100%	91%	91%	91%
Italy	94%	100%	100%	96%	91%	96%
Japan	84%	98%	100%	97%	82%	95%
Kazakhstan	100%	100%	100%	99%	99%	99%
Korea, Rep. of	99%	99%	100%	98%	97%	97%
Kosovo	100%	100%	100%	97%	97%	97%
Kuwait	98%	98%	100%	96%	93%	94%
Latvia	91%	99%	100%	93%	85%	92%
Lithuania	100%	100%	100%	94%	94%	94%
Malta	100%	100%	100%	96%	96%	96%
Montenegro	100%	100%	100%	98%	98%	98%
Morocco	100%	100%	100%	99%	99%	99%

Exhibit 9.12: **Participation Rates (Unweighted) – TIMSS 2019 – Fourth Grade (continued)**

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Netherlands	48%	75%	100%	97%	46%	73%
New Zealand	86%	99%	100%	94%	80%	93%
North Macedonia	97%	100%	100%	95%	92%	95%
Northern Ireland	61%	86%	99%	91%	55%	78%
Norway (5)	71%	90%	100%	93%	66%	83%
Oman	99%	100%	100%	98%	97%	98%
Pakistan	85%	98%	100%	96%	82%	94%
Philippines	100%	100%	100%	98%	98%	98%
Poland	96%	100%	100%	92%	88%	92%
Portugal	87%	100%	99%	93%	81%	93%
Qatar	100%	100%	100%	97%	97%	97%
Russian Federation	99%	99%	100%	97%	96%	96%
Saudi Arabia	97%	100%	100%	98%	96%	98%
Serbia	96%	100%	100%	97%	93%	97%
Singapore	100%	100%	100%	97%	97%	97%
Slovak Republic	97%	99%	100%	96%	93%	95%
South Africa (5)	96%	100%	100%	98%	94%	97%
Spain	98%	100%	100%	95%	94%	95%
Sweden	99%	100%	100%	94%	93%	94%
Turkey (5)	99%	100%	100%	98%	98%	98%
United Arab Emirates	100%	100%	100%	96%	95%	95%
United States	77%	88%	100%	95%	73%	84%
Benchmarking Participants						
Ontario, Canada	94%	96%	100%	94%	89%	90%
Quebec, Canada	81%	86%	100%	96%	78%	83%
Moscow City, Russian Fed.	98%	99%	100%	97%	95%	97%
Madrid, Spain	100%	100%	100%	96%	96%	96%
Abu Dhabi, UAE	100%	100%	100%	95%	95%	95%
Dubai, UAE	100%	100%	100%	96%	96%	96%

Exhibit 9.13: Participation Rates (Unweighted) – TIMSS 2019 – Eighth Grade

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Australia	99%	100%	100%	91%	90%	91%
Bahrain	100%	100%	100%	97%	97%	97%
Chile	88%	98%	100%	95%	84%	93%
Chinese Taipei	98%	99%	100%	98%	95%	97%
Cyprus	100%	100%	100%	96%	96%	96%
Egypt	99%	100%	100%	96%	96%	96%
England	83%	90%	100%	94%	78%	85%
Finland	100%	100%	100%	95%	95%	95%
France	100%	100%	100%	96%	96%	96%
Georgia	90%	92%	100%	97%	87%	89%
Hong Kong SAR	71%	86%	100%	93%	66%	80%
Hungary	94%	99%	100%	96%	91%	96%
Iran, Islamic Rep. of	100%	100%	100%	98%	98%	98%
Ireland	97%	98%	100%	88%	85%	86%
Israel	94%	98%	100%	92%	87%	89%
Italy	97%	100%	100%	97%	94%	97%
Japan	83%	95%	100%	94%	78%	89%
Jordan	100%	100%	100%	98%	98%	98%
Kazakhstan	100%	100%	100%	98%	98%	98%
Korea, Rep. of	100%	100%	100%	97%	97%	97%
Kuwait	99%	99%	100%	97%	96%	96%
Lebanon	88%	94%	100%	95%	83%	90%
Lithuania	99%	99%	100%	92%	92%	92%
Malaysia	99%	100%	100%	98%	97%	98%
Morocco	100%	100%	100%	97%	97%	97%
New Zealand	76%	89%	100%	92%	70%	81%
Norway (9)	80%	95%	98%	89%	70%	83%
Oman	98%	100%	100%	98%	96%	98%
Portugal	94%	99%	99%	95%	88%	92%
Qatar	100%	100%	100%	96%	96%	96%
Romania	95%	100%	100%	94%	90%	94%
Russian Federation	100%	100%	100%	97%	97%	97%
Saudi Arabia	100%	100%	100%	99%	99%	99%
Singapore	100%	100%	100%	96%	96%	96%
South Africa (9)	99%	100%	100%	96%	95%	96%
Sweden	99%	99%	100%	91%	89%	90%

Exhibit 9.13: Participation Rates (Unweighted) – TIMSS 2019 – Eighth Grade (continued)

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Turkey	99%	100%	100%	98%	98%	98%
United Arab Emirates	100%	100%	100%	95%	95%	95%
United States	72%	85%	100%	93%	67%	79%
Benchmarking Participants						
Ontario, Canada	92%	93%	100%	93%	86%	87%
Quebec, Canada	74%	77%	99%	94%	69%	72%
Moscow City, Russian Fed.	97%	99%	100%	96%	94%	96%
Gauteng, RSA (9)	99%	100%	100%	97%	95%	97%
Western Cape, RSA (9)	99%	100%	100%	95%	95%	95%
Abu Dhabi, UAE	100%	100%	100%	95%	95%	95%
Dubai, UAE	100%	100%	100%	96%	96%	96%

TIMSS 2019 National Samples – Achieved Sample Sizes

Exhibits 9.14 through 9.17 show the achieved sample sizes in terms of schools and students for each of the participants in the TIMSS 2019 fourth and eighth grade assessments, respectively.

Exhibit 9.14: School Sample Sizes – TIMSS 2019 – Fourth Grade

Country	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample that Participated	Number of Replacement Schools that Participated	Total Number of Schools that Participated
Albania	180	169	167	0	167
Armenia	150	150	150	0	150
Australia	290	288	285	2	287
Austria	197	194	193	0	193
Azerbaijan	200	199	186	8	194
Bahrain	185	185	185	0	185
Belgium (Flemish)	160	156	101	46	147
Bosnia and Herzegovina	178	178	178	0	178
Bulgaria	151	151	146	5	151
Canada	788	777	669	35	704
Chile	174	172	151	18	169
Chinese Taipei	163	163	155	7	162

Exhibit 9.14: School Sample Sizes – TIMSS 2019 – Fourth Grade (continued)

Country	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample that Participated	Number of Replacement Schools that Participated	Total Number of Schools that Participated
Croatia	159	158	150	3	153
Cyprus	152	151	150	1	151
Czech Republic	156	152	151	1	152
Denmark	175	174	123	43	166
England	150	150	129	10	139
Finland	159	158	157	1	158
France	156	155	155	0	155
Georgia	158	157	151	3	154
Germany	206	203	198	5	203
Hong Kong SAR	159	159	109	30	139
Hungary	151	150	139	10	149
Iran, Islamic Rep. of	224	224	224	0	224
Ireland	151	150	150	0	150
Italy	162	162	153	9	162
Japan	150	150	126	21	147
Kazakhstan	169	168	168	0	168
Korea, Rep. of	152	152	151	0	151
Kosovo	150	145	145	0	145
Kuwait	170	167	163	1	164
Latvia	156	156	142	12	154
Lithuania	208	207	207	0	207
Malta	99	98	98	0	98
Montenegro	140	140	140	0	140
Morocco	265	264	264	0	264
Netherlands	151	149	71	41	112
New Zealand	164	161	138	22	160
North Macedonia	150	150	146	4	150
Northern Ireland	156	156	95	39	134
Norway (5)	167	167	119	31	150
Oman	228	228	226	2	228
Pakistan	150	142	121	18	139
Philippines	184	180	180	0	180
Poland	150	149	143	6	149
Portugal	182	181	158	23	181
Qatar	242	242	242	0	242

Exhibit 9.14: School Sample Sizes – TIMSS 2019 – Fourth Grade (continued)

Country	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample that Participated	Number of Replacement Schools that Participated	Total Number of Schools that Participated
Russian Federation	202	202	200	0	200
Saudi Arabia	222	221	215	5	220
Serbia	170	165	159	6	165
Singapore	187	187	187	0	187
Slovak Republic	159	158	153	4	157
South Africa (5)	300	298	286	11	297
Spain	502	502	494	7	501
Sweden	150	145	144	1	145
Turkey (5)	181	180	179	1	180
United Arab Emirates	697	688	688	0	688
United States	329	325	249	38	287
Benchmarking Participants					
Ontario, Canada	171	170	160	3	163
Quebec, Canada	172	172	140	8	148
Moscow City, Russian Fed.	152	151	148	2	150
Madrid, Spain	167	167	167	0	167
Abu Dhabi, UAE	249	247	247	0	247
Dubai, UAE	205	199	199	0	199

Exhibit 9.15: School Sample Sizes – TIMSS 2019 – Eighth Grade

Country	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample that Participated	Number of Replacement Schools that Participated	Total Number of Schools that Participated
Australia	289	284	282	2	284
Bahrain	112	112	112	0	112
Chile	169	167	147	17	164
Chinese Taipei	206	205	200	3	203
Cyprus	99	98	98	0	98
Egypt	174	169	168	1	169
England	151	151	125	11	136
Finland	158	154	154	0	154
France	150	150	150	0	150
Georgia	158	157	142	3	145
Hong Kong SAR	158	158	112	24	136
Hungary	155	155	146	8	154
Iran, Islamic Rep. of	220	220	220	0	220
Ireland	152	152	147	2	149
Israel	161	161	152	5	157
Italy	158	158	153	5	158
Japan	150	150	125	17	142
Jordan	248	235	235	0	235
Kazakhstan	169	168	168	0	168
Korea, Rep. of	168	168	168	0	168
Kuwait	178	172	171	0	171
Lebanon	218	216	189	15	204
Lithuania	195	195	194	0	194
Malaysia	178	177	175	2	177
Morocco	253	251	251	0	251
New Zealand	154	151	115	19	134
Norway (9)	166	165	132	25	157
Oman	230	228	223	5	228
Portugal	158	158	149	7	156
Qatar	152	152	152	0	152
Romania	198	198	189	9	198
Russian Federation	204	204	203	1	204
Saudi Arabia	212	209	208	1	209
Singapore	153	153	153	0	153

Exhibit 9.15: School Sample Sizes – TIMSS 2019 – Eighth Grade (continued)

Country	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample that Participated	Number of Replacement Schools that Participated	Total Number of Schools that Participated
South Africa (9)	524	520	516	3	519
Sweden	153	151	149	1	150
Turkey	181	181	180	1	181
United Arab Emirates	631	623	623	0	623
United States	325	321	231	42	273
Benchmarking Participants					
Ontario, Canada	172	170	157	1	158
Quebec, Canada	166	161	119	5	124
Moscow City, Russian Fed.	152	151	147	3	150
Gauteng, RSA (9)	150	150	148	2	150
Western Cape, RSA (9)	150	149	148	1	149
Abu Dhabi, UAE	230	230	230	0	230
Dubai, UAE	171	163	163	0	163

Exhibit 9.16: Student Sample Sizes – TIMSS 2019 – Fourth Grade

Country	Within-School Student Participation (Weighted Percentage)	Number of Students Sampled in Participating Schools	Number of Students Withdrawn from Class/ School	Number of Students Excluded	Number of Students Eligible	Number of Students Absent	Number of Students Assessed
Albania	99%	4,548	31	25	4,492	66	4,426
Armenia	97%	5,612	32	0	5,580	181	5,399
Australia	94%	6,517	110	128	6,279	389	5,890
Austria	97%	4,901	33	256	4,612	148	4,464
Azerbaijan	95%	5,600	17	19	5,564	319	5,245
Bahrain	98%	5,903	25	22	5,856	94	5,762
Belgium (Flemish)	93%	5,113	26	114	4,973	318	4,655
Bosnia and Herzegovina	96%	6,048	61	74	5,913	285	5,628
Bulgaria	95%	4,632	70	88	4,474	206	4,268
Canada	95%	15,164	199	429	14,536	883	13,653
Chile	96%	4,578	77	112	4,389	215	4,174
Chinese Taipei	99%	3,958	65	65	3,828	63	3,765
Croatia	91%	4,395	8	148	4,239	454	3,785
Cyprus	97%	4,353	14	150	4,189	127	4,062
Czech Republic	96%	5,054	48	53	4,953	261	4,692
Denmark	87%	3,881	67	48	3,766	539	3,227
England	96%	3,759	78	127	3,554	158	3,396
Finland	97%	4,987	37	45	4,905	175	4,730
France	98%	4,456	35	104	4,317	131	4,186
Georgia	97%	4,019	28	83	3,908	121	3,787
Germany	97%	3,706	51	89	3,566	129	3,437
Hong Kong SAR	90%	3,461	18	101	3,342	374	2,968
Hungary	97%	4,867	34	89	4,744	173	4,571
Iran, Islamic Rep. of	99%	6,194	46	76	6,072	62	6,010
Ireland	91%	5,126	22	52	5,052	470	4,582
Italy	97%	4,109	22	199	3,888	147	3,741
Japan	97%	4,358	15	34	4,309	113	4,196
Kazakhstan	99%	4,932	37	38	4,857	66	4,791
Korea, Rep. of	98%	4,105	50	63	3,992	99	3,893
Kosovo	97%	4,757	43	95	4,619	123	4,496
Kuwait	96%	4,731	83	14	4,634	197	4,437
Latvia	94%	4,886	15	68	4,803	322	4,481
Lithuania	94%	4,198	12	186	4,000	259	3,741
Malta	96%	3,914	17	115	3,782	152	3,630
Montenegro	98%	5,248	49	37	5,162	86	5,076

Exhibit 9.16: Student Sample Sizes – TIMSS 2019 – Fourth Grade (continued)

Country	Within-School Student Participation (Weighted Percentage)	Number of Students Sampled in Participating Schools	Number of Students Withdrawn from Class/ School	Number of Students Excluded	Number of Students Eligible	Number of Students Absent	Number of Students Assessed
Morocco	99%	8,051	217	0	7,834	111	7,723
Netherlands	97%	3,562	69	27	3,466	111	3,355
New Zealand	94%	5,611	100	164	5,347	328	5,019
North Macedonia	95%	3,531	32	44	3,455	185	3,270
Northern Ireland	91%	3,877	21	23	3,833	336	3,497
Norway (5)	94%	4,410	27	149	4,234	283	3,951
Oman	98%	7,079	94	57	6,928	114	6,814
Pakistan	98%	4,453	315	0	4,138	158	3,980
Philippines	98%	5,693	89	0	5,604	89	5,515
Poland	93%	5,427	44	100	5,283	401	4,882
Portugal	94%	5,015	35	366	4,614	314	4,300
Qatar	97%	5,251	127	60	5,064	131	4,933
Russian Federation	98%	4,282	8	144	4,130	108	4,022
Saudi Arabia	99%	5,585	23	25	5,537	84	5,453
Serbia	97%	4,667	53	93	4,521	141	4,380
Singapore	97%	6,209	22	0	6,187	201	5,986
Slovak Republic	97%	4,477	26	24	4,427	180	4,247
South Africa (5)	98%	12,289	107	0	12,182	291	11,891
Spain	95%	10,497	48	421	10,028	473	9,555
Sweden	95%	4,407	31	160	4,216	251	3,965
Turkey (5)	99%	4,554	142	319	4,093	65	4,028
United Arab Emirates	96%	28,029	414	564	27,051	1,217	25,834
United States	96%	9,955	152	601	9,202	426	8,776
Benchmarking Participants							
Ontario, Canada	95%	4,251	83	95	4,073	243	3,830
Quebec, Canada	96%	4,047	9	37	4,001	164	3,837
Moscow City, Russian Fed.	98%	3,992	11	35	3,946	103	3,843
Madrid, Spain	96%	3,666	17	123	3,526	136	3,390
Abu Dhabi, UAE	95%	9,822	38	239	9,545	508	9,037
Dubai, UAE	97%	8,125	362	213	7,550	285	7,265

Students attending a sampled class at the time the sample was chosen but leaving the class before the assessment was administered were classified as withdrawn.

Students with a disability or language barrier that prevented them from participating in the assessment were classified as excluded.

Students not present when the assessment was administered, and not subsequently assessed in a make-up session, were classified as absent.

Exhibit 9.17: **Student Sample Sizes – TIMSS 2019 – Eighth Grade**

Country	Within-School Student Participation (Weighted Percentage)	Number of Students Sampled in Participating Schools	Number of Students Withdrawn from Class/ School	Number of Students Excluded	Number of Students Eligible	Number of Students Absent	Number of Students Assessed
Australia	91%	10,383	213	161	10,009	949	9,060
Bahrain	97%	5,947	44	12	5,891	166	5,725
Chile	96%	4,469	68	76	4,325	210	4,115
Chinese Taipei	98%	5,185	106	42	5,037	122	4,915
Cyprus	96%	3,800	23	94	3,683	162	3,521
Egypt	97%	7,700	214	0	7,486	276	7,210
England	95%	3,785	140	70	3,575	210	3,365
Finland	96%	5,252	62	62	5,128	254	4,874
France	97%	4,122	53	49	4,020	146	3,874
Georgia	97%	3,540	37	73	3,430	115	3,315
Hong Kong SAR	94%	3,612	12	73	3,527	262	3,265
Hungary	97%	4,862	23	86	4,753	184	4,569
Iran, Islamic Rep. of	98%	6,242	110	35	6,097	117	5,980
Ireland	88%	4,763	46	39	4,678	560	4,118
Israel	93%	4,154	36	51	4,067	336	3,731
Italy	97%	3,919	22	153	3,744	125	3,619
Japan	94%	4,763	3	17	4,743	297	4,446
Jordan	98%	7,856	484	13	7,359	183	7,176
Kazakhstan	99%	4,587	34	28	4,525	72	4,453
Korea, Rep. of	98%	4,025	18	37	3,970	109	3,861
Kuwait	97%	4,818	92	0	4,726	152	4,574
Lebanon	95%	5,117	151	0	4,966	236	4,730
Lithuania	93%	4,262	19	98	4,145	319	3,826
Malaysia	98%	7,323	120	0	7,203	138	7,065
Morocco	98%	9,081	395	0	8,686	228	8,458
New Zealand	91%	6,775	119	79	6,577	526	6,051
Norway (9)	89%	5,335	41	141	5,153	578	4,575
Oman	99%	7,024	132	37	6,855	104	6,751
Portugal	96%	3,752	32	152	3,568	191	3,377
Qatar	97%	4,196	138	32	4,026	142	3,884
Romania	94%	4,803	13	15	4,775	281	4,494
Russian Federation	97%	4,125	28	76	4,021	120	3,901
Saudi Arabia	99%	5,762	19	13	5,730	50	5,680
Singapore	96%	5,074	19	0	5,055	202	4,853
South Africa (9)	96%	22,658	921	0	21,737	908	20,829

Exhibit 9.17: **Student Sample Sizes – TIMSS 2019 – Eighth Grade (continued)**

Country	Within-School Student Participation (Weighted Percentage)	Number of Students Sampled in Participating Schools	Number of Students Withdrawn from Class/ School	Number of Students Excluded	Number of Students Eligible	Number of Students Absent	Number of Students Assessed
Sweden	92%	4,683	64	213	4,406	410	3,996
Turkey	99%	4,377	111	123	4,143	66	4,077
United Arab Emirates	96%	23,974	251	315	23,408	1,074	22,334
United States	94%	9,924	307	242	9,375	677	8,698
Benchmarking Participants							
Ontario, Canada	94%	4,194	63	75	4,056	280	3,776
Quebec, Canada	95%	3,411	28	7	3,376	198	3,178
Moscow City, Russian Fed.	97%	3,963	21	19	3,923	140	3,783
Gauteng, RSA (9)	97%	6,025	188	0	5,837	204	5,633
Western Cape, RSA (9)	95%	5,901	284	0	5,617	266	5,351
Abu Dhabi, UAE	96%	8,770	41	86	8,643	439	8,204
Dubai, UAE	96%	6,308	199	141	5,968	240	5,728

Students attending a sampled class at the time the sample was chosen but leaving the class before the assessment was administered were classified as withdrawn.

Students with a disability or language barrier that prevented them from participating in the assessment were classified as excluded.

Students not present when the assessment was administered, and not subsequently assessed in a make-up session, were classified as absent.

TIMSS 2019 Trends in Student Populations

Because an important goal of the TIMSS 2019 assessment was to measure changes in students' mathematics and science achievement across assessment cycles, it was important to track any changes over time in population composition and coverage that might be related to student achievement. Exhibits 9.18 and 9.19 present, for each country, trends across cycles (2019, 2015, 2011, 2007, 2003, and 1995 at the fourth grade and 2019, 2015, 2011, 2007, 2003, 1999, and 1995 at the eighth grade) in four important characteristics of the assessment populations: number of years of formal schooling, average student age, percent of students in the national target population excluded from the assessment, and overall participation rates after using replacements. Most countries and benchmarking participants were very similar with regard to these characteristics across the assessment cycles, although there have been changes in some countries in the age and grade structure of the assessed populations, in the target population coverage, and in the exclusion rate.

In terms of changes in age structure, the Russian Federation has undergone changes in the age at which children enter schools that are reflected in their samples. In 2003, the Russian fourth grade sample contained third grade students from some regions and fourth grade students from others, whereas all students were in the fourth grade by 2007. At the eighth grade, there was still a mixture of seventh and eighth grade students in 2007, but by 2011 the sample was all eighth grade students, with correspondingly a higher average age. Turkey chose to assess students at the fifth grade in 2019, breaking the trend with previous cycles.

National coverage of the international target population was generally comprehensive for most countries and has not changed across assessments, with just a few exceptions. At the fourth grade, Kuwait assessed students from both the public and private schools in the 2019 and 2015 cycles while they assessed only students from the public schools in prior cycles. As a result, the 2019 trend is only with the 2015 data.

In most countries, exclusion rates did not exceed the TIMSS 2019 guidelines of 5 percent, and have not changed very much across assessments cycles. At the fourth grade, Denmark reduced its overall exclusion rate of 4.4 percent between 2015 and 2019 by providing more precise guidelines on within-school exclusions of special needs students. During that same period, Serbia also decreased its overall exclusion rate by reducing their within-school exclusion of non-native language speakers and by excluding less very small schools. On the other hand, the student exclusion rate was higher in 2019 than in 2015 at the fourth grade by more than 1.5 percentage points in Belgium (Flemish), England, New Zealand, Russian Federation, Saudi Arabia, and Singapore. At the eighth grade, those with increases of more than 1.5 percentage points in their exclusions since 2015 included Egypt, England, Hong Kong SAR, Oman, Russian Federation, Saudi Arabia, Singapore, and Turkey.

Exhibit 9.18: Trends in Student Populations – TIMSS 2019 – Fourth Grade

Country	Years of Formal Schooling*	Average Age at Time of Testing	Overall Exclusion Rates	Overall Participation Rates (After Replacement)
Armenia				
2019	4	9.9	1.2%	97%
[b]2015	4	9.9	1.0%	96%
2011	4	10.0	2.0%	98%
2003	4	10.9	2.9%	90%
Australia				
2019	4	10.1	4.8%	94%
2015	4	10.0	4.2%	94%
2011	4	10.0	4.4%	93%
2007	4	9.9	4.0%	95%
†2003	4	9.9	2.7%	85%
≡1995	4 or 5	10.2	1.8%	66%

Exhibit 9.18: **Trends in Student Populations – TIMSS 2019 – Fourth Grade (continued)**

Country	Years of Formal Schooling*	Average Age at Time of Testing	Overall Exclusion Rates	Overall Participation Rates (After Replacement)
Austria				
2019	4	10.4	5.4%	97%
2011	4	10.3	5.1%	98%
2007	4	10.3	5.0%	97%
[≡]1995	4	10.5	2.8%	69%
Azerbaijan				
2019	4	10.3	2.6%	92%
[2]2011	4	10.2	7.2%	100%
Bahrain				
2019	4	9.8	0.8%	98%
[2]2015	4	9.9	5.6%	99%
[b]2011	4	10.4	1.1%	90%
Belgium (Flemish)				
[†]2019	4	10.0	3.0%	89%
[†]2015	4	10.1	1.4%	95%
2011	4	10.0	5.0%	92%
[2]2003	4	10.0	6.3%	97%
Bulgaria				
2019	4	10.7	3.4%	95%
2015	4	10.8	2.9%	93%
Canada				
[1 2]2019	4	9.9	7.0%	86%
[1 2 †]2015	4	9.9	6.1%	80%
Chile				
2019	4	10.1	3.8%	95%
2015	4	10.2	3.7%	88%
2011	4	10.1	3.7%	95%
Chinese Taipei				
2019	4	10.2	2.0%	98%
2015	4	10.2	2.4%	99%
2011	4	10.2	1.4%	99%
2007	4	10.2	2.8%	100%
2003	4	10.2	3.1%	99%

Exhibit 9.18: **Trends in Student Populations – TIMSS 2019 – Fourth Grade (continued)**

Country	Years of Formal Schooling*	Average Age at Time of Testing	Overall Exclusion Rates	Overall Participation Rates (After Replacement)
Croatia				
2019	4	10.5	4.2%	87%
2015	4	10.6	4.4%	94%
[2]2011	4	10.7	7.9%	95%
Cyprus				
2019	4	9.8	4.6%	97%
2015	4	9.8	4.6%	98%
2003	4	9.9	2.9%	97%
1995	4	9.8	3.2%	83%
Czech Republic				
2019	4	10.4	4.7%	96%
2015	4	10.4	4.2%	95%
2011	4	10.4	5.1%	94%
2007	4	10.3	4.9%	92%
1995	4	10.4	4.1%	86%
Denmark				
†2019	4	10.9	3.1%	83%
[2] †2015	4	10.9	7.5%	86%
[2]2011	4	11.0	6.3%	87%
[2]2007	4	11.0	4.1%	85%
England				
[2]2019	5	10.2	5.8%	89%
2015	5	10.1	2.3%	96%
2011	5	10.2	2.0%	78%
2007	5	10.2	2.1%	84%
†2003	5	10.3	1.9%	76%
[3] †1995	5	10.0	12.1%	83%
Finland				
2019	4	10.8	3.3%	97%
2015	4	10.8	2.0%	97%
2011	4	10.8	3.1%	96%

Exhibit 9.18: Trends in Student Populations – TIMSS 2019 – Fourth Grade (continued)

Country	Years of Formal Schooling*	Average Age at Time of Testing	Overall Exclusion Rates	Overall Participation Rates (After Replacement)
France				
2019	4	9.9	4.4%	98%
2015	4	9.9	5.3%	97%
Georgia				
[1]2019	4	10.1	4.7%	96%
[1]2015	4	9.7	4.9%	98%
[1]2011	4	10.0	4.9%	96%
[1]2007	4	10.1	4.8%	98%
Germany				
2019	4	10.4	3.9%	97%
2015	4	10.4	2.7%	95%
2011	4	10.4	1.9%	95%
2007	4	10.4	1.3%	96%
Hong Kong SAR				
†2019	4	10.1	3.5%	79%
†2015	4	10.1	2.2%	76%
[2]2011	4	10.1	8.5%	82%
2007	4	10.2	5.4%	81%
†2003	4	10.2	3.8%	83%
1995	4	10.1	2.7%	83%
Hungary				
2019	4	10.5	4.1%	96%
2015	4	10.7	4.8%	96%
2011	4	10.7	4.2%	96%
2007	4	10.7	4.4%	96%
[2]2003	4	10.5	8.1%	93%
1995	4	10.4	3.8%	92%
Iran, Islamic Rep. of				
2019	4	10.2	4.2%	99%
2015	4	10.2	4.0%	99%
2011	4	10.2	4.5%	99%
2007	4	10.2	3.0%	99%

Exhibit 9.18: **Trends in Student Populations – TIMSS 2019 – Fourth Grade (continued)**

Country	Years of Formal Schooling*	Average Age at Time of Testing	Overall Exclusion Rates	Overall Participation Rates (After Replacement)
[2]2003	4	10.4	5.7%	98%
1995	4	10.5	1.3%	97%
Ireland				
2019	4	10.4	3.0%	91%
2015	4	10.4	2.7%	96%
2011	4	10.3	2.5%	95%
[2]1995	4	10.3	6.9%	90%
Italy				
2019	4	9.6	4.9%	97%
[2]2015	4	9.7	6.2%	94%
2011	4	9.7	3.7%	95%
2007	4	9.8	5.3%	97%
2003	4	9.8	4.2%	97%
Japan				
2019	4	10.4	2.2%	95%
2015	4	10.5	2.9%	97%
2011	4	10.5	3.2%	96%
2007	4	10.5	1.1%	95%
2003	4	10.4	0.8%	97%
1995	4	10.4	3.0%	92%
Kazakhstan				
[2]2019	4	10.4	5.8%	99%
[2]2011	4	10.4	6.3%	99%
Korea, Rep. of				
2019	4	10.5	2.3%	97%
2015	4	10.5	2.5%	97%
2011	4	10.4	2.5%	98%
[2]1995	4	10.3	6.6%	95%
Kuwait				
2019	4	9.7	1.7%	94%
2015	4	9.7	3.0%	90%

Exhibit 9.18: **Trends in Student Populations – TIMSS 2019 – Fourth Grade (continued)**

Country	Years of Formal Schooling*	Average Age at Time of Testing	Overall Exclusion Rates	Overall Participation Rates (After Replacement)
Latvia				
[2]2019	4	10.8	6.9%	93%
2003	4	11.1	4.4%	88%
Lithuania				
[2]2019	4	10.7	6.7%	94%
[2]2015	4	10.7	6.1%	94%
[1 2]2011	4	10.7	5.6%	94%
[1]2007	4	10.8	5.4%	94%
[1]2003	4	10.9	4.6%	87%
Malta				
2019	4	9.8	4.5%	96%
2011	5	9.8	3.6%	95%
Morocco				
2019	4	10.1	1.8%	99%
2015	4	10.3	1.5%	99%
2011	4	10.5	2.0%	96%
Netherlands				
≡2019	4	10.1	3.5%	73%
†2015	4	10.0	3.2%	83%
†2011	4	10.2	4.0%	79%
‡2007	4	10.2	4.8%	91%
†2003	4	10.2	5.2%	84%
≡1995	4	10.3	4.4%	59%
New Zealand				
[2]2019	4.5 - 5.5	10.0	6.9%	93%
2015	4.5 - 5.5	10.0	4.8%	90%
2011	4.5 - 5.5	9.9	4.9%	90%
2007	4.5 - 5.5	10.0	5.4%	96%
2003	4.5 - 5.5	10.0	4.0%	93%
1995	4.5 - 5.5	10.0	1.3%	95%

Exhibit 9.18: **Trends in Student Populations – TIMSS 2019 – Fourth Grade (continued)**

Country	Years of Formal Schooling*	Average Age at Time of Testing	Overall Exclusion Rates	Overall Participation Rates (After Replacement)
Northern Ireland				
†2019	4	10.4	2.8%	78%
‡2015	4	10.4	2.7%	71%
†2011	4	10.4	3.5%	79%
Norway (5)				
†2019	5	10.7	4.7%	84%
2015	5	10.7	4.7%	89%
Oman				
2019	4	9.7	2.2%	98%
2015	4	9.6	0.8%	97%
2011	4	9.9	1.5%	96%
Philippines				
[2]2019	4	10.1	7.7%	98%
2003	4	10.8	4.5%	81%
Poland				
2019	4	10.3	3.1%	93%
2015	4	10.7	4.0%	92%
Portugal				
[2]2019	4	10.0	7.8%	94%
[2]2015	4	9.9	6.5%	92%
2011	4	10.0	2.5%	92%
[2]1995	4	10.4	7.3%	92%
Qatar				
2019	4	9.9	2.2%	97%
2015	4	10.1	3.8%	99%
[2]2011	4	10.0	6.2%	99%
Russian Federation				
[2]2019	4	10.8	6.3%	97%
2015	4	10.8	4.0%	98%
2011	4	10.8	5.3%	98%
2007	4	10.8	3.6%	98%
[2]2003	3 or 4	10.6	6.8%	97%

Exhibit 9.18: Trends in Student Populations – TIMSS 2019 – Fourth Grade (continued)

Country	Years of Formal Schooling*	Average Age at Time of Testing	Overall Exclusion Rates	Overall Participation Rates (After Replacement)
Saudi Arabia				
[2]2019	4	9.9	10.5%	98%
2015	4	10.0	1.9%	93%
2011	4	10.0	1.6%	99%
Serbia				
[2]2019	4	10.6	8.2%	97%
[3]2015	4	10.7	11.3%	96%
[2]2011	4	10.8	9.4%	97%
Singapore				
[3]2019	4	10.4	12.8%	97%
[2]2015	4	10.4	10.1%	96%
[2]2011	4	10.4	6.3%	96%
2007	4	10.4	1.5%	96%
2003	4	10.3	0.0%	98%
1995	4	10.3	0.0%	98%
Slovak Republic				
[2]2019	4	10.4	5.5%	96%
2015	4	10.4	4.2%	97%
2011	4	10.4	4.6%	96%
2007	4	10.4	3.3%	97%
South Africa (5)				
2019	5	11.5	1.1%	97%
[b]2015	5	11.5	2.2%	98%
Spain				
2019	4	9.9	5.4%	95%
[2]2015	4	9.9	5.6%	95%
2011	4	9.8	5.3%	97%
Sweden				
2019	4	10.8	5.4%	95%
[2]2015	4	10.8	5.7%	95%
2011	4	10.7	4.1%	91%
2007	4	10.8	3.1%	97%

Exhibit 9.18: **Trends in Student Populations – TIMSS 2019 – Fourth Grade (continued)**

Country	Years of Formal Schooling*	Average Age at Time of Testing	Overall Exclusion Rates	Overall Participation Rates (After Replacement)
United Arab Emirates				
2019	4	9.7	3.2%	96%
2015	4	9.8	4.7%	97%
2011	4	9.8	3.3%	97%
United States				
[2] [†]2019	4	10.2	7.2%	84%
[2] [†]2015	4	10.2	6.8%	81%
[2]2011	4	10.2	7.0%	80%
[2] [†]2007	4	10.3	9.2%	84%
[†]2003	4	10.2	5.1%	78%
1995	4	10.2	4.7%	80%
Benchmarking Participants				
Ontario, Canada				
[2]2019	4	9.8	7.0%	90%
2015	4	9.8	3.4%	90%
2011	4	9.8	5.3%	94%
[2]2007	4	9.8	6.3%	92%
2003	4	9.8	4.8%	90%
[2]1995	4	9.8	–	92%
Quebec, Canada				
2019	4	10.1	4.4%	83%
[b]2015	4	10.1	5.4%	59%
2011	4	10.1	3.7%	91%
[2]2007	4	10.1	6.4%	84%
2003	4	10.1	3.6%	91%
1995	4	10.3	–	81%
Abu Dhabi, UAE				
2019	4	9.7	3.6%	95%
[2]2015	4	9.8	5.8%	97%
2011	4	9.7	2.7%	97%

Exhibit 9.18: Trends in Student Populations – TIMSS 2019 – Fourth Grade (continued)

Country	Years of Formal Schooling*	Average Age at Time of Testing	Overall Exclusion Rates	Overall Participation Rates (After Replacement)
Dubai, UAE				
[2]2019	4	9.9	5.6%	97%
2015	4	9.8	5.3%	97%
2011	4	9.8	5.1%	96%
þ ‡2007	4	10.0	5.4%	67%

* Represents years of schooling counting from the first year of ISCED Level 1.

Data are included only for assessment years with comparable results for each country.

See Exhibit 9.3 for population coverage notes 1, 2, and 3. See Exhibit 9.10 for sampling guidelines and sampling participation notes †, ‡, and ≡.

þ Tested the same cohort of students as other countries, but later in the assessment year.

Armenia began testing younger students in 2011 due to educational reforms.

Bahrain in 2015 administered both TIMSS and TIMSS Numeracy assessments to fourth grade students. Results for 2015 in mathematics are based on the average of both.

Georgia in 2011 excluded schools in South Ossetia and Abkhazia due to lack of access and absence of official statistics. Abkhazia refugee schools in other territories of Georgia were included in the sample frame.

Iran in 2015 administered both TIMSS and TIMSS Numeracy assessments to fourth grade students. Results for 2015 in mathematics are based on the average of both.

Results for Lithuania before 2015 do not include students taught in Polish or Russian.

Morocco and the Philippines in 2019 administered the less difficult fourth grade mathematics assessment.

Kuwait and Morocco in 2015 administered both TIMSS and TIMSS Numeracy assessments to fourth grade students. Results for 2015 in mathematics are based on the average of both.

Saudi Arabia and South Africa in 2019 administered the less difficult fourth grade mathematics assessment. South Africa in 2015 participated in only TIMSS Numeracy at the fifth grade.

Ontario and Quebec in 1995 participated as part of Canada. A dash (–) indicates comparable data not available.

Exhibit 9.19: Trends in Student Populations – TIMSS 2019 – Eighth Grade

Country	Years of Formal Schooling*	Average Age at Time of Testing	Overall Exclusion Rates	Overall Participation Rates (After Replacement)
Australia				
2019	8	14.1	3.8%	91%
2015	8	14.0	3.5%	90%
2011	8	14.0	3.2%	88%
2007	8	13.9	1.9%	93%
2003	8	13.9	1.3%	83%
‡ 1995	8 or 9	14.2	0.8%	70%
Bahrain				
2019	8	13.8	0.6%	97%
2015	8	14.0	3.8%	97%
þ 2011	8	14.4	1.6%	97%
2007	8	14.1	1.5%	97%
2003	8	14.1	0.0%	98%
Chile				
2019	8	14.2	2.2%	95%
2015	8	14.3	1.9%	85%
2011	8	14.2	2.8%	95%
2003	8	14.2	2.2%	99%
1999	8	14.4	2.8%	96%
Chinese Taipei				
2019	8	14.3	1.5%	97%
2015	8	14.3	1.7%	98%
2011	8	14.2	1.3%	99%
2007	8	14.2	3.3%	99%
2003	8	14.2	4.8%	99%
1999	8	14.2	1.6%	99%
Cyprus				
2019	8	13.8	2.8%	96%
2007	8	13.8	2.5%	96%
2003	8	13.8	2.5%	96%
1999	8	13.8	0.8%	97%
1995	8	13.7	0.0%	97%
Egypt				
[2] 2019	8	13.9	9.1%	97%
2015	8	14.1	0.1%	91%

Exhibit 9.19: Trends in Student Populations – TIMSS 2019 – Eighth Grade (continued)

Country	Years of Formal Schooling*	Average Age at Time of Testing	Overall Exclusion Rates	Overall Participation Rates (After Replacement)
2007	8	14.1	0.5%	98%
2003	8	14.4	3.4%	97%
England				
2019	9	14.0	4.8%	85%
2015	9	14.1	2.3%	92%
[‡] 2011	9	14.2	2.2%	70%
[†] 2007	9	14.2	2.3%	75%
[≡] 2003	9	14.3	2.1%	46%
[†] 1999	9	14.2	5.0%	77%
[3 †] 1995	9	14.0	11.3%	77%
Finland				
2019	8	14.8	3.1%	95%
2011	8	14.8	3.4%	93%
France				
2019	8	13.9	3.8%	97%
1995	8	14.3	2.0%	82%
Georgia				
[1] 2019	8	13.8	4.3%	89%
[1 2] 2015	8	13.7	6.0%	98%
[1] 2011	8	14.2	4.5%	97%
[1] 2007	8	14.2	3.9%	97%
Hong Kong SAR				
[†] 2019	8	14.1	3.3%	81%
2015	8	14.2	1.6%	81%
2011	8	14.2	5.3%	75%
[†] 2007	8	14.4	3.8%	75%
[†] 2003	8	14.4	3.4%	80%
[†] 1999	8	14.2	0.8%	74%
1995	8	14.2	2.0%	81%
Hungary				
2019	8	14.6	4.4%	96%
2015	8	14.7	5.4%	96%
2011	8	14.7	4.4%	95%
2007	8	14.6	3.9%	96%

Exhibit 9.19: Trends in Student Populations – TIMSS 2019 – Eighth Grade (continued)

Country	Years of Formal Schooling*	Average Age at Time of Testing	Overall Exclusion Rates	Overall Participation Rates (After Replacement)
[2] 2003	8	14.5	8.5%	94%
1999	8	14.4	4.3%	93%
1995	8	14.3	3.8%	87%
Iran, Islamic Rep. of				
2019	8	14.1	0.9%	98%
2015	8	14.2	2.2%	98%
2011	8	14.3	2.2%	99%
2007	8	14.2	0.5%	98%
[2] 2003	8	14.4	6.5%	98%
1999	8	14.6	4.4%	98%
1995	8	14.6	0.3%	98%
Ireland				
2019	8	14.4	1.0%	86%
2015	8	14.4	1.2%	91%
1995	8	14.4	0.4%	81%
Israel				
[3] 2019	8	14.0	23.2%	91%
[3] 2015	8	14.0	22.8%	93%
[3] 2011	8	14.0	22.6%	92%
Italy				
2019	8	13.7	4.3%	97%
[2] 2015	8	13.8	6.1%	93%
2011	8	13.8	4.7%	93%
2007	8	13.9	5.0%	96%
2003	8	13.9	3.6%	97%
[2] 1999	8	14.0	6.7%	97%
Japan				
2019	8	14.4	1.8%	88%
2015	8	14.5	2.3%	93%
2011	8	14.5	2.8%	87%
2007	8	14.5	3.5%	91%
2003	8	14.4	0.6%	93%
1999	8	14.4	1.3%	89%
1995	8	14.4	0.6%	90%

Exhibit 9.19: Trends in Student Populations – TIMSS 2019 – Eighth Grade (continued)

Country	Years of Formal Schooling*	Average Age at Time of Testing	Overall Exclusion Rates	Overall Participation Rates (After Replacement)
Jordan				
2019	8	13.9	0.2%	98%
2015	8	13.8	1.0%	96%
2011	8	13.9	0.4%	96%
2007	8	14.0	2.0%	96%
2003	8	13.9	1.3%	96%
1999	8	14.0	3.0%	99%
Kazakhstan				
[2] 2019	8	14.3	5.8%	99%
2011	8	14.6	5.1%	98%
Korea, Rep. of				
2019	8	14.5	1.6%	98%
2015	8	14.4	2.1%	98%
2011	8	14.3	1.9%	99%
2007	8	14.3	1.6%	99%
[b] 2003	8	14.6	4.9%	98%
1999	8	14.4	4.0%	100%
1995	8	14.2	3.8%	95%
Kuwait				
2019	8	13.8	2.0%	96%
2015	8	13.7	3.3%	85%
Lebanon				
2019	8	14.0	1.2%	88%
2015	8	14.2	1.3%	88%
2011	8	14.3	1.4%	94%
2007	8	14.4	1.4%	85%
2003	8	14.6	1.4%	91%
Lithuania				
2019	8	14.7	5.3%	92%
[2] 2015	8	14.7	7.0%	93%
[1] 2011	8	14.7	4.8%	92%
[1] 2007	8	14.9	4.2%	90%
[1] 2003	8	14.9	2.6%	84%
[1 b] 1999	8	15.2	4.5%	89%
[1 2] 1995	8	14.3	6.6%	83%

Exhibit 9.19: **Trends in Student Populations – TIMSS 2019 – Eighth Grade (continued)**

Country	Years of Formal Schooling*	Average Age at Time of Testing	Overall Exclusion Rates	Overall Participation Rates (After Replacement)
Malaysia				
2019	8	14.3	3.2%	98%
2015	8	14.3	4.3%	98%
2011	8	14.4	0.1%	98%
2007	8	14.3	3.3%	98%
2003	8	14.3	4.0%	98%
1999	8	14.4	4.6%	99%
Morocco				
2019	8	14.5	0.0%	98%
2015	8	14.5	0.0%	95%
2011	8	14.7	0.1%	94%
New Zealand				
†2019	8.5 – 9.5	13.9	4.2%	81%
†2015	8.5 – 9.5	14.1	3.1%	81%
2011	8.5 – 9.5	14.1	3.2%	88%
2003	8.5 – 9.5	14.1	4.4%	90%
1999	8.5 – 9.5	14.0	2.4%	91%
1995	8.5 – 9.5	14.0	1.7%	94%
Norway (9)				
†2019	9	14.7	4.0%	84%
2015	9	14.7	3.7%	87%
Oman				
2019	8	13.9	2.2%	99%
2015	8	14.0	0.4%	96%
2011	8	14.1	1.2%	97%
2007	8	14.3	1.2%	99%
Portugal				
2019	8	14.0	5.5%	94%
1995	8	14.5	0.3%	92%
Qatar				
2019	8	14.0	2.2%	97%
2015	8	14.1	3.2%	96%
2011	8	14.0	4.5%	99%

Exhibit 9.19: **Trends in Student Populations – TIMSS 2019 – Eighth Grade (continued)**

Country	Years of Formal Schooling*	Average Age at Time of Testing	Overall Exclusion Rates	Overall Participation Rates (After Replacement)
Romania				
2019	8	14.8	3.2%	94%
2011	8	14.9	1.3%	99%
2007	8	15.0	1.8%	97%
2003	8	15.0	0.5%	98%
1999	8	14.8	3.7%	97%
1995	8	14.6	2.8%	89%
Russian Federation				
[2] 2019	8	14.8	5.7%	97%
2015	8	14.7	3.7%	97%
[2] 2011	8	14.7	6.0%	98%
2007	7 or 8	14.6	2.3%	97%
2003	7 or 8	14.2	5.5%	96%
1999	7 or 8	14.1	1.7%	97%
[2] 1995	7 or 8	14.0	6.3%	95%
Saudi Arabia				
[2] 2019	8	13.9	10.0%	99%
2015	8	14.1	2.1%	97%
2011	8	14.1	1.2%	98%
Singapore				
[2] 2019	8	14.3	10.3%	96%
[2] 2015	8	14.4	7.0%	97%
[2] 2011	8	14.4	6.0%	95%
2007	8	14.4	1.8%	95%
2003	8	14.3	0.0%	97%
1999	8	14.4	0.0%	98%
1995	8	14.5	4.6%	95%
South Africa (9)				
2019	9	15.5	1.1%	96%
[b] 2015	9	15.7	1.5%	96%
2011	9	16.0	1.4%	95%
Sweden				
[2] 2019	8	14.8	6.3%	91%
2015	8	14.7	5.4%	94%
2011	8	14.8	5.1%	92%

Exhibit 9.19: Trends in Student Populations – TIMSS 2019 – Eighth Grade (continued)

Country	Years of Formal Schooling*	Average Age at Time of Testing	Overall Exclusion Rates	Overall Participation Rates (After Replacement)
2007	8	14.8	3.6%	94%
2003	8	14.9	2.8%	87%
1995	7	14.9	0.9%	90%
Turkey				
2019	8	13.9	3.4%	99%
2015	8	13.9	1.3%	98%
2011	8	14.0	1.5%	97%
United Arab Emirates				
2019	8	13.7	2.4%	96%
2015	8	13.9	3.6%	97%
2011	8	13.9	2.8%	97%
United States				
† 2019	8	14.2	3.9%	79%
† 2015	8	14.2	5.1%	78%
2 2011	8	14.2	7.2%	81%
2 † 2007	8	14.3	7.9%	77%
‡ 2003	8	14.2	4.9%	73%
1999	8	14.2	3.9%	85%
† 1995	8	14.2	2.1%	78%
Benchmarking Participants				
Ontario, Canada				
2019	8	13.8	5.5%	88%
2015	8	13.8	2.5%	87%
2 2011	8	13.8	5.6%	93%
2 2007	8	13.8	6.2%	89%
2 2003	8	13.8	6.0%	89%
1999	8	13.9	5.1%	93%
1995	8	14.0	-	90%
Quebec, Canada				
‡ 2019	8	14.2	4.2%	73%
≡ 2015	8	14.3	5.3%	58%
2011	8	14.2	4.9%	88%
3 2007	8	14.2	13.6%	77%

Exhibit 9.19: Trends in Student Populations – TIMSS 2019 – Eighth Grade (continued)

Country	Years of Formal Schooling*	Average Age at Time of Testing	Overall Exclusion Rates	Overall Participation Rates (After Replacement)
2003	8	14.2	4.8%	85%
1999	8	14.3	1.3%	92%
1995	8	14.5	-	89%
Abu Dhabi, UAE				
2019	8	13.7	1.7%	96%
2015	8	13.9	4.1%	98%
2011	8	13.8	1.7%	96%
Dubai, UAE				
[2] 2019	8	13.9	5.5%	96%
2015	8	13.9	5.2%	97%
2011	8	13.9	4.0%	95%
þ ‡ 2007	8	14.2	5.0%	69%

* Represents years of schooling counting from the first year of ISCED Level 1.

Data are included only for assessment years with comparable results for each country.

See Exhibit 9.4 for population coverage notes 1, 2, and 3. See Exhibit 9.11 for sampling guidelines and sampling participation notes †, ‡, and ≡.

þ Tested the same cohort of students as other countries, but later in the assessment year.

Egypt's 2015 exclusion rate may be underestimated.

Georgia in 2011 excluded schools in South Ossetia and Abkhazia due to lack of access and absence of official statistics. Abkhazia refugee schools in other territories of Georgia were included in the sample frame.

Results for Lithuania before 2015 do not include students taught in Polish or Russian. Lithuania in 1999 tested the same cohort of students as other countries, but later in the assessment year.

Ontario and Quebec in 1995 and 1999 participated as part of Canada. A dash (-) indicates comparable data not available.

Characteristics of the Bridge Samples

As mentioned earlier, eTIMSS countries also provided a separate sample of bridge data in order to control for mode effects while linking the two versions to the TIMSS achievement scales and to safeguard the measurement of trends from previous assessments.

The bridge data result from administering the paper version of the trend items (eight blocks of items for each subject and grade that also were administered in 2015) to a separate, equivalent sample of students during the main data collection. The following sections of this chapter provide a summary of the major characteristics of the bridge samples for trend countries that participated in eTIMSS.

Overlap between the Bridge and eTIMSS Samples

As mentioned in Chapter 3, it was important that the eTIMSS countries' bridge samples mirror their main eTIMSS samples as closely as possible. For operational reasons, it was not possible to administer both the eTIMSS assessment and bridge assessment to the same students or in the same class. Consequently, bridge samples were obtained by a) selecting an additional class from a subset of the sampled schools, or b) selecting a separate sample of schools, or c) a combination of the two approaches. The sampling experts from Statistics Canada worked with each country during the sampling development stage to develop an optimal strategy for selecting the bridge sample.

Exhibits 9.20 and 9.21 present the number and percentage of students from the bridge sample that came from the eTIMSS schools.

Exhibit 9.20: Bridge – Percentage of Students from the Bridge Sample from eTIMSS schools – TIMSS 2019 – Fourth Grade

Country	Number of Students	Unweighted Percentage	Weighted Percentage
Austria	753	38.3%	36.4%
Canada	891	55.5%	60.3%
Chile	0	0.0%	0.0%
Chinese Taipei	1394	83.8%	85.6%
Croatia	683	46.4%	48.6%
Czech Republic	0	0.0%	0.0%
Denmark	450	31.4%	37.2%
England	0	0.0%	0.0%
Finland	0	0.0%	0.0%
France	0	0.0%	0.0%
Georgia	0	0.0%	0.0%
Germany	731	48.6%	44.9%
Hong Kong SAR	1304	98.1%	97.9%
Hungary	0	0.0%	0.0%
Italy	0	0.0%	0.0%
Korea, Rep. of	1215	78.8%	82.8%
Lithuania	0	0.0%	0.0%
Netherlands	0	0.0%	0.0%
Norway (5)	0	0.0%	0.0%
Portugal	1326	82.3%	87.8%
Qatar	1379	92.8%	98.5%
Russian Federation	0	0.0%	0.0%
Singapore	1881	100.0%	100.0%
Slovak Republic	460	28.6%	34.5%
Spain	840	50.3%	53.0%
Sweden	0	0.0%	0.0%
United Arab Emirates	2124	94.7%	98.9%
United States	1456	88.1%	90.7%

Exhibit 9.21: Bridge – Percentage of Students from the Bridge Sample from eTIMSS schools – TIMSS 2019 – Eighth Grade

Country	Number of Students	Unweighted Percentage	Weighted Percentage
Chile	0	0.0%	0.0%
Chinese Taipei	1530	97.0%	98.0%
England	0	0.0%	0.0%
Georgia	0	0.0%	0.0%
Hong Kong SAR	1423	100.0%	100.0%
Hungary	1751	100.0%	100.0%
Israel	1772	95.1%	91.2%
Italy	0	0.0%	0.0%
Korea, Rep. of	1548	91.4%	95.8%
Lithuania	1687	100.0%	100.0%
Malaysia	0	0.0%	0.0%
Norway (9)	0	0.0%	0.0%
Qatar	1408	94.5%	98.6%
Russian Federation	0	0.0%	0.0%
Singapore	1871	100.0%	100.0%
Sweden	0	0.0%	0.0%
Turkey	1218	67.0%	69.7%
United Arab Emirates	1936	92.7%	98.0%
United States	1307	88.1%	91.6%

National Coverage and Exclusions of the Bridge Samples

The coverage and school exclusions prior to school sampling are the same for the bridge and eTIMSS samples as they took place before the drawing of the samples. Although the within-school exclusion estimates for the bridge and eTIMSS samples could be different because the students in the two samples were not the same, in general the within-school exclusion rates estimated from the bridge samples were very similar to those estimated from the eTIMSS samples. However, because of the smaller sample sizes for the bridge, the within-school exclusion rate estimates from the eTIMSS samples are more precise than those estimated from the bridge samples. Exhibits 9.22 and 9.23 summarize population coverage and exclusions resulting from the bridge samples.

Exhibit 9.22: Bridge – Coverage for TIMSS 2019 – Fourth Grade Target Population

Country	International Target Population		Exclusions from National Target Population		
	Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
Austria	100%		0.9%	3.8%	4.8%
Canada	79%	Students from the provinces of Alberta, Manitoba, Newfoundland, Ontario, and Quebec	3.1%	3.5%	6.7%
Chile	100%		1.2%	2.4%	3.6%
Chinese Taipei	100%		0.3%	1.8%	2.1%
Croatia	100%		1.1%	3.3%	4.4%
Czech Republic	100%		2.5%	1.7%	4.2%
Denmark	100%		1.6%	1.1%	2.7%
England	100%		2.2%	7.1%	9.3%
Finland	100%		1.8%	1.5%	3.2%
France	100%		2.5%	2.3%	4.8%
Georgia	92%	Students taught in Georgian	2.8%	1.7%	4.6%
Germany	100%		1.7%	1.4%	3.1%
Hong Kong SAR	100%		1.1%	2.1%	3.2%
Hungary	100%		2.1%	1.5%	3.6%
Italy	100%		0.9%	4.8%	5.7%
Korea, Rep. of	100%		0.9%	1.7%	2.5%
Lithuania	100%		2.6%	3.3%	6.0%
Netherlands	100%		2.6%	2.9%	5.5%
Norway (5)	100%		1.4%	2.6%	4.0%
Portugal	100%		0.9%	7.2%	8.1%
Qatar	100%		1.2%	1.5%	2.7%
Russian Federation	100%		2.4%	3.6%	6.0%
Singapore	100%		12.5%	0.3%	12.7%
Slovak Republic	100%		3.6%	1.3%	4.9%
Spain	100%		1.6%	4.2%	5.8%
Sweden	100%		1.6%	3.4%	5.0%
United Arab Emirates	100%		1.1%	1.3%	2.4%
United States	100%		0.0%	5.8%	5.8%

Exhibit 9.23: Bridge – Coverage for TIMSS 2019 Eighth Grade Target Population

Country	International Target Population		Exclusions from National Target Population		
	Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
Chile	100%		0.3%	3.1%	3.4%
Chinese Taipei	100%		0.1%	1.1%	1.2%
England	100%		2.9%	4.2%	7.1%
Georgia	91%	Students taught in Georgian	2.2%	2.4%	4.6%
Hong Kong SAR	100%		1.2%	1.3%	2.5%
Hungary	100%		2.5%	1.0%	3.5%
Israel	100%		19.5%	3.5%	22.9%
Italy	100%		0.8%	4.4%	5.1%
Korea, Rep. of	100%		0.7%	1.0%	1.7%
Lithuania	100%		3.2%	3.0%	6.2%
Malaysia	100%		1.9%	0.0%	1.9%
Norway (9)	100%		1.4%	2.8%	4.2%
Qatar	100%		1.3%	1.2%	2.5%
Russian Federation	100%		2.8%	3.1%	5.9%
Singapore	100%		10.1%	0.0%	10.1%
Sweden	100%		1.7%	4.0%	5.7%
Turkey	100%		1.1%	1.1%	2.1%
United Arab Emirates	100%		1.1%	1.7%	2.9%
United States	100%		0.0%	2.6%	2.6%

Target Population Size

Exhibits 9.24 and 9.25 show the number of schools and students in each country's target population[3] and bridge sample, as well as an estimate of the student population size based on the bridge sample data. The target population figures are derived from the sampling frame used to select the TIMSS 2019 samples, while the sample figures are based on the number of sampled schools and students that participated in the bridge assessments. The student population size estimated from the sample were computed using sampling weights, which are explained in more detail in Chapter 3.

3 After school-level exclusions.

Exhibit 9.24: Bridge – Population and Sample Sizes – TIMSS 2019 – Fourth Grade

Country	Population		Sample		
	Schools	Students	Schools	Students	Population Size Estimated from Sample
Austria	3,095	81,406	99	1,964	81,357
Canada	9,796	304,798	83	1,604	292,209
Chile	6,081	252,190	58	1,612	261,962
Chinese Taipei	2,476	190,975	68	1,663	187,133
Croatia	1,571	39,244	74	1,472	38,094
Czech Republic	3,578	114,774	58	2,030	110,694
Denmark	1,644	66,225	61	1,432	64,611
England	15,349	644,127	46	1,242	620,802
Finland	1,840	59,755	71	1,983	60,028
France	31,716	822,438	60	1,948	880,469
Georgia	1,678	42,980	58	1,632	41,954
Germany	17,584	716,091	78	1,505	713,942
Hong Kong SAR	564	60,786	54	1,329	60,002
Hungary	2,888	94,673	50	1,778	89,056
Italy	6,809	556,298	60	1,921	539,069
Korea, Rep. of	5,478	472,130	68	1,541	420,658
Lithuania	827	28,035	74	1,587	27,628
Netherlands	6,291	178,200	41	1,295	158,792
Norway (5)	1,945	62,012	55	1,899	66,756
Portugal	1,245	99,927	90	1,612	100,853
Qatar	247	25,506	63	1,486	25,664
Russian Federation	40,575	1,414,240	92	2,128	1,661,895
Singapore	187	39,934	56	1,881	39,978
Slovak Republic	2,000	52,222	70	1,610	52,347
Spain	12,861	489,765	69	1,670	464,033
Sweden	3,276	114,494	52	1,697	119,524
United Arab Emirates	754	85,609	98	2,243	88,435
United States	72,902	4,153,454	79	1,652	4,099,214

Exhibit 9.25: Bridge – Population and Sample Sizes – TIMSS 2019 – Eighth Grade

Country	Population		Sample		
	Schools	Students	Schools	Students	Population Size Estimated from Sample
Chile	5,767	246,120	56	1,526	230,321
Chinese Taipei	931	214,516	57	1,578	198,632
England	3,706	584,697	47	1,592	582,799
Georgia	1,837	45,339	53	1,314	44,074
Hong Kong SAR	478	54,160	54	1,423	55,605
Hungary	2,724	87,805	52	1,751	83,778
Israel	979	106,971	69	1,863	115,855
Italy	5,775	566,636	58	2,032	609,747
Korea, Rep. of	3,006	465,626	65	1,693	449,648
Lithuania	706	25,394	72	1,687	25,305
Malaysia	2,565	423,150	44	1,560	413,205
Norway (9)	1,012	60,847	51	2,018	61,630
Qatar	156	19,513	63	1,490	19,480
Russian Federation	37,308	1,326,933	92	2,083	1,424,446
Singapore	153	38,517	55	1,871	38,436
Sweden	1,600	108,164	52	1,582	114,578
Turkey	16,179	1,204,063	72	1,819	1,176,396
United Arab Emirates	685	68,113	88	2,089	64,713
United States	48,557	4,059,757	65	1,484	3,804,681

Sampling Participation for the TIMSS 2019 Bridge Samples

The bridge samples for TIMSS 2019 were subject to the same quality requirements as the TIMSS 2019 samples, as summarized in Exhibit 9.9.

Exhibits 9.26 through 9.29 present the school, classroom, student, and overall weighted and unweighted participation rates for each country's fourth and eighth grade bridge samples, respectively. At the fourth grade, all but two countries were in the same participation category for the bridge sample as for the eTIMSS samples. The Netherlands improved their rating and achieved the minimum acceptable participation rate after including replacement schools for their bridge sample and Norway met the participation requirement for their bridge sample without the use of replacement schools.

At the eighth grade, all but one country had the same participation category for the bridge sample as for the eTIMSS sample. The United States nearly met the required sampling participation rate at the eighth grade with the use of replacement for their bridge sample.

Exhibit 9.26: **Bridge – Participation Rates (Weighted) – TIMSS 2019 – Fourth Grade**

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Austria	99%	100%	100%	97%	96%	97%
Canada	85%	90%	100%	94%	80%	85%
Chile	100%	100%	100%	96%	95%	96%
Chinese Taipei	94%	100%	100%	99%	93%	99%
Croatia	98%	98%	100%	89%	87%	87%
Czech Republic	100%	100%	100%	94%	94%	94%
Denmark	65%	98%	100%	86%	56%	84%
England	82%	90%	100%	95%	77%	85%
Finland	98%	100%	100%	97%	96%	97%
France	98%	98%	100%	98%	96%	96%
Georgia	100%	100%	100%	96%	96%	96%
Germany	97%	100%	100%	96%	92%	96%
Hong Kong SAR	67%	85%	100%	87%	58%	74%
Hungary	100%	100%	100%	96%	96%	96%
Italy	100%	100%	100%	97%	97%	97%
Korea, Rep. of	100%	100%	100%	98%	98%	98%
Lithuania	100%	100%	100%	93%	93%	93%
Netherlands	55%	82%	99%	96%	53%	78%
Norway (5)	100%	100%	100%	95%	95%	95%
Portugal	91%	100%	100%	92%	84%	92%
Qatar	100%	100%	100%	97%	97%	97%
Russian Federation	100%	100%	100%	97%	97%	97%
Singapore	100%	100%	100%	96%	96%	96%
Slovak Republic	99%	100%	100%	96%	95%	96%
Spain	97%	98%	100%	97%	94%	96%
Sweden	97%	100%	100%	96%	93%	96%
United Arab Emirates	97%	97%	100%	96%	92%	92%
United States	77%	93%	100%	96%	74%	89%

Exhibit 9.27: Bridge – Participation Rates (Weighted) – TIMSS 2019 – Eighth Grade

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Chile	97%	100%	100%	96%	92%	96%
Chinese Taipei	100%	100%	100%	98%	98%	98%
England	81%	95%	100%	92%	74%	87%
Georgia	91%	91%	100%	97%	88%	88%
Hong Kong SAR	63%	85%	100%	90%	57%	77%
Hungary	100%	100%	100%	96%	96%	96%
Israel	91%	97%	100%	90%	82%	87%
Italy	100%	100%	100%	97%	97%	97%
Korea, Rep. of	100%	100%	100%	97%	97%	97%
Lithuania	100%	100%	100%	92%	92%	92%
Malaysia	100%	100%	100%	98%	98%	98%
Norway (9)	91%	93%	100%	91%	82%	85%
Qatar	100%	100%	100%	97%	97%	97%
Russian Federation	100%	100%	100%	97%	97%	97%
Singapore	98%	98%	100%	97%	95%	95%
Sweden	97%	100%	100%	92%	89%	92%
Turkey	100%	100%	100%	97%	97%	97%
United Arab Emirates	98%	98%	100%	97%	95%	95%
United States	72%	80%	100%	93%	67%	74%

Exhibit 9.28: **Bridge – Participation Rates (Unweighted) – TIMSS 2019 – Fourth Grade**

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Austria	99%	100%	100%	97%	96%	97%
Canada	84%	90%	100%	94%	78%	84%
Chile	98%	100%	100%	96%	94%	96%
Chinese Taipei	94%	100%	100%	99%	93%	99%
Croatia	97%	97%	100%	89%	87%	87%
Czech Republic	100%	100%	100%	95%	95%	95%
Denmark	65%	98%	100%	86%	56%	85%
England	81%	88%	100%	95%	77%	84%
Finland	99%	100%	100%	97%	96%	97%
France	98%	98%	100%	98%	96%	96%
Georgia	100%	100%	100%	96%	96%	96%
Germany	97%	100%	100%	96%	93%	96%
Hong Kong SAR	65%	82%	100%	87%	57%	71%
Hungary	100%	100%	100%	96%	96%	96%
Italy	98%	100%	100%	97%	95%	97%
Korea, Rep. of	100%	100%	100%	98%	98%	98%
Lithuania	100%	100%	100%	93%	93%	93%
Netherlands	54%	82%	97%	96%	50%	77%
Norway (5)	100%	100%	100%	95%	95%	95%
Portugal	92%	100%	100%	92%	85%	92%
Qatar	100%	100%	100%	97%	97%	97%
Russian Federation	100%	100%	100%	97%	97%	97%
Singapore	100%	100%	100%	96%	96%	96%
Slovak Republic	99%	100%	100%	96%	94%	96%
Spain	97%	99%	100%	97%	94%	95%
Sweden	96%	100%	100%	95%	92%	95%
United Arab Emirates	98%	98%	100%	96%	94%	94%
United States	76%	93%	100%	95%	73%	89%

Exhibit 9.29: Bridge – Participation Rates (Unweighted) – TIMSS 2019 – Eighth Grade

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Chile	95%	100%	100%	95%	90%	95%
Chinese Taipei	100%	100%	100%	98%	98%	98%
England	82%	94%	100%	91%	75%	86%
Georgia	91%	91%	100%	96%	88%	88%
Hong Kong SAR	65%	86%	100%	92%	60%	78%
Hungary	100%	100%	100%	96%	96%	96%
Israel	93%	97%	100%	90%	84%	87%
Italy	100%	100%	100%	97%	97%	97%
Korea, Rep. of	100%	100%	100%	97%	97%	97%
Lithuania	100%	100%	100%	92%	92%	92%
Malaysia	100%	100%	100%	98%	98%	98%
Norway (9)	93%	94%	100%	91%	84%	86%
Qatar	100%	100%	100%	97%	97%	97%
Russian Federation	100%	100%	100%	97%	97%	97%
Singapore	98%	98%	100%	97%	95%	95%
Sweden	98%	100%	100%	92%	90%	92%
Turkey	100%	100%	100%	97%	97%	97%
United Arab Emirates	99%	99%	100%	97%	96%	96%
United States	71%	79%	100%	93%	66%	74%

Exhibits 9.30 through 9.33 show the achieved bridge sample sizes in terms of schools and students for each country at fourth and eighth grade, respectively.

Exhibit 9.30: Bridge – School Sample Sizes – TIMSS 2019 – Fourth Grade

Country	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample that Participated	Number of Replacement Schools that Participated	Total Number of Schools that Participated
Austria	100	99	98	1	99
Canada	94	92	77	6	83
Chile	58	58	57	1	58
Chinese Taipei	68	68	64	4	68
Croatia	76	76	74	0	74
Czech Republic	60	58	58	0	58
Denmark	62	62	40	21	61
England	52	52	42	4	46
Finland	72	71	70	1	71
France	62	61	60	0	60
Georgia	58	58	58	0	58
Germany	78	78	76	2	78
Hong Kong SAR	66	66	43	11	54
Hungary	52	50	50	0	50
Italy	60	60	59	1	60
Korea, Rep. of	68	68	68	0	68
Lithuania	74	74	74	0	74
Netherlands	50	50	27	14	41
Norway (5)	56	55	55	0	55
Portugal	90	90	83	7	90
Qatar	63	63	63	0	63
Russian Federation	92	92	92	0	92
Singapore	56	56	56	0	56
Slovak Republic	70	70	69	1	70
Spain	70	70	68	1	69
Sweden	52	52	50	2	52
United Arab Emirates	101	100	98	0	98
United States	86	85	65	14	79

Exhibit 9.31: **Bridge – School Sample Sizes – TIMSS 2019 – Eighth Grade**

Country	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample that Participated	Number of Replacement Schools that Participated	Total Number of Schools that Participated
Chile	56	56	53	3	56
Chinese Taipei	57	57	57	0	57
England	50	50	41	6	47
Georgia	58	58	53	0	53
Hong Kong SAR	63	63	41	13	54
Hungary	54	52	52	0	52
Israel	71	71	66	3	69
Italy	58	58	58	0	58
Korea, Rep. of	66	65	65	0	65
Lithuania	72	72	72	0	72
Malaysia	44	44	44	0	44
Norway (9)	54	54	50	1	51
Qatar	63	63	63	0	63
Russian Federation	92	92	92	0	92
Singapore	56	56	55	0	55
Sweden	54	52	51	1	52
Turkey	72	72	72	0	72
United Arab Emirates	93	89	88	0	88
United States	83	82	58	7	65

Exhibit 9.32: Bridge – Student Sample Sizes – TIMSS 2019 – Fourth Grade

Country	Within-School Student Participation (Weighted Percentage)	Number of Students Sampled in Participating Schools	Number of Students Withdrawn from Class/ School	Number of Students Excluded	Number of Eligible Students	Number of Students Absent	Number of Students Assessed
Austria	97%	2,125	19	83	2,023	59	1,964
Canada	94%	1,773	19	40	1,714	110	1,604
Chile	96%	1,740	10	49	1,681	69	1,612
Chinese Taipei	99%	1,737	22	28	1,687	24	1,663
Croatia	89%	1,711	0	56	1,655	183	1,472
Czech Republic	94%	2,193	18	27	2,148	118	2,030
Denmark	86%	1,706	29	16	1,661	229	1,432
England	95%	1,445	14	121	1,310	68	1,242
Finland	97%	2,067	9	11	2,047	64	1,983
France	98%	2,050	13	46	1,991	43	1,948
Georgia	96%	1,742	12	30	1,700	68	1,632
Germany	96%	1,630	31	25	1,574	69	1,505
Hong Kong SAR	87%	1,561	2	34	1,525	196	1,329
Hungary	96%	1,898	16	27	1,855	77	1,778
Italy	97%	2,111	7	120	1,984	63	1,921
Korea, Rep. of	98%	1,616	16	24	1,576	35	1,541
Lithuania	93%	1,769	4	56	1,709	122	1,587
Netherlands	96%	1,389	17	27	1,345	50	1,295
Norway (5)	95%	2,057	9	44	2,004	105	1,899
Portugal	92%	1,891	10	125	1,756	144	1,612
Qatar	97%	1,600	40	24	1,536	50	1,486
Russian Federation	97%	2,264	5	65	2,194	66	2,128
Singapore	96%	1,962	5	0	1,957	76	1,881
Slovak Republic	96%	1,699	9	8	1,682	72	1,610
Spain	97%	1,810	4	79	1,727	57	1,670
Sweden	96%	1,845	14	53	1,778	81	1,697
United Arab Emirates	96%	2,385	12	37	2,336	93	2,243
United States	96%	1,827	22	74	1,731	79	1,652

Exhibit 9.33: **Bridge – Student Sample Sizes – TIMSS 2019 – Eighth Grade**

Country	Within-School Student Participation (Weighted Percentage)	Number of Students Sampled in Participating Schools	Number of Students Withdrawn from Class/ School	Number of Students Excluded	Number of Eligible Students	Number of Students Absent	Number of Students Assessed
Chile	96%	1,666	23	43	1,600	74	1,526
Chinese Taipei	98%	1,666	34	16	1,616	38	1,578
England	92%	1,834	47	43	1,744	152	1,592
Georgia	97%	1,412	13	34	1,365	51	1,314
Hong Kong SAR	90%	1,578	2	21	1,555	132	1,423
Hungary	96%	1,847	8	15	1,824	73	1,751
Israel	90%	2,101	15	16	2,070	207	1,863
Italy	97%	2,206	17	95	2,094	62	2,032
Korea, Rep. of	97%	1,775	7	19	1,749	56	1,693
Lithuania	92%	1,880	9	37	1,834	147	1,687
Malaysia	98%	1,613	25	0	1,588	28	1,560
Norway (9)	91%	2,288	10	65	2,213	195	2,018
Qatar	97%	1,613	53	20	1,540	50	1,490
Russian Federation	97%	2,195	11	41	2,143	60	2,083
Singapore	97%	1,945	10	0	1,935	64	1,871
Sweden	92%	1,800	16	63	1,721	139	1,582
Turkey	97%	1,956	51	25	1,880	61	1,819
United Arab Emirates	97%	2,203	6	35	2,162	73	2,089
United States	93%	1,644	32	18	1,594	110	1,484

Appendix 9A: Characteristics of National Samples

Albania

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 3), special needs schools, very small schools - Grade 3, language not Albanian, and special curriculum
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private), and urbanization (urban, rural) within public schools
- Implicit stratification by urbanization (urban, rural) within private school stratum
- Sampled two classrooms in large schools (measure of size > 100)

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Public - Urban	88	1	85	0	0	2	0
Public - Rural	78	10	68	0	0	0	0
Private	14	0	14	0	0	0	0
Total	**180**	**11**	**167**	**0**	**0**	**2**	**0**

Armenia

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), and special needs schools
- Within-school exclusions consisted of non-native language speakers

Sample Design

- Explicit stratification by region (10)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 60)
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Aragatsotn	8	0	8	0	0	0	0
Ararat	14	0	14	0	0	0	0
Armavir	14	0	14	0	0	0	0
Gegharkunik	12	0	12	0	0	0	0
Kotayk	16	0	16	0	0	0	0
Lori	12	0	12	0	0	0	0
Shirak	12	0	12	0	0	0	0
Syunik & Vayots Dzor	10	0	10	0	0	0	0
Tavush	8	0	8	0	0	0	0
Yerevan	44	0	44	0	0	0	0
Total	**150**	**0**	**150**	**0**	**0**	**0**	**0**

Australia

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), very remote schools, and special and non-mainstream schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by state or territory (8)
- Implicit stratification by school type (Catholic, government, independent), geographic location (metropolitan, provincial, remote), and socioeconomic index (low socioeconomic status, high socioeconomic status)
- Sampled one classroom per school. In tracked schools, classrooms were grouped according to the ability level of students prior to sampling and one classroom was sampled per class group.
- The TIMSS Grade 4 and Grade 8 samples were selected sequentially.
- The TIMSS sample at Grade 4 was selected by controlling the overlap with the PISA and the TIMSS Grade 8 samples using the Chowdhury approach
- Schools were oversampled at the state/territory level

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Australian Capital Territory	29	0	29	0	0	0	0
New South Wales	46	0	44	2	0	0	0
Northern Territory	14	0	14	0	0	0	0
Queensland	45	0	45	0	0	0	0
South Australia	41	0	41	0	0	0	0
Tasmania	31	0	30	0	0	1	0
Victoria	44	0	44	0	0	0	1
Western Australia	38	0	38	0	0	0	1
Total	**288**	**0**	**285**	**2**	**0**	**1**	**2**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), very remote schools, and special and non-mainstream schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by state or territory (8)
- Implicit stratification by school type (Catholic, government, independent), geographic location (metropolitan, provincial, remote), and socioeconomic index (low socioeconomic status, high socioeconomic status)
- Sampled one classroom per school. In tracked schools, classrooms were grouped according to the ability level of students prior to sampling and one classroom was sampled per class group.
- The TIMSS Grade 4 and Grade 8 samples were selected sequentially.
- The TIMSS sample at Grade 8 was selected by controlling the overlap with the PISA sample using the Chowdhury approach
- Schools were oversampled at the state/territory level

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Australian Capital Territory	30	0	30	0	0	0	0
New South Wales	45	0	43	2	0	0	0
Northern Territory	14	1	13	0	0	0	1
Queensland	47	0	47	0	0	0	0
South Australia	39	0	39	0	0	0	1
Tasmania	29	1	28	0	0	0	1
Victoria	45	0	45	0	0	0	0
Western Australia	37	0	37	0	0	0	0
Total	**286**	**2**	**282**	**2**	**0**	**0**	**3**

Austria

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by urbanization (area with more than 50,000 inhabitants, area with 50,000 inhabitants or less), achievement score (low, medium, high), and school size (small, large)
- Implicit stratification by region (9)
- Sampled two classrooms per school. In schools sampled for eTIMSS and bridge, one classroom sampled per assessment
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- The bridge sample was obtained using a combination of strategies. In the large school strata, the bridge sample was selected as a subset of the eTIMSS school sample and classes were randomly assigned to either the eTIMSS or bridge samples. In the small school strata, a distinct sample of schools was selected for the bridge sample. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.
- Austria provided a bridge sample for their own national analyses only since they did not have a trend with TIMSS 2015
- 36.4 % of students in the bridge sample were in schools selected for the eTIMSS sample

School Participation Status

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacements	Participating Schools: 2nd Replacements	Refusal Schools	Excluded Schools
Rural (no more than 50,000 inhabitants) - Low - Small	28	0	28	0	0	0	0
Rural (no more than 50,000 inhabitants) - Low - Large	12	0	12	0	0	0	0
Rural (no more than 50,000 inhabitants) - Medium - Small	38	2	36	0	0	0	0
Rural (no more than 50,000 inhabitants) - Medium - Large	12	0	12	0	0	0	0
Rural (no more than 50,000 inhabitants) - High - Small	32	0	31	0	0	1	0
Rural (no more than 50,000 inhabitants) - High - Large	10	0	10	0	0	0	0
Urban (more than 50,000 inhabitants) - Low - Small	9	0	9	0	0	0	0
Urban (more than 50,000 inhabitants) - Low - Large	17	0	17	0	0	0	0
Urban (more than 50,000 inhabitants) - Medium - Small	8	0	8	0	0	0	1
Urban (more than 50,000 inhabitants) - Medium - Large	12	0	12	0	0	0	0
Urban (more than 50,000 inhabitants) - High - Small	8	0	8	0	0	0	0
Urban (more than 50,000 inhabitants) - High - Large	10	0	10	0	0	0	0
Total	**196**	**2**	**193**	**0**	**0**	**1**	**1**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Rural (no more than 50,000 inhabitants) - Low - Large	12	0	12	0	0	0	0
Rural (no more than 50,000 inhabitants) - Medium - Small	38	2	36	0	0	0	0
Rural (no more than 50,000 inhabitants) - Medium - Large	12	0	12	0	0	0	0
Rural (no more than 50,000 inhabitants) - High - Small	32	0	31	0	0	1	0
Rural (no more than 50,000 inhabitants) - High - Large	10	0	10	0	0	0	0
Urban (more than 50,000 inhabitants) - Low - Small	9	0	9	0	0	0	0
Urban (more than 50,000 inhabitants) - Low - Large	17	0	17	0	0	0	0
Urban (more than 50,000 inhabitants) - Medium - Small	8	0	8	0	0	0	1
Urban (more than 50,000 inhabitants) - Medium - Large	12	0	12	0	0	0	0
Urban (more than 50,000 inhabitants) - High - Small	8	0	8	0	0	0	0
Urban (more than 50,000 inhabitants) - High - Large	10	0	10	0	0	0	0
Total	**196**	**2**	**193**	**0**	**0**	**1**	**1**

Azerbaijan

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, language of instruction other than Azerbaijani or Russian, and private schools
- Within-school exclusions consisted of students with intellectual disabilities, and students with functional disabilities

Sample Design

- Explicit stratification by language (Azerbaijani only, Russian or Azerbaijani and Russian), urbanization (urban, rural) within Azerbaijani only strata, and city (Baku, other) within urban stratum
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 120) and in schools where class grouping is applied
- Class group option was used in bilingual schools

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Azerbaijani only - Urban - Baku	30	1	26	2	0	1	0
Azerbaijani only - Urban - Other cities	40	0	39	1	0	0	0
Azerbaijani only - Rural	80	0	78	0	0	2	0
Russian, Azerbaijani and Russian	50	0	43	5	0	2	0
Total	**200**	**1**	**186**	**8**	**0**	**5**	**0**

Bahrain

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools, and schools with students taught in French
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private), and by governorate (4) and gender (girls, boys) within public schools
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 155)
- No overlap between Grade 4 and Grade 8 samples
- Schools or classes were used as variance estimation strata and classes or half classes were used to build jacknife replicates

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Public Muharraq - Girls	10	0	10	0	0	0	0
Public Muharraq - Boys	10	0	10	0	0	0	0
Public Capital - Girls	19	0	19	0	0	0	0
Public Capital - Boys	20	0	20	0	0	0	0
Public Northern - Girls	22	0	22	0	0	0	0
Public Northern - Boys	17	0	17	0	0	0	0
Public Southern - Girls	11	0	11	0	0	0	0
Public Southern - Boys	12	0	12	0	0	0	0
Private	64	0	64	0	0	0	0
Total	**185**	**0**	**185**	**0**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools, and schools with students taught in French
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private), and by governorate (4) and gender (girls, boys) within public schools
- No implicit stratification
- Sampled two classrooms per school
- No overlap between Grade 4 and Grade 8 samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Public Muharraq - Girls	6	0	6	0	0	0	0
Public Muharraq - Boys	6	0	6	0	0	0	0
Public Capital - Girls	8	0	8	0	0	0	0
Public Capital - Boys	8	0	8	0	0	0	0
Public Northern - Girls	8	0	8	0	0	0	0
Public Northern - Boys	9	0	9	0	0	0	0
Public Southern - Girls	7	0	7	0	0	0	0
Public Southern - Boys	10	0	10	0	0	0	0
Private	50	0	50	0	0	0	0
Total	**112**	**0**	**112**	**0**	**0**	**0**	**0**

Belgium (Flemish)

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), language other than Dutch, and special needs schools other than type 1, 3, 8, 9
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by region (6), school type (official, private) within Antwerpen region, and socioeconomic status (4). Special needs schools were grouped into one separate stratum.
- No implicit stratification
- Sampled two classrooms per school

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Antwerpen - Official - High SES	8	0	8	0	0	0	0
Antwerpen - Official - Medium SES	8	0	6	1	1	0	0
Antwerpen - Official - Low SES	8	0	7	1	0	0	0
Antwerpen - Private - High SES	8	0	5	1	2	0	0
Antwerpen - Private - Medium SES	8	0	3	0	3	2	0
Antwerpen - Private - Low SES	8	0	5	2	0	1	0
Brussels Hoofdstedelijk Gewest - Low and Medium SES	8	0	6	2	0	0	0
Limburg - High SES	8	0	5	2	0	1	0
Limburg - Medium SES	8	0	2	4	2	0	0

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Limburg - Low SES	8	0	4	1	2	1	0
Oost-Vlaanderen - High SES	8	0	7	1	0	0	0
Oost-Vlaanderen - Medium SES	8	0	6	1	1	0	0
Oost-Vlaanderen - Low SES	8	0	4	2	1	1	0
Vlaams-Brabant - High SES	8	0	5	2	1	0	0
Vlaams-Brabant - Medium SES	8	0	7	1	0	0	0
Vlaams-Brabant - Low SES	8	0	4	3	0	1	0
West-Vlaanderen - High SES	8	1	4	3	0	0	0
West-Vlaanderen - Medium SES	8	0	6	1	1	0	0
West-Vlaanderen - Low SES	8	0	5	3	0	0	0
Special Needs Schools	8	3	2	1	0	2	0
Total	**160**	**4**	**101**	**32**	**14**	**9**	**0**

Bosnia and Herzegovina

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, and international schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by region (12) and urbanization (rural, urban) within 3 larger regions
- Implicit stratification by urbanization (urban, rural) within two other larger regions
- Sampled two classrooms per school
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- A census of schools was taken in two small regions
- In census strata, schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Republika Srpska - Rural	10	0	10	0	0	0	0
Republika Srpska - Urban	36	0	36	0	0	0	0
Brcko District	8	0	8	0	0	0	0
Bosnian Podrinje Canton	6	0	6	0	0	0	0
Herzeg-Bosnia Canton	8	0	8	0	0	0	0
Herzegovina-Neretva Canton	8	0	8	0	0	0	0

School Participation Status (continued)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacements	Participating Schools: 2nd Replacements	Refusal Schools	Excluded Schools
Sarajevo Canton	20	0	20	0	0	0	0
Central Bosnia Canton - Rural	8	0	8	0	0	0	0
Central Bosnia Canton - Urban	9	0	9	0	0	0	0
Tuzla Canton - Rural	12	0	12	0	0	0	0
Tuzla Canton - Urban	8	0	8	0	0	0	0
Una-Sana Canton	12	0	12	0	0	0	0
Zenica-Doboj Canton - Rural	8	0	8	0	0	0	0
Zenica-Doboj Canton - Urban	10	0	10	0	0	0	0
Posavina Canton	7	0	7	0	0	0	0
West Herzegovina Canton	8	0	8	0	0	0	0
Total	**178**	**0**	**178**	**0**	**0**	**0**	**0**

Bulgaria

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (elementary, basic, general) and urbanization (capital, large cities, other)
- Implicit stratification by score (4)
- Sampled two classrooms in large schools (measure of size > 80)
- The Field Test and Main Data Collection TIMSS samples were selected sequentially. The TIMSS Main Data Collection sample was selected by controlling for the overlap with the TIMSS Field Test and TALIS samples using the Chowdhury approach.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Elementary School - Capital and Large Cities	8	0	7	1	0	0	0
Elementary School - Other	8	0	8	0	0	0	0
Basic School - Capital	10	0	10	0	0	0	0
Basic School - Large Cities	30	0	28	1	1	0	0
Basic School - Other	36	0	35	1	0	0	0
General School - Capital	15	0	15	0	0	0	0
General School - Large Cities	20	0	19	1	0	0	0
General School - Other	24	0	24	0	0	0	0
Total	**151**	**0**	**146**	**4**	**1**	**0**	**0**

Canada

Fourth Grade

Coverage and Exclusions

- Coverage is 79.3 percent. Coverage in Canada is restricted to students from the provinces of Alberta, Manitoba, Newfoundland, Ontario, and Quebec.
- School-level exclusions consisted of very small schools (measure of size < 10 in Quebec, measure of size < 6 in Ontario, Alberta, and Newfoundland, measure of size < 4 in Manitoba), special needs schools, First Nation schools and federal schools. French schools, non-ministry schools and remote or hard to access schools (in Newfoundland). Home schools (in Alberta and Manitoba), not funded schools (in Manitoba). International schools and school boards with special status (in Quebec).
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by provinces (5). Within the province of Alberta, explicit stratification was done by school system (French, English), school type (public, private, separate), and by school size (small, large). Within the province of Ontario, explicit stratification was done by language (English, French), school type (private, Catholic, public), and by school size (small, large) within Catholic and public schools. Within Quebec and Manitoba, explicit stratification was done by language (French, English), school type (public, private), and school size (small with less than three classes, large with three or more classes).
- Implicit stratification by region (6) in English public and Catholic schools explicit strata within Ontario
- Sampled two classrooms in large schools (measure of size > 40 in Ontario and measure of size > 75 in Quebec). In schools sampled for eTIMSS and bridge, one classroom sampled per assessment
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- All Alberta and Manitoba French schools were selected
- The bridge sample was obtained using a combination of strategies. In the large school strata, the bridge sample was selected as a subset of the eTIMSS sample and classes were randomly

assigned to either the eTIMSS or bridge samples. In the small school strata, a distinct sample of schools was selected for the bridge sample. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.

- 60.3 % of students in the bridge sample were in schools selected for the eTIMSS sample

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Newfoundland	100	2	91	0	0	7	0
Quebec - English - Public - Small	12	0	8	1	0	3	0
Quebec - English - Public - Large	14	0	13	1	0	0	0
Quebec - English - Private	8	0	6	0	0	2	0
Quebec - French - Public - Small	44	0	35	1	0	8	0
Quebec - French - Public - Large	84	0	70	4	0	10	0
Quebec - French - Private - Small	4	0	3	1	0	0	0
Quebec - French - Private - Large	6	0	5	0	0	1	0
Manitoba - English - Public - Small	78	1	76	0	0	1	0
Manitoba - English - Public - Large	58	0	55	0	0	3	0
Manitoba - English - Private	10	0	10	0	0	0	0
Manitoba - French - Public	19	0	17	0	0	2	0
Alberta - English - Public - Small	36	0	25	4	2	5	0
Alberta - English - Public - Large	68	1	46	5	3	13	2
Alberta - English - Private	7	0	2	3	0	2	1
Alberta - English - Separate - Small	16	1	11	1	1	2	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Alberta - English - Separate - Large	22	2	13	3	2	2	0
Alberta - French - Public	28	0	23	0	0	5	0
Ontario - English - Public - Small	24	1	23	0	0	0	0
Ontario - English - Public - Large	72	0	70	1	0	1	0
Ontario - English - Catholic - Small	18	0	18	0	0	0	0
Ontario - English - Catholic - Large	19	0	19	0	0	0	0
Ontario - Private	8	0	0	1	1	6	0
Ontario - French - Catholic & Public - Small	18	0	18	0	0	0	0
Ontario - French - Catholic & Public - Large	12	0	12	0	0	0	0
Total	**785**	**8**	**669**	**26**	**9**	**73**	**3**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Newfoundland	4	0	4	0	0	0	0
Quebec - English - Public - Small	2	0	0	0	0	2	0
Quebec - English - Public - Large	2	0	2	0	0	0	0
Quebec - English - Private	2	0	2	0	0	0	0
Quebec - French - Public - Small	6	0	6	0	0	0	0
Quebec - French - Public - Large	12	0	11	0	0	1	0
Quebec - French - Private - Small	2	0	2	0	0	0	0

Bridge Sample School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Quebec - French - Private - Large	2	0	1	0	0	1	0
Manitoba - English - Public - Small	2	0	2	0	0	0	0
Manitoba - English - Public - Large	2	0	2	0	0	0	0
Manitoba - English - Private	2	0	2	0	0	0	0
Manitoba - French - Public	2	0	2	0	0	0	0
Alberta - English - Public - Small	4	0	1	1	0	2	0
Alberta - English - Public - Large	6	0	2	2	0	2	0
Alberta - English - Private	1	0	0	0	1	0	1
Alberta - English - Separate - Small	2	0	2	0	0	0	0
Alberta - English - Separate - Large	2	1	1	0	0	0	0
Alberta - French - Public	2	0	2	0	0	0	0
Ontario - English - Public - Small	6	0	6	0	0	0	0
Ontario - English - Public - Large	16	0	15	1	0	0	0
Ontario - English - Catholic - Small	4	0	4	0	0	0	0
Ontario - English - Catholic - Large	4	0	4	0	0	0	0
Ontario - Private	2	0	0	1	0	1	0
Ontario - French - Catholic & Public - Small	2	0	2	0	0	0	0
Ontario - French - Catholic & Public - Large	2	0	2	0	0	0	0
Total	**93**	**1**	**77**	**5**	**1**	**9**	**1**

Chile

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, and geographically inaccessible schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 4'/'Grade 4 and Grade 8' schools, school type (public, private subsidized, private), and urbanization (rural, urban)
- Implicit stratification by national assessment score level (4)
- Sampled one classroom per school
- The Field Test and Main Data Collection TIMSS samples were selected sequentially. The TIMSS Main Data Collection sample was selected by controlling for the overlap with the TIMSS Field Test, ICILS, and PISA samples using the Chowdhury approach.
- Private schools were oversampled
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 - Urban	8	1	7	0	0	0	0
Grade 4 - Rural	8	0	7	1	0	0	0
Grade 4 & Grade 8 - Public - Urban	40	0	37	3	0	0	0
Grade 4 & Grade 8 - Public - Rural	10	0	10	0	0	0	0
Grade 4 & Grade 8 - Private Subsidized - Urban	71	1	61	8	1	0	0

School Participation Status (continued)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 & Grade 8 - Private Subsidized - Rural	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Private Paid	29	0	21	4	1	3	0
Total	**174**	**2**	**151**	**16**	**2**	**3**	**0**

Bridge Sample School Participation Status

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 - Urban	4	0	4	0	0	0	0
Grade 4 - Rural	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Public - Urban	12	0	12	0	0	0	0
Grade 4 & Grade 8 - Public - Rural	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Private Subsidized - Urban	22	0	22	0	0	0	0
Grade 4 & Grade 8 - Private Subsidized - Rural	4	0	3	1	0	0	0
Grade 4 & Grade 8 - Private Paid	8	0	8	0	0	0	0
Total	**58**	**0**	**57**	**1**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, and geographically inaccessible schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 8'/'Grade 4 and Grade 8' schools, school type (public, private subsidized, private), and urbanization (rural, urban)
- Implicit stratification by national assessment score level (4)
- Sampled one classroom per school
- The Field Test and Main Data Collection TIMSS samples were selected sequentially. The TIMSS Main Data Collection sample was selected by controlling for the overlap with the TIMSS Field Test, ICILS, and PISA samples using the Chowdhury approach.
- Private schools were oversampled
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8	11	0	11	0	0	0	0
Grade 4 & Grade 8 - Public - Urban School	40	0	37	3	0	0	0
Grade 4 & Grade 8 - Public - Rural School	10	0	10	0	0	0	0
Grade 4 & Grade 8 - Private-Subsidized - Urban School	71	2	60	8	1	0	0
Grade 4 & Grade 8 - Private-Subsidized - Rural School	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Private	29	0	21	4	1	3	0
Total	**169**	**2**	**147**	**15**	**2**	**3**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8	6	0	4	2	0	0	0
Grade 4 & Grade 8 - Public - Urban School	12	0	12	0	0	0	0
Grade 4 & Grade 8 - Public - Rural School	4	0	4	0	0	0	0

Bridge Sample School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 & Grade 8 - Private-Subsidized - Urban School	22	0	22	0	0	0	0
Grade 4 & Grade 8 - Private-Subsidized - Rural School	4	0	3	1	0	0	0
Grade 4 & Grade 8 - Private	8	0	8	0	0	0	0
Total	**56**	**0**	**53**	**3**	**0**	**0**	**0**

Chinese Taipei

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), and schools that do not follow the national curriculum
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by urbanization (metropolitan area, city area, developing city area, rural and remote area), region (north, other), and school size (small, large)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 290 for eTIMSS, measure of size > 68 for bridge). In schools sampled for eTIMSS and bridge, one classroom sampled per assessment
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- The bridge sample was obtained using a combination of strategies. In the large school strata, the bridge sample was selected as a subset of the eTIMSS sample and classes were randomly assigned to either the eTIMSS or bridge samples. In the small school strata, a distinct sample of schools was selected for the bridge sample. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.
- 85.6 % of students in the bridge sample were in schools selected for the eTIMSS sample

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Metropolitan Area - North	24	0	20	4	0	0	0
Metropolitan Area - Other	10	0	10	0	0	0	0
City Area - North	24	0	23	1	0	0	0
City Area - Other	22	0	22	0	0	0	0
Developing City Area - North	22	0	22	0	0	0	0
Developing City Area - Other - Large	20	0	20	0	0	0	0
Developing City Area - Other - Small	8	0	8	0	0	0	0
Rural and Remote Area - North - Large	6	0	5	1	0	0	0
Rural and Remote Area - North - Small	4	0	4	0	0	0	0
Rural and Remote Area - Other - Large	8	0	7	1	0	0	0
Rural and Remote Area - Other - Small	15	0	14	0	0	1	0
Total	**163**	**0**	**155**	**7**	**0**	**1**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Metropolitan Area - North	10	0	8	2	0	0	0
Metropolitan Area - Other	4	0	4	0	0	0	0
City Area - North	10	0	9	1	0	0	0
City Area - Other	8	0	8	0	0	0	0
Developing City Area - North	8	0	8	0	0	0	0
Developing City Area - Other - Large	8	0	8	0	0	0	0
Developing City Area - Other - Small	4	0	4	0	0	0	0

Bridge Sample School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Rural and Remote Area - North - Large	4	0	4	0	0	0	0
Rural and Remote Area - North - Small	2	0	2	0	0	0	0
Rural and Remote Area - Other - Large	4	0	3	1	0	0	0
Rural and Remote Area - Other - Small	6	0	6	0	0	0	0
Total	**68**	**0**	**64**	**4**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), and schools that do not follow the national curriculum
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by urbanization (metropolitan area, city area, developing city area, rural and remote area), region (north, other), and school size (small, large)
- Implicit stratification by performance (5)
- Sampled one classroom per school
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- Rural and remote schools were oversampled.
- The bridge sample was obtained using a combination of strategies. In the large school strata, the bridge sample was selected as a subset of the eTIMSS sample and classes were randomly assigned to either the eTIMSS or bridge samples. In the small school strata, a distinct sample

of schools was selected for the bridge sample. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.

- 98 % of students in the bridge sample were in schools selected for the eTIMSS sample

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8 - Metropolitan Area - North	23	0	22	1	0	0	0
Grade 8 - Metropolitan Area - Other	13	0	13	0	0	0	0
Grade 8 - City Area - North	24	0	23	0	0	1	0
Grade 8 - City Area - Other	24	0	24	0	0	0	0
Grade 8 - Developing City Area - North	18	1	17	0	0	0	0
Grade 8 - Developing City Area - Other	30	0	30	0	0	0	0
Grade 8 - Rural and Remote Area - North	19	0	18	1	0	0	0
Grade 8 - Rural and Remote Area - Other - Large	44	0	42	1	0	1	0
Grade 8 - Rural and Remote Area - Other - Small	11	0	11	0	0	0	0
Total	**206**	**1**	**200**	**3**	**0**	**2**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8 - Metropolitan Area - North	8	0	8	0	0	0	0
Grade 8 - Metropolitan Area - Other	4	0	4	0	0	0	0
Grade 8 - City Area - North	8	0	8	0	0	0	0

Bridge Sample School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8 - City Area - Other	8	0	8	0	0	0	0
Grade 8 - Developing City Area - North	6	0	6	0	0	0	0
Grade 8 - Developing City Area - Other	10	0	10	0	0	0	0
Grade 8 - Rural and Remote Area - North	3	0	3	0	0	0	0
Grade 8 - Rural and Remote Area - Other - Large	8	0	8	0	0	0	0
Grade 8 - Rural and Remote Area - Other - Small	2	0	2	0	0	0	0
Total	**57**	**0**	**57**	**0**	**0**	**0**	**0**

Croatia

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 3), and private schools
- Within-school exclusions consisted of students with intellectual disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (mother/single building, satellite), and by grouped regions (6) and school size (small, large) within mother/single building strata
- Implicit stratification by urbanization (urban, rural)
- Sampled two classrooms whenever possible. In schools selected for both the eTIMSS and Bridge samples, two classrooms selected for eTIMSS and one classroom selected for Bridge sample whenever possible
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- The bridge sample was obtained using a combination of strategies. In the large school strata, the bridge sample was selected as a subset of the eTIMSS school sample and classes were randomly assigned to either the eTIMSS or bridge samples. In the small school strata, a distinct sample of schools was selected for the bridge sample. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.
- 48.6 % of students in the bridge sample were in schools selected for the eTIMSS sample

School Participation Status

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacements	Participating Schools: 2nd Replacements	Refusal Schools	Excluded Schools
Mother/Single Building School - Central Croatia - Small	10	0	10	0	0	0	0
Mother/Single Building School - Central Croatia - Large	14	0	13	1	0	0	0
Mother/Single Building School - Eastern Croatia - Small	10	0	9	0	0	1	0
Mother/Single Building School - Eastern Croatia - Large	8	0	8	0	0	0	0
Mother/Single Building School - Northern Croatia - Small	8	0	8	0	0	0	0
Mother/Single Building School - Northern Croatia - Large	8	0	8	0	0	0	0
Mother/Single Building School - Western Croatia - Small	11	0	10	1	0	0	0
Mother/Single Building School - Western Croatia - Large	9	0	8	0	0	1	0
Mother/Single Building School - Southern Croatia - Small	8	0	7	0	0	1	0
Mother/Single Building School - Southern Croatia - Large	16	0	16	0	0	0	0
Mother/Single Building School - City of Zagreb - Small	9	0	9	0	0	0	0
Mother/Single Building School - City of Zagreb - Large	24	0	22	1	0	1	0
Satellite Schools	24	1	22	0	0	1	0
Total	**159**	**1**	**150**	**3**	**0**	**5**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Mother/Single Building School - Central Croatia - Small	4	0	4	0	0	0	0
Mother/Single Building School - Central Croatia - Large	6	0	6	0	0	0	0
Mother/Single Building School - Eastern Croatia - Small	6	0	6	0	0	0	0
Mother/Single Building School - Eastern Croatia - Large	4	0	4	0	0	0	0
Mother/Single Building School - Northern Croatia - Small	4	0	4	0	0	0	0
Mother/Single Building School - Northern Croatia - Large	4	0	4	0	0	0	0
Mother/Single Building School - Western Croatia - Small	4	0	4	0	0	0	0
Mother/Single Building School - Western Croatia - Large	4	0	3	0	0	1	0
Mother/Single Building School - Southern Croatia - Small	4	0	4	0	0	0	0
Mother/Single Building School - Southern Croatia - Large	8	0	8	0	0	0	0
Mother/Single Building School - City of Zagreb - Small	4	0	4	0	0	0	0
Mother/Single Building School - City of Zagreb - Large	12	0	12	0	0	0	0
Satellite Schools	12	0	11	0	0	1	0
Total	**76**	**0**	**74**	**0**	**0**	**2**	**0**

Cyprus

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, Turkish Occupied Area, and language of instruction other than Greek or English
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private), curriculum (national curriculum, other), and district (4)
- Implicit stratification by urbanization (urban, rural)
- Sampled three classrooms whenever possible in large schools (measure of size > 65)
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Private with other curriculum	12	1	10	1	0	0	0
Public and private with national curriculum - Famagusta-Larnaca	34	0	34	0	0	0	0
Public and private with national curriculum - Limassol	38	0	38	0	0	0	0
Public and private with national curriculum - Nicosia	52	0	52	0	0	0	0
Public and private with national curriculum - Paphos	16	0	16	0	0	0	0
Total	**152**	**1**	**150**	**1**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), Turkish Occupied Area, and language of instruction other than Greek or English
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private), curriculum (national curriculum, other), and district (4)
- Implicit stratification by urbanization (urban, rural)
- Sampled three classrooms whenever possible in large schools (measure of size > 120)
- All Grade 8 schools were selected for the Main Data Collection
- Schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Private with other curriculum	24	1	23	0	0	0	0
Public and private with national curriculum - Famagusta-Larnaca	15	0	15	0	0	0	0
Public and private with national curriculum - Limassol	23	0	23	0	0	0	0
Public and private with national curriculum - Nicosia	28	0	28	0	0	0	0
Public and private with national curriculum - Paphos	9	0	9	0	0	0	0
Total	**99**	**1**	**98**	**0**	**0**	**0**	**0**

Czech Republic

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), special needs schools, and Polish language schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by region (14). One additional stratum created for schools with no Grade 4 students on the frame but expected to have some during the Main Data Collection.
- No implicit stratification
- Sampled two classrooms per school
- The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Praha	16	0	16	0	0	0	0
Stredoceský	20	0	20	0	0	0	0
Plzenský	8	0	8	0	0	0	0
Karlovarský	7	0	7	0	0	0	1
Ústecký	12	0	12	0	0	0	0
Jihoceský	8	0	7	1	0	0	0
Liberecký	8	0	8	0	0	0	0
Královéhradecký	8	0	8	0	0	0	0
Pardubický	8	0	8	0	0	0	0
Vysocina	8	0	8	0	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Jihomoravský	15	0	15	0	0	0	1
Olomoucký	8	0	8	0	0	0	0
Moravskoslezský	16	0	16	0	0	0	0
Zlínský	8	0	8	0	0	0	0
Empty Schools	4	2	2	0	0	0	0
Total	**154**	**2**	**151**	**1**	**0**	**0**	**2**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Praha	4	0	4	0	0	0	0
Stredoceský	6	1	5	0	0	0	0
Plzenský	4	0	4	0	0	0	0
Karlovarský	4	0	4	0	0	0	0
Ústecký	4	0	4	0	0	0	0
Jihoceský	4	0	4	0	0	0	0
Liberecký	4	0	4	0	0	0	0
Královéhradecký	4	0	4	0	0	0	0
Pardubický	4	0	4	0	0	0	0
Vysocina	4	0	4	0	0	0	0
Jihomoravský	4	0	4	0	0	0	0
Olomoucký	4	0	4	0	0	0	0
Moravskoslezský	4	0	4	0	0	0	0
Zlínský	3	0	3	0	0	0	1
Empty Schools	2	0	2	0	0	0	0
Total	**59**	**1**	**58**	**0**	**0**	**0**	**1**

Denmark

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private), and school size (small, large) within public schools
- No implicit stratification
- Sampled two classrooms in large schools that participate in eTIMSS or both eTIMSS and bridge (measure of size > 85). Sampled two classrooms in large schools that participate in bridge only (measure of size > 44).
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- The bridge sample was obtained using a combination of strategies. In the large school stratum, the bridge sample was selected as a subset of the eTIMSS school sample and classes were randomly assigned to either the eTIMSS or bridge samples. In the small school and private school strata, a distinct sample of schools was selected for the bridge sample. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.
- 37.2 % of students in the bridge sample were in schools selected for the eTIMSS sample

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Private	30	1	12	7	4	6	0
Public - Small	80	0	63	15	1	1	0
Public - Large	65	0	48	12	4	1	0
Total	**175**	**1**	**123**	**34**	**9**	**8**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Private	10	0	6	2	1	1	0
Public - Small	28	0	17	7	4	0	0
Public - Large	24	0	17	4	3	0	0
Total	**62**	**0**	**40**	**13**	**8**	**1**	**0**

Egypt

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 20), Al-Azhar schools, special needs schools, and sports prep schools
- Within-school exclusions consisted of students with intellectual disabilities

Sample Design

- Explicit stratification by region (3), school type (4) and school gender (3).
- Implicit stratification by school shift (4) within governmental schools
- Sampled one classroom per school
- The Field Test and Main Data Collection school samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Capital - Governmental Schools - Girls	10	0	10	0	0	0	0
Capital - Governmental Schools - Boys	10	0	10	0	0	0	0
Capital - Governmental Schools - Mixed	8	0	8	0	0	0	0
North - Governmental Schools - Girls	11	0	11	0	0	0	1
North - Governmental Schools - Boys	12	0	12	0	0	0	0
North - Governmental Schools - Mixed	40	0	39	1	0	0	0
South - Governmental Schools - Girls	8	0	8	0	0	0	0
South - Governmental Schools - Boys	8	0	8	0	0	0	0

School Participation Status (continued)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacements	Participating Schools: 2nd Replacements	Refusal Schools	Excluded Schools
South - Governmental Schools - Mixed	30	0	30	0	0	0	0
Private Funded Schools (without fees)	11	0	11	0	0	0	1
Private Schools (with fees)	12	0	12	0	0	0	0
Private Language Schools	9	0	9	0	0	0	3
Total	**169**	**0**	**168**	**1**	**0**	**0**	**5**

England

Fifth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 10), and special schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (academy, all other state funded, independent), and attainment level (5)
- Implicit stratification by attainment level (7)
- Sampled two classrooms in large schools selected for eTIMSS (measure of size > 90) and in large schools selected for bridge (measure of size > 65)
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample for Grade 5 was selected by controlling for the overlap with the Main Data Collection sample at Grade 8 and the Field Test sample using the Chowdhury approach.
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
All Other State-Funded - Low	14	0	12	0	0	2	0
All Other State-Funded - Low/Middle	22	0	21	0	0	1	0
All Other State-Funded - Middle/High	24	0	20	2	0	2	0
All Other State-Funded - High	20	0	17	1	0	2	0
All Other State-Funded - Middle and N.A.	24	0	21	0	0	3	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Academy - Low and Low/Middle	16	0	14	2	0	0	0
Academy - Middle and N.A.	8	0	8	0	0	0	0
Academy - Mid-dle/ High and High	14	0	11	2	0	1	0
Independent	8	0	5	3	0	0	0
Total	**150**	**0**	**129**	**10**	**0**	**11**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
All Other State-Funded - Low	4	0	4	0	0	0	0
All Other State-Funded - Low/Middle	8	0	5	2	0	1	0
All Other State-Funded - Middle/High	8	0	7	0	0	1	0
All Other State-Funded - High	6	0	6	0	0	0	0
All Other State-Funded - Middle and N.A.	8	0	7	0	0	1	0
Academy - Low and Low/Middle	6	0	4	1	0	1	0
Academy - Middle and N.A.	4	0	3	0	0	1	0
Academy - Mid-dle/ High and High	4	0	4	0	0	0	0
Independent	4	0	2	0	1	1	0
Total	**52**	**0**	**42**	**3**	**1**	**6**	**0**

Ninth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 20), and special schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (academy, all other state funded, independent), and attainment level (5)
- Implicit stratification by attainment level (7)
- Sampled two classrooms in large schools selected for eTIMSS (measure of size > 245) and in large schools selected for bridge (measure of size > 200)
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample for Grade 9 was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
All Other State Funded - Middle/High	12	0	9	3	0	0	0
All Other State Funded - High	8	0	8	0	0	0	0
All Other State Funded - Low and Low/Middle	10	0	8	1	0	1	0
All Other State Funded - Middle and N.A.	16	0	14	2	0	0	0
Academy - Mid-dle/ High	26	0	24	0	0	2	0
Academy - High	22	0	18	1	0	3	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Academy - Low and Low/Middle	21	0	17	1	0	3	0
Academy - Middle and N.A.	26	0	22	0	0	4	0
Independent	10	0	5	3	0	2	0
Total	**151**	**0**	**125**	**11**	**0**	**15**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
All Other State Funded - Middle/High	4	0	4	0	0	0	0
All Other State Funded - High	4	0	4	0	0	0	0
All Other State Funded - Low and Low/Middle	4	0	3	0	0	1	0
All Other State Funded - Middle and N.A.	4	0	3	1	0	0	0
Academy - Mid-dle/ High	8	0	5	3	0	0	0
Academy - High	8	0	7	0	0	1	0
Academy - Low and Low/Middle	6	0	6	0	0	0	0
Academy - Middle and N.A.	8	0	7	1	0	0	0
Independent	4	0	2	1	0	1	0
Total	**50**	**0**	**41**	**6**	**0**	**3**	**0**

Finland

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, and language schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by language (Finnish, Swedish), and major region (4) and urbanization (urban/semi-urban, rural) within Finnish schools
- Implicit stratification by regional state administrative agency (6)
- Sampled two classrooms per school
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection was selected by controlling for the overlap with the Field Test sample and Main Data Collection Grade 8 sample using the Chowdhury approach.
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Finnish Speaking - Helsinki/Uusimaa	42	1	40	1	0	0	0
Finnish Speaking - Southern - Urban and Semi-Urban	24	0	24	0	0	0	0
Finnish Speaking - Southern - Rural	8	0	8	0	0	0	0
Finnish Speaking - Western - Urban and Semi-Urban	31	0	31	0	0	0	0

School Participation Status (continued)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacements	Participating Schools: 2nd Replacements	Refusal Schools	Excluded Schools
Finnish Speaking - Western - Rural	8	0	8	0	0	0	0
Finnish Speaking - Northern & Eastern - Urban and Semi-Urban	28	0	28	0	0	0	0
Finnish Speaking - Northern & Eastern - Rural	8	0	8	0	0	0	0
Swedish Speaking	10	0	10	0	0	0	0
Total	**159**	**1**	**157**	**1**	**0**	**0**	**0**

Bridge Sample School Participation Status

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacements	Participating Schools: 2nd Replacements	Refusal Schools	Excluded Schools
Finnish Speaking - Helsinki/Uusimaa	18	0	17	1	0	0	0
Finnish Speaking - Southern - Urban and Semi-Urban	12	0	12	0	0	0	0
Finnish Speaking - Southern - Rural	4	1	3	0	0	0	0
Finnish Speaking - Western - Urban and Semi-Urban	14	0	14	0	0	0	0
Finnish Speaking - Western - Rural	4	0	4	0	0	0	0
Finnish Speaking - Northern & Eastern - Urban and Semi-Urban	12	0	12	0	0	0	0
Finnish Speaking - Northern & Eastern - Rural	4	0	4	0	0	0	0
Swedish Speaking	4	0	4	0	0	0	0
Total	**72**	**1**	**70**	**1**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 6), special needs schools, and language schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by language (Finnish, Swedish), and major region (4) and urbanization (urban/semi-urban, rural) within Finnish schools
- Implicit stratification by regional state administrative agency (6)
- Sampled two classrooms per school
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- No bridge sample required at Grade 8 as they did not participate in TIMSS 2015

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Finnish Speaking - Helsinki/Uusimaa	39	0	39	0	0	0	0
Finnish Speaking - Southern - Urban and Semi-Urban	24	2	22	0	0	0	0
Finnish Speaking - Southern - Rural	9	0	9	0	0	0	0
Finnish Speaking - Western - Urban and Semi-Urban	30	1	29	0	0	0	0
Finnish Speaking - Western - Rural	8	0	8	0	0	0	0
Finnish Speaking - Northern & Eastern - Urban and Semi-Urban	28	1	27	0	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Finnish Speaking - Northern & Eastern - Rural	10	0	10	0	0	0	0
Swedish Speaking	10	0	10	0	0	0	0
Total	**158**	**4**	**154**	**0**	**0**	**0**	**0**

France

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 3), special needs schools, overseas territories, Mayotte, and private schools without a contract
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public – other, public – priority education zone, private)
- No implicit stratification
- Sampled two classrooms per school
- No overlap between Grade 4 and Grade 8 schools
- The Main Data Collection sample was selected by controlling for the overlap with the Field Test using the Chowdhury approach.
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Public - Other	108	1	107	0	0	0	0
Public - Priority Education Zone	24	0	24	0	0	0	0
Private	24	0	24	0	0	0	0
Total	**156**	**1**	**155**	**0**	**0**	**0**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Public - Other	42	1	40	0	0	1	0
Public - Priority Education Zone	10	0	10	0	0	0	0
Private	10	0	10	0	0	0	0
Total	**62**	**1**	**60**	**0**	**0**	**1**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, overseas territories, Mayotte, and private schools without a contract
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public – other, public – priority education zone, private)
- No implicit stratification
- Sampled two classrooms per school
- No overlap between Grade 4 and Grade 8 schools
- The Main Data Collection sample was selected by controlling for the overlap with the Field Test using the Chowdhury approach.
- No bridge sample required at Grade 8 as they did not participate in TIMSS 2015

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Public-Priority Educa-tion Zone	24	0	24	0	0	0	0
Public-Other	94	0	94	0	0	0	0
Private	32	0	32	0	0	0	0
Total	**150**	**0**	**150**	**0**	**0**	**0**	**0**

Georgia

Fourth Grade

Coverage and Exclusions

- Coverage is 92 percent. Coverage in Georgia is restricted to students taught in Georgian.
- School-level exclusions consisted of very small schools (measure of size < 5), and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 4'/'Grade 4 and Grade 8' schools, region (Tbilisi, other), and Mathematics average score (low, high, N.A.)
- Implicit stratification by urbanization (town, village), and school type (public, private)
- Sampled two classrooms in large schools (measure of size > 70)
- Grade 4 and Grade 8 school samples were selected simultaneously with maximum overlap
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 only	8	1	6	0	0	1	0
Grade 4 & Grade 8 - Missing Average Math Score	9	0	9	0	0	0	0
Grade 4 & Grade 8 - Tbilisi - Low Average Math Score	7	0	7	0	0	0	0
Grade 4 & Grade 8 - Tbilisi - High Average Math Score	48	0	45	3	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 & Grade 8 - Other Region - Low Average Math Score	36	0	36	0	0	0	0
Grade 4 & Grade 8 - Other Region - High Average Math Score	50	0	48	0	0	2	0
Total	**158**	**1**	**151**	**3**	**0**	**3**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 only	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Missing Average Math Score	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Tbilisi - Low Average Math Score	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Tbilisi - High Average Math Score	16	0	16	0	0	0	0
Grade 4 & Grade 8 - Other Region - Low Average Math Score	12	0	12	0	0	0	0
Grade 4 & Grade 8 - Other Region - High Average Math Score	18	0	18	0	0	0	0
Total	**58**	**0**	**58**	**0**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 91.3 percent. Coverage in Georgia is restricted to students taught in Georgian.
- School-level exclusions consisted of very small schools (measure of size < 5), and special needs schools

- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 8'/'Grade 4 and Grade 8' schools, region (Tbilisi, other), and Mathematics average score (low, high, N.A.)
- Implicit stratification by urbanization (town, village), and school type (public, private)
- Sampled two classrooms in large schools (measure of size > 95)
- Grade 4 and Grade 8 school samples were selected simultaneously with maximum overlap
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Missing Achievement score	9	1	8	0	0	0	0
Grade 4 & Grade 8 - Tbilisi - Low Achievement	7	0	7	0	0	0	0
Grade 4 & Grade 8 - Tbilisi - High Achievement	48	0	38	3	0	7	0
Grade 4 & Grade 8 - Other - Low Achievement	36	0	36	0	0	0	0
Grade 4 & Grade 8 - Other - High Achievement	50	0	45	0	0	5	0
Total	**158**	**1**	**142**	**3**	**0**	**12**	**0**

Bridge Sample School Participation Status

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacements	Participating Schools: 2nd Replacements	Refusal Schools	Excluded Schools
Grade 8	4	0	3	0	0	1	0
Grade 4 & Grade 8 - Missing Achievement score	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Tbilisi - Low Achievement	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Tbilisi - High Achievement	16	0	12	0	0	4	0
Grade 4 & Grade 8 - Other - Low Achievement	12	0	12	0	0	0	0
Grade 4 & Grade 8 - Other - High Achievement	18	0	18	0	0	0	0
Total	**58**	**0**	**53**	**0**	**0**	**5**	**0**

Germany

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (regular, special needs). Within regular school, explicit stratification by socioeconomic status estimated by the percentage of migrants (low, medium, high), and school size (small, large).
- No implicit stratification
- Sampled one classroom per school. In schools sampled for eTIMSS and bridge, one classroom sampled per assessment
- The bridge sample was obtained using a combination of strategies. In the large school strata, the bridge sample was selected as a subset of the eTIMSS school sample and classes were randomly assigned to either the eTIMSS or bridge samples. In the small school strata, a distinct sample of schools was selected for the bridge sample. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.
- 44.9 % of students in the bridge sample were in schools selected for the eTIMSS sample

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Regular Schools - Very low percentage of im-migrants	24	0	19	3	2	0	0
Regular Schools - Low percentage of immi-grants - Small	50	0	50	0	0	0	0
Regular Schools - Low percentage of immi-grants - Large	50	0	50	0	0	0	0
Regular Schools - Me-dium percentage of immigrants - Small	18	0	18	0	0	0	0
Regular Schools - Me-dium percentage of immigrants - Large	30	0	30	0	0	0	0
Regular Schools - High percentage of immi-grants - Small	10	1	9	0	0	0	0
Regular Schools - High percentage of immi-grants - Large	14	0	14	0	0	0	0
Special Needs Schools	10	2	8	0	0	0	0
Total	**206**	**3**	**198**	**3**	**2**	**0**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Regular Schools - Very low percentage of im-migrants	10	0	10	0	0	0	0
Regular Schools - Low percentage of immi-grants - Small	20	0	18	2	0	0	0
Regular Schools - Low percentage of immi-grants - Large	18	0	18	0	0	0	0
Regular Schools - Me-dium percentage of immigrants - Small	6	0	6	0	0	0	0

Bridge Sample School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Regular Schools - Medium percentage of immigrants - Large	12	0	12	0	0	0	0
Regular Schools - High percentage of immi-grants - Small	4	0	4	0	0	0	0
Regular Schools - High percentage of immi-grants - Large	6	0	6	0	0	0	0
Special Needs Schools	2	0	2	0	0	0	0
Total	**78**	**0**	**76**	**2**	**0**	**0**	**0**

Hong Kong SAR

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 3), special needs schools, Japanese school, and remote school
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school finance type (5)
- No implicit stratification
- Sampled one classroom per school. One additional classroom selected in schools sampled for the bridge
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- The bridge sample was selected as a subset of the eTIMSS school sample and classes were randomly assigned to either the eTIMSS or bridge samples. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.
- 97.9 % of students in the bridge sample were in schools selected for the eTIMSS sample

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Aided	113	0	82	18	6	7	0
Direct Subsidy	10	0	7	2	0	1	0
Government	12	0	12	0	0	0	0
Private	12	0	6	1	1	4	0
Non-Local	12	0	2	2	0	8	0
Total	**159**	**0**	**109**	**23**	**7**	**20**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Aided	46	0	32	8	2	4	0
Direct Subsidy	4	0	3	0	0	1	0
Government	4	0	4	0	0	0	0
Private	6	0	2	1	0	3	0
Non-Local	6	0	2	0	0	4	0
Total	**66**	**0**	**43**	**9**	**2**	**12**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools, and Japanese school
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school finance type (4)
- Implicit stratification by other school characteristic (3)
- Sampled one classroom per school. One additional classroom selected in schools sampled for the bridge
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- The bridge sample was selected as a subset of the eTIMSS school sample and classes were randomly assigned to either the eTIMSS or bridge samples. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.
- 100 % of students in the bridge sample were in schools selected for the eTIMSS sample

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Aided	112	0	83	15	4	10	0
Direct Subsidy	22	0	16	3	0	3	0
Government	12	0	10	0	0	2	0
Private	12	0	3	2	0	7	0
Non-Local	158	0	112	20	4	22	0
Total	**159**	**0**	**109**	**23**	**7**	**20**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Aided	45	0	31	8	1	5	0
Direct Subsidy	10	0	6	2	0	2	0
Government	4	0	4	0	0	0	0
Private	4	0	0	2	0	2	0
Non-Local	63	0	41	12	1	9	0
Total	**66**	**0**	**43**	**9**	**2**	**12**	**0**

Hungary

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools, and schools with students taught in foreign language
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 4 only'/'Grade 4 and Grade 8' schools, type of community (capital, county town, town, rural area) and national assessment score (low, medium, high) within 'Grade 4 and Grade 8' stratum
- No implicit stratification
- Sampled two classrooms per school
- Grade 4 and Grade 8 samples were selected simultaneously with maximum overlap
- The Main Data Collection school samples for Grade 4 and Grade 8 were selected by controlling for the overlap with the Field test samples using the Chowdhury approach
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4	20	1	17	0	2	0	0
Grade 4 & Grade 8 - Capital - High Performance	13	0	10	2	1	0	0
Grade 4 & Grade 8 - Capital - Low or Medium Performance	10	0	9	1	0	0	0
Grade 4 & Grade 8 - County Town - High Performance	10	0	9	1	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 & Grade 8 - County Town - Low or Medium Performance	14	0	14	0	0	0	0
Grade 4 & Grade 8 - Town - Low Performance	14	0	13	0	1	0	0
Grade 4 & Grade 8 - Town - Medium Performance	26	0	25	0	1	0	0
Grade 4 & Grade 8 - Town - High Performance	8	0	6	1	0	1	0
Grade 4 & Grade 8 - Rural Area - Low Performance	16	0	16	0	0	0	0
Grade 4 & Grade 8 - Rural Area - Medium Performance	12	0	12	0	0	0	0
Grade 4 & Grade 8 - Rural Area - High Performance	8	0	8	0	0	0	0
Total	**151**	**1**	**139**	**5**	**5**	**1**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4	6	1	5	0	0	0	0
Grade 4 & Grade 8 - Capital - High Performance	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Capital - Low or Medium Performance	4	0	4	0	0	0	0
Grade 4 & Grade 8 - County Town - High Performance	4	0	4	0	0	0	0
Grade 4 & Grade 8 - County Town - Low or Medium Performance	4	0	4	0	0	0	0

Bridge Sample School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 & Grade 8 - Town - Low Performance	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Town - Medium Performance	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Town - High Performance	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Rural Area - Low Performance	6	0	6	0	0	0	0
Grade 4 & Grade 8 - Rural Area - Medium Performance	4	1	3	0	0	0	0
Grade 4 & Grade 8 - Rural Area - High Performance	4	0	4	0	0	0	0
Total	**52**	**2**	**50**	**0**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools, and schools with students taught in foreign language
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 8 only'/'Grade 4 and Grade 8' schools, type of community (capital, county town, town, rural area) and national assessment score (low, medium, high) within 'Grade 4 and Grade 8' stratum
- No implicit stratification

- Sampled two classrooms per school
- Grade 4 and Grade 8 samples were selected simultaneously with maximum overlap
- The Main Data Collection school samples for Grade 4 and Grade 8 were selected by controlling for the overlap with the Field test samples using the Chowdhury approach
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8	24	0	24	0	0	0	0
Grade 4 & Grade 8 - Capital - High	13	0	10	2	1	0	0
Grade 4 & Grade 8 - Capital - Low or Medium	10	0	9	1	0	0	0
Grade 4 & Grade 8 - County town - High	10	0	9	1	0	0	0
Grade 4 & Grade 8 - County town - Low or Medium	14	0	14	0	0	0	0
Grade 4 & Grade 8 - Town - Low	14	0	13	0	1	0	0
Grade 4 & Grade 8 - Town - Medium	26	0	25	0	1	0	0
Grade 4 & Grade 8 - Town - High	8	0	6	1	0	1	0
Grade 4 & Grade 8 - Rural area - Low	16	0	16	0	0	0	0
Grade 4 & Grade 8 - Rural area - Medium	12	0	12	0	0	0	0
Grade 4 & Grade 8 - Rural area - High	8	0	8	0	0	0	0
Total	**155**	**0**	**146**	**5**	**3**	**1**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Capital - High	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Capital - Low or Medium	4	1	3	0	0	0	0
Grade 4 & Grade 8 - County town - High	4	0	4	0	0	0	0
Grade 4 & Grade 8 - County town - Low or Medium	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Town - Low	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Town - Medium	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Town - High	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Rural area - Low	6	0	6	0	0	0	0
Grade 4 & Grade 8 - Rural area - Medium	4	1	3	0	0	0	0
Grade 4 & Grade 8 - Rural area - High	4	0	4	0	0	0	0
Total	**54**	**2**	**52**	**0**	**0**	**0**	**0**

Iran, Islamic Rep. of

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), and geographically inaccessible schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private), gender (mixed, girls, boys), and province or grouped provinces (7)
- No implicit stratification
- Sampled one classroom per school
- No overlap between Grade 4 and Grade 8 samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8	24	0	24	0	0	0	0
Grade 4 & Grade 8 - Capital - High	13	0	10	2	1	0	0
Grade 4 & Grade 8 - Capital - Low or Medium	10	0	9	1	0	0	0
Grade 4 & Grade 8 - County town - High	10	0	9	1	0	0	0
Grade 4 & Grade 8 - County town - Low or Medium	14	0	14	0	0	0	0
Grade 4 & Grade 8 - Town - Low	14	0	13	0	1	0	0
Grade 4 & Grade 8 - Town - Medium	26	0	25	0	1	0	0
Grade 4 - Private	22	0	22	0	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 - Public - Boys - Esfahan	10	0	10	0	0	0	0
Grade 4 - Public - Boys - Fars	10	0	10	0	0	0	0
Grade 4 - Public - Boys - Khozestan	10	0	10	0	0	0	0
Grade 4 - Public - Boys - Tehran Province	10	0	10	0	0	0	0
Grade 4 - Public - Boys - Tehran City	10	0	10	0	0	0	0
Grade 4 - Public - Boys - Khorasan Razavi	10	0	10	0	0	0	0
Grade 4 - Public - Boys - Other Provinces	30	0	30	0	0	0	0
Grade 4 - Public - Girls - Esfahan	10	0	10	0	0	0	0
Grade 4 - Public - Girls - Fars	10	0	10	0	0	0	0
Grade 4 - Public - Girls - Khozestan	10	0	10	0	0	0	0
Grade 4 - Public - Girls - Tehran Province	10	0	10	0	0	0	0
Grade 4 - Public - Girls - Tehran City	10	0	10	0	0	0	0
Grade 4 - Public - Girls - Khorasan Razavi	10	0	10	0	0	0	0
Grade 4 - Public - Girls - Other Provinces	30	0	30	0	0	0	0
Grade 4 - Public - Mixed	22	0	22	0	0	0	0
Total	**224**	**0**	**224**	**0**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), and geographically inaccessible schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private), gender (mixed, girls, boys), and province or grouped provinces (7)
- No implicit stratification
- Sampled one classroom per school
- No overlap between Grade 4 and Grade 8 samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8 - Private	14	0	14	0	0	0	0
Grade 8 - Public - Boys - Esfahan	10	0	10	0	0	0	0
Grade 8 - Public - Boys - Fars	10	0	10	0	0	0	0
Grade 8 - Public - Boys - Khozestan	10	0	10	0	0	0	0
Grade 8 - Public - Boys - Tehran Province	10	0	10	0	0	0	0
Grade 8 - Public - Boys - Tehran City	10	0	10	0	0	0	0
Grade 8 - Public - Boys - Khorasan Razavi	10	0	10	0	0	0	0
Grade 8 - Public - Boys - Other Provinces	38	0	38	0	0	0	0
Grade 8 - Public - Girls - Esfahan	10	0	10	0	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8 - Public - Girls - Fars	10	0	10	0	0	0	0
Grade 8 - Public - Girls - Khozestan	10	0	10	0	0	0	0
Grade 8 - Public - Girls - Tehran Province	10	0	10	0	0	0	0
Grade 8 - Public - Girls - Tehran City	10	0	10	0	0	0	0
Grade 8 - Public - Girls - Khorasan Razavi	10	0	10	0	0	0	0
Grade 8 - Public - Girls - Other Provinces	38	0	38	0	0	0	0
Grade 8 - Public - Mixed	10	0	10	0	0	0	0
Total	**220**	**0**	**220**	**0**	**0**	**0**	**0**

Ireland

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 3), special needs schools, and non-aided (private) schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school level socioeconomic status DEIS (urban band 1, urban band 2, rural), language of instruction (Gaelscoil, Gaeltacht, ordinary), and gender (boys, girls, mixed)
- Implicit stratification by location (cities, rural)
- Sampled two classrooms per school
- No overlap between Grade 4 and Grade 8 schools.
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection school sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
All Irish Schools	10	0	10	0	0	0	0
Gaeltacht Schools	8	0	8	0	0	0	0
DEIS Urban Band 1 - Ordinary School	14	0	14	0	0	0	0
DEIS Urban Band 2 - Ordinary School	8	0	8	0	0	0	0
DEIS Rural - Ordinary School	8	0	8	0	0	0	0
Non-DEIS - Ordinary School - Boys	10	0	10	0	0	0	0

School Participation Status (continued)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacements	Participating Schools: 2nd Replacements	Refusal Schools	Excluded Schools
Non-DEIS - Ordinary School - Girls	12	0	12	0	0	0	0
Non-DEIS - Ordinary School - Mixed	81	1	80	0	0	0	0
Total	**151**	**1**	**150**	**0**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of island schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school sector (community/comprehensive, secondary, vocational), socioeconomic status (high, medium, low) and gender (boys, girls, mixed)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 150)
- No overlap between Grade 4 and Grade 8 schools.
- The Field Test and Main Data Collection samples were selected sequentially. The TIMSS Main Data Collection school sample was selected by controlling for the overlap with the TIMSS Field Test sample and the PISA Feasibility study sample using the Chowdhury approach.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Secondary - High SES - Girls	15	0	15	0	0	0	0
Secondary - High SES - Boys	14	0	14	0	0	0	0
Secondary - High SES - Mixed	14	0	14	0	0	0	0
Secondary - Medium SES - Girls	10	0	9	1	0	0	0
Secondary - Medium SES - Boys	9	0	9	0	0	0	0
Secondary - Medium SES - Mixed	8	0	7	0	0	1	0
Secondary - Low SES	10	0	10	0	0	0	0
Vocational - High SES	8	0	7	1	0	0	0
Vocational - Medium SES	17	0	17	0	0	0	0
Vocational - Low SES	19	0	18	0	0	1	0
Community/ Comprehensive - High SES	8	0	7	0	0	1	0
Community/ Comprehensive - Medium SES	10	0	10	0	0	0	0
Community/ Comprehensive - Low SES	10	0	10	0	0	0	0
Total	**152**	**0**	**147**	**2**	**0**	**3**	**0**

Israel

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, ultra Orthodox schools, and schools teaching in English or French
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school sector (3), socioeconomic status (high, medium, low), subgroups within Arab sector (Arab, Druze, Bedouin), and school size (small, large)
- Implicit stratification by gender (male, female, mixed), and region (north, south, all)
- Sampled one classroom per school in schools that are selected to do eTIMSS or Bridge only and two classes in schools that are selected to do both assessments.
- The Field Test and Main Data Collection TIMSS samples were selected sequentially. The TIMSS Main Data Collection sample was selected by controlling for the overlap with the TIMSS Field Test sample and a national study (Mitzav) using the Chowdhury approach.
- The bridge sample was obtained using a combination of strategies. In the large school strata, the bridge sample was selected as a subset of the eTIMSS school sample and classes were randomly assigned to either the eTIMSS or bridge samples. In the small school strata, a distinct sample of schools was selected for the bridge sample. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.
- 91.2 % of students in the bridge sample were in schools selected for the eTIMSS sample

School Participation Status

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacements	Participating Schools: 2nd Replacements	Refusal Schools	Excluded Schools
Hebrew-Secular - High SES	39	0	36	2	0	1	0
Hebrew-Secular - Medium SES	32	0	32	0	0	0	0
Hebrew-Secular - Low SES	11	0	10	0	0	1	0
Hebrew-Religious - High SES	10	0	9	1	0	0	0
Hebrew-Religious - Medium SES - Large	9	0	8	0	1	0	0
Hebrew-Religious - Medium SES - Small	4	0	2	1	0	1	0
Hebrew-Religious - Low SES	8	0	7	0	0	1	0
Arabic-Arabs - Medium SES	10	0	10	0	0	0	0
Arabic-Arabs - Low SES	18	0	18	0	0	0	0
Arabic-Druze	8	0	8	0	0	0	0
Arabic-Bedouin	12	0	12	0	0	0	0
Total	**161**	**0**	**152**	**4**	**1**	**4**	**0**

Bridge Sample School Participation Status

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacements	Participating Schools: 2nd Replacements	Refusal Schools	Excluded Schools
Hebrew-Secular - High SES	14	0	12	1	0	1	0
Hebrew-Secular - Medium SES	12	0	11	1	0	0	0
Hebrew-Secular - Low SES	4	0	3	0	0	1	0
Hebrew-Religious - High SES	4	0	3	1	0	0	0
Hebrew-Religious - Medium SES - Large	4	0	4	0	0	0	0

Bridge Sample School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Hebrew-Religious - Medium SES - Small	2	0	2	0	0	0	0
Hebrew-Religious - Low SES	3	0	3	0	0	0	0
Arabic-Arabs - Medium SES	5	0	5	0	0	0	0
Arabic-Arabs - Low SES	11	0	11	0	0	0	0
Arabic-Druze	4	0	4	0	0	0	0
Arabic-Bedouin	8	0	8	0	0	0	0
Total	**71**	**0**	**66**	**3**	**0**	**2**	**0**

Italy

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 3), special needs schools, Slovenian, Ladin and German language schools
- Within-school exclusions consisted of students with functional disabilities

Sample Design

- Explicit stratification by 'Grade 4 only'/'Grade 4 and Grade 8' schools, school type (private, public), region (center, islands, north east, north west, and south) within Grade 4 and Grade 8 public schools
- Implicit stratification by region (center, islands, north east, north west, and south) within Grade 4 only
- Sampled two classrooms in large schools (measure of size > 112)
- Grade 4 and Grade 8 school samples were selected simultaneously with maximum overlap
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 - Public	16	0	14	2	0	0	0
Grade 4 - Private	10	0	10	0	0	0	0
Grade 4 & Grade 8 - Public - Center	26	0	25	1	0	0	0
Grade 4 & Grade 8 - Public - Islands	20	0	20	0	0	0	0
Grade 4 & Grade 8 - Public - North East	24	0	22	2	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 & Grade 8 - Public - North West	34	0	33	1	0	0	0
Grade 4 & Grade 8 - Public - South	24	0	24	0	0	0	0
Grade 4 & Grade 8 - Private	8	0	5	3	0	0	0
Total	**162**	**0**	**153**	**9**	**0**	**0**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 - Public	6	0	6	0	0	0	0
Grade 4 - Private	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Public - Center	10	0	10	0	0	0	0
Grade 4 & Grade 8 - Public - Islands	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Public - North East	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Public - North West	12	0	12	0	0	0	0
Grade 4 & Grade 8 - Public - South	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Private	4	0	3	1	0	0	0
Total	**60**	**0**	**59**	**1**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 3), special needs schools, Slovenian, Ladin and German language schools
- Within-school exclusions consisted of students with functional disabilities

Sample Design

- Explicit stratification by 'Grade 8 only'/'Grade 4 and Grade 8' schools, school type (private, public), region (center, islands, north east, north west, and south) within Grade 4 and Grade 8 public schools
- Implicit stratification by region (center, islands, north east, north west, and south) within Grade 8 only
- Sampled two classrooms in large schools (measure of size > 135)
- Grade 4 and Grade 8 school samples were selected simultaneously with maximum overlap
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8 - Public	14	0	13	1	0	0	0
Grade 8 - Private	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Public - Center	26	0	25	1	0	0	0
Grade 4 & Grade 8 - Public - Islands	20	0	20	0	0	0	0
Grade 4 & Grade 8 - Public - North East	24	0	22	2	0	0	0
Grade 4 & Grade 8 - Public - North West	34	0	33	1	0	0	0
Grade 4 & Grade 8 - Public - South	24	0	24	0	0	0	0
Grade 4 & Grade 8 - Private	8	0	8	0	0	0	0
Total	**158**	**0**	**153**	**5**	**0**	**0**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8 - Public	4	0	4	0	0	0	0
Grade 8 - Private	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Public - Center	10	0	10	0	0	0	0
Grade 4 & Grade 8 - Public - Islands	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Public - North East	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Public - North West	12	0	12	0	0	0	0
Grade 4 & Grade 8 - Public - South	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Private	4	0	4	0	0	0	0
Total	**58**	**0**	**58**	**0**	**0**	**0**	**0**

Japan

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school location (4)
- No implicit stratification
- Sampled one classroom per school
- No overlap between Grade 4 and Grade 8 samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Very Large City	39	0	33	4	1	1	0
Large City	22	0	17	0	4	1	0
Small City	74	0	61	11	1	1	0
Non-City Area	15	0	15	0	0	0	0
Total	**150**	**0**	**126**	**15**	**6**	**3**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private) and school location (4)
- No implicit stratification
- Sampled one classroom per school
- No overlap between Grade 4 and Grade 8 samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Public - Very large city	32	0	26	2	3	1	0
Public - Large city	20	0	16	1	3	0	0
Public - Small city	69	0	63	2	1	3	0
Public - Non-city area	14	0	13	1	0	0	0
Private or National school	15	0	7	3	1	4	0
Total	**150**	**0**	**125**	**9**	**8**	**8**	**0**

Jordan

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- No school-level exclusions
- Within-school exclusions consisted of students with intellectual disabilities, and students with functional disabilities

Sample Design

- Explicit stratification by school type (6) and achievement level (4)
- Implicit stratification by region (south, north, middle, all)
- Sampled one classroom per school
- The Field Test and Main Data Collection samples were selected simultaneously to avoid overlap.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Discovery - Low	8	1	7	0	0	0	0
Discovery - Medium	8	1	7	0	0	0	0
Discovery - High	8	0	8	0	0	0	0
Discovery - Very High	8	0	8	0	0	0	0
Madrasati - Low	8	0	8	0	0	0	0
Madrasati - Medium	8	0	8	0	0	0	0
Madrasati - High	8	0	8	0	0	0	0
Madrasati - Very High	8	0	8	0	0	0	0
Syria - Low	8	1	7	0	0	0	0
Syria - Medium	8	1	7	0	0	0	0
Syria - High	8	0	8	0	0	0	0
Syria - Very High	8	1	7	0	0	0	0
Public - Low	20	1	19	0	0	0	0
Public - Medium	24	2	22	0	0	0	0
Public - High	20	2	18	0	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Public - Very High	24	1	23	0	0	0	0
UNRWA - Low	8	0	8	0	0	0	0
UNRWA - Medium	8	0	8	0	0	0	0
UNRWA - High	8	0	8	0	0	0	0
UNRWA - Very High	8	0	8	0	0	0	0
Private - Low	8	0	8	0	0	0	0
Private - Medium	8	1	7	0	0	0	0
Private - High	8	0	8	0	0	0	0
Private - Very High	8	1	7	0	0	0	0
Total	**248**	**13**	**235**	**0**	**0**	**0**	**0**

Kazakhstan

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), special needs schools, and Uzbek, Uighur, Tadjik only schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 4 only'/'Grade 4 and Grade 8' schools, region (4), urbanization (urban, rural), and language (Kazakh, Russian)
- No implicit stratification
- Sampled two classrooms per school
- Grade 4 and Grade 8 school samples were selected simultaneously with maximum overlap.
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Region A - Urban - Kazakh	10	0	10	0	0	0	0
Grade 4 & Grade 8 - Region A - Urban - Kazakh and Russian	10	0	10	0	0	0	0
Grade 4 & Grade 8 - Region A - Rural - Kazakh	20	0	20	0	0	0	0
Grade 4 & Grade 8 - Region A - Rural - Kazakh and Russian	8	0	8	0	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 & Grade 8 - Region B - Urban - Kazakh	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Region B - Urban - Kazakh and Russian	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Region B - Rural - Kazakh or Kazakh and Russian	10	0	10	0	0	0	0
Grade 4 & Grade 8 - Region C - Urban - Kazakh	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Region C - Urban - Kazakh and Russian	10	0	10	0	0	0	0
Grade 4 & Grade 8 - Region C - Rural - Kazakh	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Region C - Rural - Kazakh and Russian	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Region D - Urban - Kazakh	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Region D - Urban - Kazakh and Russian	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Region D - Rural - Kazakh or Kazakh and Russian	7	0	7	0	0	0	0
Grade 4 & Grade 8 - Urban - Russian	14	0	14	0	0	0	0
Grade 4 & Grade 8 - Rural - Russian	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Other Languages	7	0	7	0	0	0	1
Total	**168**	**0**	**168**	**0**	**0**	**0**	**1**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), special needs schools, and Uzbek, Uighur, Tadjik only schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 8 only'/'Grade 4 and Grade 8' schools, region (4), urbanization (urban, rural), and language (Kazakh, Russian)
- No implicit stratification
- Sampled two classrooms per school
- Grade 4 and Grade 8 school samples were selected simultaneously with maximum overlap.
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Region A - Urban - Kazakh	10	0	10	0	0	0	0
Grade 4 & Grade 8 - Region A - Urban - Kazakh and Russian	10	0	10	0	0	0	0
Grade 4 & Grade 8 - Region A - Rural - Ka-zakh	20	0	20	0	0	0	0
Grade 4 & Grade 8 - Region A - Rural - Ka-zakh and Russian	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Region B - Urban - Kazakh	8	0	8	0	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 & Grade 8 - Region B - Urban - Kazakh and Russian	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Region B - Rural - Ka-zakh or Kazakh and Russian	10	0	10	0	0	0	0
Grade 4 & Grade 8 - Region C - Urban - Kazakh	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Region C - Urban - Kazakh and Russian	10	0	10	0	0	0	0
Grade 4 & Grade 8 - Region C - Rural - Kazakh	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Region C - Rural - Kazakh and Russian	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Region D - Urban - Kazakh	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Region D - Urban - Kazakh and Russian	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Region D - Rural - Kazakh or Kazakh and Russian	7	0	7	0	0	0	0
Grade 4 & Grade 8 - Urban - Russian	14	0	14	0	0	0	0
Grade 4 & Grade 8 - Rural - Russian	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Other	7	0	7	0	0	0	1
Total	**168**	**0**	**168**	**0**	**0**	**0**	**1**

Korea, Rep. of

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), and remote schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by urbanization (3) and school size (small, large)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 190). In schools sampled for eTIMSS and bridge, one additional classroom was selected for the bridge
- No overlap between Grade 4 and Grade 8 school samples
- The Main Data Collection school sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach
- The bridge sample was obtained using a combination of strategies. In the large school strata, the bridge sample was selected as a subset of the eTIMSS school sample and classes were randomly assigned to either the eTIMSS or bridge samples. In the small school strata, a distinct sample of schools was selected for the bridge sample. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.
- 82.8 % of students in the bridge sample were in schools selected for the eTIMSS sample

School Participation Status

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Big City - Small	8	0	8	0	0	0	0
Big City - Large	52	0	52	0	0	0	0
Medium/Small City - Small	10	0	10	0	0	0	0
Medium/Small City - Large	56	0	56	0	0	0	0
Small Town or Village - Small	10	0	9	0	0	1	0
Small Town or Village - Large	16	0	16	0	0	0	0
Total	**152**	**0**	**151**	**0**	**0**	**1**	**0**

Bridge Sample School Participation Status

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Big City - Small	6	0	6	0	0	0	0
Big City - Large	22	0	22	0	0	0	0
Medium/Small City - Small	6	0	6	0	0	0	0
Medium/Small City - Large	22	0	22	0	0	0	0
Small Town or Village - Small	6	0	6	0	0	0	0
Small Town or Village - Large	6	0	6	0	0	0	0
Total	**68**	**0**	**68**	**0**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), remote schools, and physical education middle schools

- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by urbanization (3), school gender (girls, boys, mixed), and school size (small, large) within small town or village strata
- No implicit stratification
- Sampled one classroom per school. In schools sampled for eTIMSS and bridge, one additional classroom was selected for the bridge
- No overlap between Grade 4 and Grade 8 school samples
- The Main Data Collection school sample for TIMSS was selected by controlling for the overlap with the TIMSS Field Test, PISA, and ICILS samples using the Chowdhury approach
- The bridge sample was obtained using a combination of strategies. In the large school strata, the bridge sample was selected as a subset of the eTIMSS school sample and classes were randomly assigned to either the eTIMSS or bridge samples. In the small school strata, a distinct sample of schools was selected for the bridge sample. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.
- 95.8 % of students in the bridge sample were in schools selected for the eTIMSS sample

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Big City - Boy - Large	10	0	10	0	0	0	0
Big City - Girl - Large	10	0	10	0	0	0	0
Big City - Mixed - Large	46	0	46	0	0	0	0
Medium/Small City - Boy - Large	10	0	10	0	0	0	0
Medium/Small City - Girl - Large	10	0	10	0	0	0	0
Medium/Small City - Mixed - Large	48	0	48	0	0	0	0
Small Town or Village - Boy - Small	2	0	2	0	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Small Town or Village - Boy - Large	5	0	5	0	0	0	0
Small Town or Village - Girl - Small	2	0	2	0	0	0	0
Small Town or Village - Girl - Large	7	0	7	0	0	0	0
Small Town or Village - Mixed - Small	6	0	6	0	0	0	0
Small Town or Village - Mixed - Large	12	0	12	0	0	0	0
Total	**168**	**0**	**168**	**0**	**0**	**0**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Big City - Boy - Large	4	0	4	0	0	0	0
Big City - Girl - Large	4	0	4	0	0	0	0
Big City - Mixed - Large	16	0	16	0	0	0	0
Medium/Small City - Boy - Large	4	0	4	0	0	0	0
Medium/Small City - Girl - Large	4	0	4	0	0	0	0
Medium/Small City - Mixed - Large	18	0	18	0	0	0	0
Small Town or Village - Boy - Small	2	0	2	0	0	0	0
Small Town or Village - Boy - Large	2	0	2	0	0	0	0
Small Town or Village - Girl - Small	2	0	2	0	0	0	0
Small Town or Village - Girl - Large	2	0	2	0	0	0	0
Small Town or Village - Mixed - Small	4	1	3	0	0	0	0
Small Town or Village - Mixed - Large	4	0	4	0	0	0	0
Total	**66**	**1**	**65**	**0**	**0**	**0**	**0**

Kosovo

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of Bosnian schools, and Serbian schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by urbanization (urban, rural) and shifts (one, two or more)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 49)
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Rural - One Shift	16	0	16	0	0	0	0
Rural - Two or more Shifts	58	0	58	0	0	0	0
Urban - One Shift	7	0	7	0	0	0	1
Urban - Two or more Shifts	66	2	64	0	0	0	2
Total	**147**	**2**	**145**	**0**	**0**	**0**	**3**

Kuwait

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools, and minority language schools
- Within-school exclusions consisted of students with intellectual disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 4 only'/'Grade 4 and Grade 8' schools, school type (public, private), region (6), and gender (male, female) within public Grade 4 only schools and language (3) within private Grade 4 and Grade 8 schools
- No implicit stratification
- Sampled one classroom per school
- The Grade 4 and Grade 8 samples were selected with maximum overlap

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 - Private	26	2	24	0	0	0	0
Grade 4 - Public - Ahmedi - Female	12	0	12	0	0	0	0
Grade 4 - Public - Ahmedi - Male	10	0	10	0	0	0	0
Grade 4 - Public - Asima - Female	8	0	8	0	0	0	0
Grade 4 - Public - Asima - Male	8	0	8	0	0	0	0
Grade 4 - Public - Farwaniya - Female	10	0	10	0	0	0	0
Grade 4 - Public - Farwaniya - Male	8	0	8	0	0	0	0
Grade 4 - Public - Hawally - Female	8	0	8	0	0	0	0
Grade 4 - Public - Hawally - Male	8	0	8	0	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 - Public - Jahraa - Female	10	0	10	0	0	0	0
Grade 4 - Public - Jahraa - Male	8	0	8	0	0	0	0
Grade 4 - Public - Mubarak Al-Kabeer - Female	8	0	8	0	0	0	0
Grade 4 - Public - Mubarak Al-Kabeer - Male	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Private - Pakistani and Indian Schools	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Private - English, American, and Bilingual Schools	29	0	25	1	0	3	1
Total	**169**	**2**	**163**	**1**	**0**	**3**	**1**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools, and minority language schools
- No within-school exclusions

Sample Design

- Explicit stratification by 'Grade 8 only'/'Grade 4 and Grade 8' schools, school type (public, private), region (6), and gender (male, female) within public Grade 4 only schools and language (3) within private Grade 4 and Grade 8 schools
- No implicit stratification
- Sampled one classroom per school
- The Grade 4 and Grade 8 samples were selected with maximum overlap

School Participation Status

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Grade 8 - Private	32	2	30	0	0	0	0
Grade 8 - Public - Ahmedi - Female	14	0	14	0	0	0	0
Grade 8 - Public - Ahmedi - Male	12	1	11	0	0	0	0
Grade 8 - Public - Asima - Female	10	0	10	0	0	0	0
Grade 8 - Public - Asima - Male	10	0	10	0	0	0	0
Grade 8 - Public - Farwaniya - Female	12	0	12	0	0	0	0
Grade 8 - Public - Farwaniya - Male	10	0	10	0	0	0	0
Grade 8 - Public - Hawally - Female	10	0	10	0	0	0	0
Grade 8 - Public - Hawally - Male	10	1	9	0	0	0	0
Grade 8 - Public - Jahraa - Female	10	0	10	0	0	0	0
Grade 8 - Public - Jahraa - Male	10	0	10	0	0	0	0
Grade 8 - Public - Mubarak Al-Kabeer - Female	8	0	8	0	0	0	0
Grade 8 - Public - Mubarak Al-Kabeer - Male	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Private - Pakistani and Indian Schools	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Private - English, American, and Bilingual Schools	12	0	11	0	0	1	2
Total	**176**	**4**	**171**	**0**	**0**	**1**	**2**

Latvia

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools, other Language schools, and distance learning schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by urbanization (Riga, city, town and rural area), language (Latvian, Russian), and school type (basic-beginners, secondary) within town and rural area Latvian schools
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 85) and in schools where class grouping was applied
- Class group option was used in bilingual schools

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Riga - Latvian	26	0	25	1	0	0	0
Riga - Russian	24	0	22	1	0	1	0
Other Cities - Latvian	20	0	20	0	0	0	0
Other Cities - Russian	12	0	11	1	0	0	0
Town-Rural - Latvian - Basic/Beginners	30	0	23	5	2	0	0
Town-Rural - Latvian - Secondary	36	0	34	2	0	0	0
Town-Rural - Russian	8	0	7	0	0	1	0
Total	**156**	**0**	**142**	**10**	**2**	**2**	**0**

Lebanon

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 8)
- No within-school exclusions

Sample Design

- Explicit stratification by regions or grouped regions (6), school type (public, private), and school size (small, large)
- No implicit stratification
- Sampled one classroom per school
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- Census of schools in Beirut and Mont Liban large public school strata
- In census strata, schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Bekaa Baalbak Hermel - Public - Small	8	1	7	0	0	0	0
Bekaa Baalbak Hermel - Public - Large	8	0	8	0	0	0	0
Bekaa Baalbak Hermel - Private - Small	8	0	7	1	0	0	0
Bekaa Baalbak Hermel - Private - Large	8	0	7	1	0	0	0
Nord Aakaar - Public - Small	10	0	10	0	0	0	0

School Participation Status (continued)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacements	Participating Schools: 2nd Replacements	Refusal Schools	Excluded Schools
Nord Aakaar - Public - Large	8	0	8	0	0	0	0
Nord Aakaar - Private - Small	8	0	7	1	0	0	0
Nord Aakaar - Private - Large	14	0	12	1	0	1	0
Sud Nabatyeh - Public - Small	8	0	8	0	0	0	0
Sud Nabatyeh - Public - Large	8	0	8	0	0	0	0
Sud Nabatyeh - Private - Small	8	0	8	0	0	0	0
Sud Nabatyeh - Private - Large	12	0	9	1	0	2	0
Beirut - Public - Small	8	1	7	0	0	0	0
Beirut - Public - Large	8	0	8	0	0	0	0
Beirut - Private - Small	8	0	7	0	0	1	0
Beirut - Private - Large	8	0	4	1	0	3	0
Mont Liban - Public - Small	8	0	8	0	0	0	0
Mont Liban - Public - Large	8	0	8	0	0	0	0
Mont Liban - Private - Small	8	0	7	1	0	0	0
Mont Liban - Private - Large	8	0	6	1	0	1	0
Mont Liban Suburb - Public - Small	8	0	8	0	0	0	0
Mont Liban Suburb - Public - Large	8	0	8	0	0	0	0
Mont Liban Suburb - Private - Small	8	0	5	2	0	1	0
Mont Liban Suburb - Private - Large	22	0	14	2	3	3	0
Total	**218**	**2**	**189**	**12**	**3**	**12**	**0**

Lithuania

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), special needs schools, schools with students taught in a language other than Lithuanian, Polish, or Russian, and schools providing remote studying
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 4'/'Grade 4 and Grade 8' schools, and language (5)
- Implicit stratification by urbanization (4), and school type (4)
- Sampled two classrooms in large schools (more than 4 classes)
- Grade 4 and Grade 8 school samples were selected simultaneously with minimum overlap.
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 - Lithuanian only	31	0	31	0	0	0	0
Grade 4 - Other	11	0	11	0	0	0	1
Grade 4 & Grade 8 - Lithuanian only	112	0	112	0	0	0	0
Grade 4 & Grade 8 - Russian	14	0	14	0	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 & Grade 8 - Polish	26	0	26	0	0	0	0
Grade 4 & Grade 8 - Lithuanian and Russian and/or Polish	13	0	13	0	0	0	0
Total	**207**	**0**	**207**	**0**	**0**	**0**	**1**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 - Lithuanian only	12	0	12	0	0	0	0
Grade 4 - Other	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Lithuanian only	46	0	46	0	0	0	0
Grade 4 & Grade 8 - Russian	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Polish	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Lithuanian and Russian and/or Polish	4	0	4	0	0	0	0
Total	**74**	**0**	**74**	**0**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), special needs schools, schools with students taught in a language other than Lithuanian, Polish, or Russian, and schools providing remote studying
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 8'/'Grade 4 and Grade 8' schools, and language (5)
- Implicit stratification by urbanization (4), and school type (4)
- Sampled two classrooms in large schools (more than 4 classes)
- Grade 4 and Grade 8 school samples were selected simultaneously with minimum overlap.
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8 - Lithuanian only	14	0	14	0	0	0	0
Grade 8 - Other	6	0	6	0	0	0	0
Grade 4 & Grade 8 - Lithuanian only	122	0	121	0	0	1	0
Grade 4 & Grade 8 - Russian	14	0	14	0	0	0	0
Grade 4 & Grade 8 - Polish	26	0	26	0	0	0	0
Grade 4 & Grade 8 - Lithuanian and Russian and/or Polish	13	0	13	0	0	0	0
Total	**195**	**0**	**194**	**0**	**0**	**1**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8 - Lithuanian only	8	0	8	0	0	0	0
Grade 8 - Other	2	0	2	0	0	0	0
Grade 4 & Grade 8 - Lithuanian only	50	0	50	0	0	0	0
Grade 4 & Grade 8 - Russian	4	0	4	0	0	0	0

Bridge Sample School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 & Grade 8 - Polish	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Lithuanian and Russian and/or Polish	4	0	4	0	0	0	0
Total	**72**	**0**	**72**	**0**	**0**	**0**	**0**

Malaysia

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 15), special needs schools, schools located at remote area, and schools that do not follow the national curriculum
- Within-school exclusions consisted of students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (6), score level (3) within Ministry of Education daily school strata and urbanization (rural, urban) within all Ministry of Education strata
- No implicit stratification
- Sampled two classrooms in Ministry of Education daily schools
- The Field Test and Main Data Collection samples were selected sequentially. The TIMSS Main Data Collection sample was selected by controlling for the overlap with the TIMSS Field Test and PISA samples using the Chowdhury approach.
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
MOE Daily School - High to mid-High - Urban	9	0	9	0	0	0	0
MOE Daily School - Intermediate or N.A. - Urban	31	0	30	1	0	0	0
MOE Daily School - High to intermediate - Rural	19	0	19	0	0	0	0

School Participation Status (continued)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
MOE Daily School - Mid-Low to Very Low - Rural	31	0	31	0	0	0	0
MOE Daily School - Mid-Low to Very Low - Urban	24	0	24	0	0	0	0
MOE Fully Residential School - Rural	11	0	11	0	0	0	0
MOE Fully Residential School - Urban	10	0	10	0	0	0	0
MOE Religious School - Rural	10	0	9	1	0	0	0
MOE Religious School - Urban	10	0	10	0	0	0	0
MARA Junior Science College	8	0	8	0	0	0	0
Non-MOE Religious School	8	0	8	0	0	0	0
Private School	6	0	6	0	0	0	1
Total	**177**	**0**	**175**	**2**	**0**	**0**	**1**

Bridge Sample School Participation Status

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
MOE Daily School - High to mid-High - Urban	2	0	2	0	0	0	0
MOE Daily School - Intermediate or N.A. - Urban	8	0	8	0	0	0	0
MOE Daily School - High to intermediate - Rural	6	0	6	0	0	0	0
MOE Daily School - Mid-Low to Very Low - Rural	8	0	8	0	0	0	0
MOE Daily School - Mid-Low to Very Low - Urban	6	0	6	0	0	0	0

Bridge Sample School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
MOE Fully Residential School - Rural	2	0	2	0	0	0	0
MOE Fully Residential School - Urban	2	0	2	0	0	0	0
MOE Religious School - Rural	2	0	2	0	0	0	0
MOE Religious School - Urban	2	0	2	0	0	0	0
MARA Junior Science College	2	0	2	0	0	0	0
Non-MOE Religious School	2	0	2	0	0	0	0
Private School	2	0	2	0	0	0	0
Total	**44**	**0**	**44**	**0**	**0**	**0**	**0**

Malta

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools, and language schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (church, independent, state)
- No implicit stratification
- Sampled all classrooms
- Classes were used as variance estimation strata and half classes were used to build jackknife replicates

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Church	25	0	25	0	0	0	0
Independent	11	0	11	0	0	0	0
State	63	1	62	0	0	0	0
Total	**99**	**1**	**98**	**0**	**0**	**0**	**0**

Montenegro

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 2), and language of instruction not Montenegrin
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by region (north, central, south)
- Implicit stratification by urbanization (rural, urban)
- Sampled three classrooms in large schools (measure of size > 80) and two classrooms elsewhere
- All schools at Grade 4 were selected
- Schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
North	66	0	66	0	0	0	0
Central	48	0	48	0	0	0	0
South	26	0	26	0	0	0	0
Total	**140**	**0**	**140**	**0**	**0**	**0**	**0**

Morocco

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 6)
- No within-school exclusions

Sample Design

- Explicit stratification by school type (private, public) and region (12)
- Implicit stratification by urbanization (urban, rural) within public sector
- Sampled two classrooms in public schools from the region of Oued eddahab Lagouira where all schools were taken
- No overlap between Grade 4 and Grade 8 samples
- Schools at the regional level were oversampled. Census in the region of Oued eddahab Lagouira.
- In census strata, schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Private - Casablanca-Settat	8	0	8	0	0	0	0
Private - All other re-gions	16	1	15	0	0	0	0
Public - Tanger-Tetouan-Al Hoceima	20	0	20	0	0	0	0
Public - Oriental	20	0	20	0	0	0	0
Public - Fes-Meknes	20	0	20	0	0	0	0
Public - Rabat-Sale-Kenitra	20	0	20	0	0	0	0
Public - Beni Mellal-Khenifra	20	0	20	0	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Public - Casablanca-Settat	20	0	20	0	0	0	0
Public - Marrakech-Safi	20	0	20	0	0	0	0
Public - Draa-Tafilalet	20	0	20	0	0	0	0
Public - Souss-Massa	20	0	20	0	0	0	0
Public - Guelmim-Oued Noun	20	0	20	0	0	0	0
Public - Laayoune-Sakia El Hamra	20	0	20	0	0	0	0
Public - Eddakhla-Oued Eddahab	21	0	21	0	0	0	0
Total	**265**	**1**	**264**	**0**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 6)
- No within-school exclusions

Sample Design

- Explicit stratification by school type (private, public) and region (12)
- Implicit stratification by urbanization (urban, rural) within public sector
- Sampled two classrooms in public schools from the region of Oued eddahab Lagouira where all schools were taken
- No overlap between Grade 4 and Grade 8 samples
- Schools at the regional level were oversampled. Census in the region of Oued eddahab Lagouira.
- In census strata, schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Private - Casablanca-Settat	8	0	8	0	0	0	0
Private - All other re-gions	12	0	12	0	0	0	0
Public - Tanger Te-touan Al Hoceima	20	0	20	0	0	0	0
Public - Oriental	20	1	19	0	0	0	0
Public - Fes Meknes	20	0	20	0	0	0	0
Public - Rabat Sale Kenitra	20	1	19	0	0	0	0
Public - Beni Mellal Khenifra	20	0	20	0	0	0	0
Public - Casablanca Settat	24	0	24	0	0	0	0
Public - Marrakech Safi	20	0	20	0	0	0	0
Public - Draa Tafilalet	20	0	20	0	0	0	0
Public - Souss Massa	20	0	20	0	0	0	0
Public - Guelmim Oued Noun	20	0	20	0	0	0	0
Public - Laayoune Sa-kia El Hamra	20	0	20	0	0	0	0
Public - Eddakhla Oued Eddahab	9	0	9	0	0	0	0
Total	**253**	**2**	**251**	**0**	**0**	**0**	**0**

Netherlands

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by socioeconomic status (low, medium, high)
- No implicit stratification
- Sampled all classrooms
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
High Mean SES	99	1	40	23	10	25	0
Medium Mean SES	44	1	25	5	3	10	0
Low Mean SES	8	0	6	0	0	2	0
Total	**151**	**2**	**71**	**28**	**13**	**37**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
High Mean SES	32	0	20	6	2	4	0
Medium Mean SES	14	0	6	3	1	4	0
Low Mean SES	4	0	1	1	1	1	0
Total	**50**	**0**	**27**	**10**	**4**	**9**	**0**

New Zealand

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), correspondence schools, Maori-medium Level 1 immersion schools, and mostly students in Level 1-2 immersion units schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- No explicit stratification
- Implicit stratification by school type (state, independent), socioeconomic status (4) and urbanization (major urban centers, smaller urban centers) within state schools
- Sampled two classrooms per school
- The sample at Grade 4 was selected by controlling for the overlap with Grade 8 Field Test and Main Data Collection samples using the Chowdhury approach.

School Participation Status

			Participating Schools				
Explicit Strata	**Total Sampled Schools**	**Ineligible Schools**	**Original Schools**	**1st Replacements**	**2nd Replacements**	**Refusal Schools**	**Excluded Schools**
None	163	2	138	18	4	1	1
Total	**163**	**2**	**138**	**18**	**4**	**1**	**1**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), correspondence schools, Maori-medium Level 1 immersion schools, and mostly students in Level 1-2 immersion units schools

- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (state, independent), and socioeconomic status (4), and urbanization (major urban centers, smaller urban centers) within state schools. One additional stratum created for newly created schools
- Implicit stratification by school gender (coeducational, boys, girls)
- Sampled two classrooms per school
- Class group option was used in schools by ability level (advanced, other).

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Independent school	8	0	6	1	0	1	0
State - Low SES school -Major urban centers	12	0	10	2	0	0	0
State - Low SES school -Smaller centers	8	0	4	2	0	2	0
State - Moderately low SES school - Major urban centers	23	0	18	1	0	4	1
State - Moderately low SES school - Smaller centers	12	0	9	1	1	1	0
State - Moderately high SES school - Major urban centers	40	0	31	4	0	5	0
State - Moderately high SES school - Smaller centers	16	0	11	4	0	1	0
State - High SES school	30	0	25	2	0	3	0
New School	3	1	1	0	1	0	1
Total	**152**	**1**	**115**	**17**	**2**	**17**	**2**

North Macedonia

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools, and Turkish language schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by urbanization (urban, rural, mixed) and language (Macedonian, Albanian, mixed)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 70) and in schools with more than one language of instruction
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Urban - Macedonian	40	0	40	0	0	0	0
Urban - Albanian	8	0	8	0	0	0	0
Urban - Mixed	10	0	10	0	0	0	0
Rural - Macedonian	16	0	14	2	0	0	0
Rural - Albanian	22	0	20	2	0	0	0
Rural - Mixed	10	0	10	0	0	0	0
Mixed - Macedonian	30	0	30	0	0	0	0
Mixed - Albanian or Mixed	14	0	14	0	0	0	0
Total	**150**	**0**	**146**	**4**	**0**	**0**	**0**

Northern Ireland

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 6), and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by region (5), and deprivation group (9)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 90), and in schools with composite classes.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Belfast - Lower Deprivation	8	0	5	0	1	2	0
Belfast - Highest Deprivation	14	0	8	0	3	3	0
Western - Lower Deprivation	10	0	6	4	0	0	0
Western - Moderate to high Deprivation	10	0	5	2	0	3	0
Western - Highest Deprivation	8	0	6	1	0	1	0
North Eastern - Lowest Deprivation	10	0	8	1	1	0	0
North Eastern - Low to moderate Deprivation	10	0	6	4	0	0	0
North Eastern - Higher Deprivation	14	0	7	5	1	1	0
South Eastern - Lowest Deprivation	12	0	6	3	2	1	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
South Eastern - Low to moderate Deprivation	12	0	10	0	1	1	0
South Eastern - Higher Deprivation	10	0	7	1	0	2	0
Southern - Lower Deprivation	14	0	6	2	3	3	0
Southern - Moderate Deprivation	10	0	8	0	1	1	0
Southern - Higher Deprivation	14	0	7	1	2	4	0
Total	**156**	**0**	**95**	**24**	**15**	**22**	**0**

Norway

Fifth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, immigrant schools, Sami schools, and international schools
- Within-school exclusions consisted of students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 5'/'Grade 5 and Grade 9' schools, city (Oslo, other), and municipality size (small, medium, large)
- Implicit stratification by national numeracy test score (4)
- Sampled two classrooms per school
- Grade 5 and Grade 9 school samples were selected simultaneously with minimum overlap.
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples
- Additional replacement schools were used for eTIMSS in the case that schools did not have adequate technology infrastructure

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 5 - Oslo	22	0	16	3	1	2	0
Grade 5 - Other - Small Municipalities	8	0	7	0	0	1	0
Grade 5 - Other - Me-dium Municipalities	36	0	21	7	4	4	0
Grade 5 - Other - Large Municipalities	68	0	52	11	2	3	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 5 & Grade 9 - Oslo	7	0	6	0	0	1	0
Grade 5 & Grade 9 - Other - Small Munici-palities	8	0	6	1	0	1	0
Grade 5 & Grade 9 - Other - Medium Munic-ipalities	8	0	5	1	0	2	0
Grade 5 & Grade 9 - Other - Large Munici-palities	10	0	6	0	1	3	0
Total	**167**	**0**	**119**	**23**	**8**	**17**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 5 - Oslo	4	0	4	0	0	0	0
Grade 5 - Other - Small Municipalities	4	0	4	0	0	0	0
Grade 5 - Other - Me-dium Municipalities	12	0	12	0	0	0	0
Grade 5 - Other - Large Municipalities	20	0	20	0	0	0	0
Grade 5 & Grade 9 - Oslo	4	1	3	0	0	0	0
Grade 5 & Grade 9 - Other - Small Munici-palities	4	0	4	0	0	0	0
Grade 5 & Grade 9 - Other - Medium Munic-ipalities	4	0	4	0	0	0	0
Grade 5 & Grade 9 - Other - Large Munici-palities	4	0	4	0	0	0	0
Total	**56**	**1**	**55**	**0**	**0**	**0**	**0**

Ninth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, immigrant schools, and international schools
- Within-school exclusions consisted of students with functional disabilities

Sample Design

- Explicit stratification by 'Grade 9'/'Grade 5 and Grade 9' schools, city (Oslo, other), and municipality size (small, medium, large)
- Implicit stratification by national numeracy test score (4)
- Sampled two classrooms per school
- Grade 5 and Grade 9 school samples were selected simultaneously with minimum overlap.
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples
- Additional replacement schools were used for eTIMSS in the case that schools did not have adequate technology infrastructure

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 9 - Oslo	20	0	20	0	0	0	0
Grade 9 - Other - Small municipalities	8	0	5	3	0	0	0
Grade 9 - Other - Me-dium municipalities	36	0	24	7	1	4	0
Grade 9 - Other - Large municipalities	62	0	48	11	1	2	0
Grade 5 & Grade 9 - Oslo	9	0	9	0	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 5 & Grade 9 - Other - Small municipalities	10	0	10	0	0	0	0
Grade 5 & Grade 9 - Other - Medium municipalities	8	0	7	1	0	0	0
Grade 5 & Grade 9 - Other - Large municipalities	13	1	9	1	0	2	0
Total	**166**	**1**	**132**	**23**	**2**	**8**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 9 - Oslo	4	0	4	0	0	0	0
Grade 9 - Other - Small municipalities	4	0	4	0	0	0	0
Grade 9 - Other - Medium municipalities	12	0	11	0	0	1	0
Grade 9 - Other - Large municipalities	18	0	16	1	0	1	0
Grade 5 & Grade 9 - Oslo	4	0	4	0	0	0	0
Grade 5 & Grade 9 - Other - Small municipalities	4	0	4	0	0	0	0
Grade 5 & Grade 9 - Other - Medium municipalities	4	0	4	0	0	0	0
Grade 5 & Grade 9 - Other - Large municipalities	4	0	3	0	0	1	0
Total	**54**	**0**	**50**	**1**	**0**	**3**	**0**

Oman

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 7), special needs schools, and evening schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by governorates (11) and school type (private, international)
- No implicit stratification
- Sampled two classrooms in census stratum (Musandam Governorate) and in large schools (measure of size > 250)
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Grade 4 and Grade 8 Field Test samples using the Chowdhury approach.
- Census of schools in Musandam Governorate
- In census strata, schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Muscat Governorate	20	0	20	0	0	0	0
Ash Sharqiyah North Governorate	19	0	19	0	0	0	0
Ash Sharqiyah South Governorate	20	0	20	0	0	0	0
Ad Dakhliyah Governorate	20	0	20	0	0	0	0
Adh Dhahirah Governorate	20	0	20	0	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Al Batinah North Governorate	26	0	26	0	0	0	0
Al Batinah South Governorate	19	0	19	0	0	0	0
Al Buraimi Governorate	8	0	8	0	0	0	0
Musandam Governorate	8	0	8	0	0	0	0
Dhofar Governorate	20	0	20	0	0	0	0
Al Wusta Governorate	8	0	8	0	0	0	0
Private Schools	20	0	18	1	1	0	0
International Schools	20	0	20	0	0	0	0
Total	**228**	**0**	**226**	**1**	**1**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 7), special needs schools, and evening schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by governorates (11) and school type (private, international)
- Implicit stratification by gender (3)
- Sampled two classrooms in large schools (measure of size > 250)
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for minimum overlap with the Grade 4 and Grade 8 Field Test samples and maximum overlap with the Grade 4 Main Data Collection sample using the Chowdhury approach.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Muscat Governorate	23	0	23	0	0	0	0
Ash Sharqiyah North Governorate	20	0	20	0	0	0	0
Ash Sharqiyah South Governorate	20	0	20	0	0	0	0
Ad Dakhliyah Governorate	20	0	20	0	0	0	0
Adh Dhahirah Governorate	20	0	20	0	0	0	0
Al Batinah North Governorate	22	0	22	0	0	0	1
Al Batinah South Governorate	20	0	20	0	0	0	0
Al Buraimi Governorate	8	0	8	0	0	0	0
Musandam Governorate	8	0	8	0	0	0	0
Dhofar Governorate	19	0	19	0	0	0	1
Al Wusta Governorate	8	0	8	0	0	0	0
Private Schools	20	0	15	5	0	0	0
International Schools	20	0	20	0	0	0	0
Total	**228**	**0**	**223**	**5**	**0**	**0**	**2**

Pakistan

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), and language of instruction other than English, Urdu or Sindhi
- No within-school exclusions

Sample Design

- Explicit stratification by school type (public, private), region (Khyber Pakhtunkhwa, Punjab, Sindh, other small regions) within public schools and region (Punjab, other regions) within private schools
- Implicit stratification by region (5), urbanization (urban, rural) and gender (boys, girls) within public schools, and by regions (6) within private schools
- Sampled two classrooms per school
- Private schools were sampled with equal probability as no measure of size was available

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Public - Smaller Re-gions	20	0	15	1	4	0	0
Public - Khyber Pakh-tunkhwa	24	0	22	0	1	1	0
Public - Punjab	50	0	50	0	0	0	0
Public - Sindh	22	0	16	4	0	2	0
Private - Punjab	26	7	13	3	3	0	0
Private - All Other Re-gions	8	1	5	1	1	0	0
Total	**150**	**8**	**121**	**9**	**9**	**3**	**0**

Philippines

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 12), special needs schools, and schools in community with armed conflict
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private), socioeconomic index (high, medium, low), geographic location (urban, rural), and unknown
- No implicit stratification
- Sampled one classroom per school

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Public - High SES - Urban	30	0	30	0	0	0	0
Public - High SES - Rural	45	0	45	0	0	0	1
Public - Medium SES - Urban	34	0	34	0	0	0	0
Public - Medium SES - Rural	20	0	20	0	0	0	0
Public - Low SES - Urban	14	0	14	0	0	0	0
Public - Low SES - Rural	8	0	8	0	0	0	0
Private - High SES	8	1	7	0	0	0	0
Private - Medium SES	8	0	8	0	0	0	0
Private - Low SES	8	0	8	0	0	0	0
Unknown - No available data	8	2	6	0	0	0	0
Total	**183**	**3**	**180**	**0**	**0**	**0**	**1**

Poland

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, and schools with language of instruction other than Polish
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by urbanization (4)
- No implicit stratification
- Sampled two classrooms whenever possible

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Village	54	1	51	2	0	0	0
Town (Up to 20 Thousand Inhabitants)	26	0	26	0	0	0	0
City (20 to 100 Thousand Inhabitants)	30	0	30	0	0	0	0
City (Above 100 Thousand Inhabitants)	40	0	36	4	0	0	0
Total	**150**	**1**	**143**	**6**	**0**	**0**	**0**

Portugal

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, and non Portuguese instruction language or not following national curriculum
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private), NUTS 2 region within public schools (8), and school size (2) within private schools
- Implicit stratification by NUTS 3 region within public schools (25) and NUTS 2 region within private schools (8)
- Sampled two classrooms in large schools (measure of size > 110). In schools sampled for eTIMSS and bridge, one additional classroom sampled for the bridge
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- The bridge sample was obtained using a combination of strategies. In the large school strata, the bridge sample was selected as a subset of the eTIMSS school sample and classes were randomly assigned to either the eTIMSS or bridge samples. In the small school strata, a distinct sample of schools was selected for the bridge sample. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.
- 87.8 % of students in the bridge sample were in schools selected for the eTIMSS sample

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Public - Alentejo	12	0	10	2	0	0	0
Public - Algarve	10	0	9	1	0	0	0
Public - Centro	30	0	26	3	1	0	0
Public - Lisboa	38	0	33	4	1	0	0
Public - Norte - Porto	24	0	22	2	0	0	0
Public - Norte - Other	26	0	26	0	0	0	0
Public - R. A. Açores	8	0	7	1	0	0	0
Public - R. A. Madeira	8	0	7	1	0	0	0
Private - Small	16	1	10	4	1	0	0
Private - Large	10	0	8	2	0	0	0
Total	**182**	**1**	**158**	**20**	**3**	**0**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Public - Alentejo	6	0	6	0	0	0	0
Public - Algarve	4	0	3	1	0	0	0
Public - Centro	16	0	15	1	0	0	0
Public - Lisboa	20	0	16	3	1	0	0
Public - Norte - Porto	12	0	12	0	0	0	0
Public - Norte - Other	12	0	12	0	0	0	0
Public - R. A. Açores	4	0	4	0	0	0	0
Public - R. A. Madeira	4	0	4	0	0	0	0
Private - Small	8	0	7	1	0	0	0
Private - Large	4	0	4	0	0	0	0
Total	**90**	**0**	**83**	**6**	**1**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 10), and non Portuguese instruction language or not following national curriculum
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private) and NUTS 2 region (8) within public schools
- Implicit stratification by NUTS 3 region within public schools (25) and grouped NUTS 2 region within private schools (5)
- Sampled two classrooms in large schools (measure of size > 190)
- The Main Data Collection Grade 8 sample was selected by controlling for the overlap with the Field Test and Grade 4 Main Data Collection samples using the Chowdhury approach.
- No bridge sample required at Grade 8 as they did not participate in TIMSS 2015

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Public - Alentejo	10	0	9	1	0	0	0
Public - Algarve	8	0	8	0	0	0	0
Public - Centro	24	0	22	2	0	0	0
Public - Lisboa	36	0	35	1	0	0	0
Public - Porto	22	0	21	1	0	0	0
Public - Norte	24	0	23	1	0	0	0
Public - R.A. Açores	8	0	5	1	0	2	0
Public - R.A. Madeira	8	0	8	0	0	0	0
Private	18	0	18	0	0	0	0
Total	**158**	**0**	**149**	**7**	**0**	**2**	**0**

Qatar

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools, and instruction not in English or Arabic
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 4 only'/'Grade 4 and Grade 8', gender (3) within Grade 4 only stratum
- Implicit stratification by gender (3) within Grade 4 and 8 schools, and school type (4)
- Sampled one classroom per school. In schools sampled for eTIMSS and bridge, one additional classroom selected for the bridge
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- The bridge sample was selected as a subset of the eTIMSS school sample and classes were randomly assigned to either the eTIMSS or bridge samples. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.
- 98.5 % of students in the bridge sample were in schools selected for the eTIMSS sample
- Schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 - Boys	42	0	42	0	0	0	0
Grade 4 - Girls	46	0	46	0	0	0	0
Grade 4 - Mixed	58	0	58	0	0	0	0
Grade 4 & Grade 8	96	0	96	0	0	0	0
Total	**242**	**0**	**242**	**0**	**0**	**0**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 - Boys	11	0	11	0	0	0	0
Grade 4 - Girls	11	0	11	0	0	0	0
Grade 4 - Mixed	9	0	9	0	0	0	0
Grade 4 & Grade 8	32	0	32	0	0	0	0
Total	**63**	**0**	**63**	**0**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools, and instruction not in English or Arabic
- Within-school exclusions consisted of students with intellectual disabilities, and students with functional disabilities

Sample Design

- Explicit stratification by ‘Grade 8 only’/’Grade 4 and Grade 8’.
- Implicit stratification by gender (3), and school type (4)
- Sampled two classrooms in large schools (measure of size >100) selected for eTIMSS only and in schools selected for eTIMSS and bridge

- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- The bridge sample was selected as a subset of the eTIMSS school sample and classes were randomly assigned to either the eTIMSS or bridge samples. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.
- 98.6 % of students in the bridge sample were in schools selected for the eTIMSS sample
- Schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8	57	0	57	0	0	0	0
Grade 4 & Grade 8	95	0	95	0	0	0	0
Total	**152**	**0**	**152**	**0**	**0**	**0**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8	33	0	33	0	0	0	0
Grade 4 & Grade 8	30	0	30	0	0	0	0
Total	**63**	**0**	**63**	**0**	**0**	**0**	**0**

Romania

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, and schools with different curriculum
- Within-school exclusions consisted of students with intellectual disabilities, and students with functional disabilities

Sample Design

- Explicit stratification by urbanization (rural, urban) and regions (5)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 100)

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Rural - Bucharest-Ilfov	8	0	8	0	0	0	0
Rural - North	32	0	28	3	1	0	0
Rural - Center	12	0	10	1	1	0	0
Rural - South	36	0	35	1	0	0	0
Rural - West	8	0	8	0	0	0	0
Urban - Bucharest-Ilfov	16	0	16	0	0	0	0
Urban - North	28	0	28	0	0	0	0
Urban - Center	12	0	11	1	0	0	0
Urban - South	34	0	34	0	0	0	0
Urban - West	12	0	11	1	0	0	0
Total	**198**	**0**	**189**	**7**	**2**	**0**	**0**

Russian Federation

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by region (43).
- No implicit stratification
- Sampled one classroom per school
- No overlap control between Grade 4 and Grade 8 samples.
- The Main Data Collection sample was selected by controlling for the overlap with the Moscow benchmarking sample using the Chowdhury approach
- An extra sampling stage (regions) was required prior to sampling schools. 43 regions out of 85 were selected with probability proportional to the region size. 14 bigger regions were selected with certainty. Each certainty region make up an explicit stratum. The other sampled regions make up one other large explicit stratum for variance purposes. In this latter stratum of sampled regions, a sample of schools is selected within each region.
- Within regions, schools were selected with probability proportional to (school) size systematic sampling. Schools were sorted (serpentine) by location (up to 7 levels) before being sorted by school size.
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples
- Special treatment is required for variance calculation due to the first sampling stage (region). Within each explicit stratum made up from a certainty region, schools are paired together as in the standard procedure. In the larger explicit stratum composed of sampled regions, regions are paired for variance calculation purposes.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Sankt-Petersburg*	6	0	6	0	0	0	0
City of Moscow*	12	0	11	0	0	1	0
Moscow region*	10	0	10	0	0	0	0
Nizhni Novgorod re-gion*	4	0	4	0	0	0	0
Samara region*	4	0	4	0	0	0	0
Republic of Tatarstan*	6	0	6	0	0	0	0
Republic of Bashkorto-stan*	6	0	6	0	0	0	0
Krasnodar territory*	8	0	8	0	0	0	0
Rostov region*	6	0	6	0	0	0	0
Chelyabinsk region*	4	0	4	0	0	0	0
Sverdlovsk region*	6	0	6	0	0	0	0
Kemerovo region*	4	0	4	0	0	0	0
Krasnoyarsk territory*	4	0	4	0	0	0	0
Republic of Dagestan*	6	0	6	0	0	0	0
Pskov region	4	0	4	0	0	0	0
Republic of Komi	4	0	4	0	0	0	0
Vologda region	4	0	4	0	0	0	0
Voronezh region	4	0	4	0	0	0	0
Belgorod region	4	0	4	0	0	0	0
Tula region	4	0	4	0	0	0	0
Yaroslavl region	4	0	4	0	0	0	0
Ryazan region	4	0	4	0	0	0	0
Tambov region	4	0	4	0	0	0	0
Kostroma region	4	0	4	0	0	0	0
Penza region	4	0	4	0	0	0	0
Chuvashi Republic	4	0	4	0	0	0	0
Orenburg region	4	0	4	0	0	0	0
Saratov region	4	0	4	0	0	0	0
Perm territory	4	0	4	0	0	0	0
Volgograd region	4	0	4	0	0	0	0
Astrakhan region	4	0	4	0	0	0	0
Kurgan region	4	0	4	0	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Khanty-Mansijsk autonomous district	4	0	4	0	0	0	0
Novosibirsk region	4	0	4	0	0	0	0
Irkutsk region	4	0	4	0	0	0	0
Altai territory	4	0	4	0	0	0	0
Zabaikalsk territory	4	0	4	0	0	0	0
Tomsk region	4	0	4	0	0	0	0
Sakhalin region	4	0	4	0	0	0	0
Khabarovsk territory	4	0	4	0	0	0	0
Primorsky territory	4	0	3	0	0	1	0
Stavropol territory	4	0	4	0	0	0	0
Kabardino-Balkarian Republic	4	0	4	0	0	0	0
Total	**202**	**0**	**200**	**0**	**0**	**2**	**0**

* Certainty regions

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Sankt-Petersburg*	2	0	2	0	0	0	0
City of Moscow*	4	0	4	0	0	0	0
Moscow region*	4	0	4	0	0	0	0
Nizhni Novgorod re-gion*	2	0	2	0	0	0	0
Samara region*	2	0	2	0	0	0	0
Republic of Tatarstan*	2	0	2	0	0	0	0
Republic of Bashkortostan*	2	0	2	0	0	0	0
Krasnodar territory*	4	0	4	0	0	0	0
Rostov region*	2	0	2	0	0	0	0
Chelyabinsk region*	2	0	2	0	0	0	0
Sverdlovsk region*	2	0	2	0	0	0	0
Kemerovo region*	2	0	2	0	0	0	0
Krasnoyarsk territory*	2	0	2	0	0	0	0

Bridge Sample School Participation Status (continued)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacements	Participating Schools: 2nd Replacements	Refusal Schools	Excluded Schools
Republic of Dagestan*	2	0	2	0	0	0	0
Pskov region	2	0	2	0	0	0	0
Republic of Komi	2	0	2	0	0	0	0
Vologda region	2	0	2	0	0	0	0
Voronezh region	2	0	2	0	0	0	0
Belgorod region	2	0	2	0	0	0	0
Tula region	2	0	2	0	0	0	0
Yaroslavl region	2	0	2	0	0	0	0
Ryazan region	2	0	2	0	0	0	0
Tambov region	2	0	2	0	0	0	0
Kostroma region	2	0	2	0	0	0	0
Penza region	2	0	2	0	0	0	0
Chuvashi Republic	2	0	2	0	0	0	0
Orenburg region	2	0	2	0	0	0	0
Saratov region	2	0	2	0	0	0	0
Perm territory	2	0	2	0	0	0	0
Volgograd region	2	0	2	0	0	0	0
Astrakhan region	2	0	2	0	0	0	0
Kurgan region	2	0	2	0	0	0	0
Khanty-Mansijsk autonomous district	2	0	2	0	0	0	0
Novosibirsk region	2	0	2	0	0	0	0
Irkutsk region	2	0	2	0	0	0	0
Altai territory	2	0	2	0	0	0	0
Zabaikalsk territory	2	0	2	0	0	0	0
Tomsk region	2	0	2	0	0	0	0
Sakhalin region	2	0	2	0	0	0	0
Khabarovsk territory	2	0	2	0	0	0	0
Primorsky territory	2	0	2	0	0	0	0
Stavropol territory	2	0	2	0	0	0	0
Kabardino-Balkarian Republic	2	0	2	0	0	0	0
Total	**92**	**0**	**92**	**0**	**0**	**0**	**0**

* Certainty regions

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by region (43).
- No implicit stratification
- Sampled one classroom per school
- No overlap control between Grade 4 and Grade 8 samples.
- An extra sampling stage (regions) was required prior to sampling schools. 43 regions out of 85 were selected with probability proportional to the region size. 14 bigger regions were selected with certainty. Each certainty region make up an explicit stratum. The other sampled regions make up one other large explicit stratum for variance purposes. In this latter stratum of sampled regions, a sample of schools is selected within each region.
- Within regions, schools were selected with probability proportional to (school) size systematic sampling. Schools were sorted (serpentine) by location (up to 7 levels) before being sorted by school size.
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples
- Special treatment is required for variance calculation due to the first sampling stage (region). Within each explicit stratum made up from a certainty region, schools are paired together as in the standard procedure. In the larger explicit stratum composed of sampled regions, regions are paired for variance calculation purposes.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Sankt-Petersburg*	6	0	6	0	0	0	0
City of Moscow*	12	0	11	1	0	0	0
Moscow region*	10	0	10	0	0	0	0
Perm territory*	4	0	4	0	0	0	0
Nizhni Novgorod re-gion*	4	0	4	0	0	0	0
Republic of Tatarstan*	6	0	6	0	0	0	0
Republic of Bashkorto-stan*	6	0	6	0	0	0	0
Krasnodar territory*	8	0	8	0	0	0	0
Rostov region*	6	0	6	0	0	0	0
Chelyabinsk region*	6	0	6	0	0	0	0
Sverdlovsk region*	6	0	6	0	0	0	0
Krasnoyarsk territory*	4	0	4	0	0	0	0
Republic of Dagestan*	6	0	6	0	0	0	0
Pskov region	4	0	4	0	0	0	0
Republic of Komi	4	0	4	0	0	0	0
Vologda region	4	0	4	0	0	0	0
Voronezh region	4	0	4	0	0	0	0
Vladimir region	4	0	4	0	0	0	0
Tver region	4	0	4	0	0	0	0
Yaroslavl region	4	0	4	0	0	0	0
Ryazan region	4	0	4	0	0	0	0
Ivanovo region	4	0	4	0	0	0	0
Republic of Marij El	4	0	4	0	0	0	0
Penza region	4	0	4	0	0	0	0
Chuvashi Republic	4	0	4	0	0	0	0
Orenburg region	4	0	4	0	0	0	0
Saratov region	4	0	4	0	0	0	0
Volgograd region	4	0	4	0	0	0	0
Astrakhan region	4	0	4	0	0	0	0
Kurgan region	4	0	4	0	0	0	0
Khanty-Mansijsk auton-omous district	4	0	4	0	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Novosibirsk region	4	0	4	0	0	0	0
Kemerovo region	4	0	4	0	0	0	0
Irkutsk region	4	0	4	0	0	0	0
Altai territory	4	0	4	0	0	0	0
Zabaikalsk territory	4	0	4	0	0	0	0
Tomsk region	4	0	4	0	0	0	0
Sakhalin region	4	0	4	0	0	0	0
Republic of Sakha	4	0	4	0	0	0	0
Primorsky territory	4	0	4	0	0	0	0
Chechen Republic	4	0	4	0	0	0	0
Kabardino-Balkarian Republic	4	0	4	0	0	0	0
Total	**204**	**0**	**203**	**1**	**0**	**0**	**0**

* Certainty regions

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Sankt-Petersburg*	2	0	2	0	0	0	0
City of Moscow*	4	0	4	0	0	0	0
Moscow region*	4	0	4	0	0	0	0
Perm territory*	2	0	2	0	0	0	0
Nizhni Novgorod re-gion*	2	0	2	0	0	0	0
Samara region*	2	0	2	0	0	0	0
Republic of Tatarstan*	2	0	2	0	0	0	0
Republic of Bashkorto-stan*	2	0	2	0	0	0	0
Krasnodar territory*	4	0	4	0	0	0	0
Rostov region*	2	0	2	0	0	0	0
Chelyabinsk region*	2	0	2	0	0	0	0
Sverdlovsk region*	2	0	2	0	0	0	0
Krasnoyarsk territory*	2	0	2	0	0	0	0

Bridge Sample School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Republic of Dagestan*	2	0	2	0	0	0	0
Pskov region	2	0	2	0	0	0	0
Republic of Komi	2	0	2	0	0	0	0
Vologda region	2	0	2	0	0	0	0
Voronezh region	2	0	2	0	0	0	0
Vladimir region	2	0	2	0	0	0	0
Tver region	2	0	2	0	0	0	0
Yaroslavl region	2	0	2	0	0	0	0
Ryazan region	2	0	2	0	0	0	0
Ivanovo region	2	0	2	0	0	0	0
Republic of Marij El	2	0	2	0	0	0	0
Penza region	2	0	2	0	0	0	0
Chuvashi Republic	2	0	2	0	0	0	0
Orenburg region	2	0	2	0	0	0	0
Saratov region	2	0	2	0	0	0	0
Volgograd region	2	0	2	0	0	0	0
Astrakhan region	2	0	2	0	0	0	0
Kurgan region	2	0	2	0	0	0	0
Khanty-Mansijsk autonomous district	2	0	2	0	0	0	0
Novosibirsk region	2	0	2	0	0	0	0
Kemerovo region	2	0	2	0	0	0	0
Irkutsk region	2	0	2	0	0	0	0
Altai territory	2	0	2	0	0	0	0
Zabaikalsk territory	2	0	2	0	0	0	0
Tomsk region	2	0	2	0	0	0	0
Sakhalin region	2	0	2	0	0	0	0
Republic of Sakha	2	0	2	0	0	0	0
Primorsky territory	2	0	2	0	0	0	0
Chechen Republic	2	0	2	0	0	0	0
Kabardino-Balkarian Republic	2	0	2	0	0	0	0
Total	**92**	**0**	**92**	**0**	**0**	**0**	**0**

* Certainty regions

Saudi Arabia

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 6), special needs schools, schools using different language other than Arabic or English, and schools in Jizan, Najran and part of Asir
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private, international/foreign) and gender (boys, girls)
- No implicit stratification
- Sampled one classroom per school
- No overlap between Grade 4 and Grade 8 samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Public - Girls	84	0	82	0	2	0	0
Public - Boys	80	0	78	1	0	1	0
Private - Girls	14	0	13	1	0	0	0
Private - Boys	16	0	16	0	0	0	0
International/Foreign - Girls	14	1	12	1	0	0	0
International/Foreign - Boys	14	0	14	0	0	0	0
Total	**222**	**1**	**215**	**3**	**2**	**1**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 6), special needs schools, schools using different language other than Arabic or English, and schools in Jizan, Najran and part of Asir
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private, international/foreign) and gender (boys, girls)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 215)
- No overlap between Grade 4 and Grade 8 samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Public - Girls	78	0	78	0	0	0	0
Public - Boys	74	1	73	0	0	0	0
Private - Girls	16	1	15	0	0	0	0
Private - Boys	16	0	16	0	0	0	0
International or Foreign - Girls	14	1	12	0	1	0	0
International or Foreign - Boys	14	0	14	0	0	0	0
Total	**212**	**3**	**208**	**0**	**1**	**0**	**0**

Serbia

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 3), special needs schools, and schools with students taught in language other than Serbian
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by region (Belgrade, Vojvodina, Central Serbia), urbanization (urban, rural), and school hierarchy (main school, branch department) within Central Serbia rural schools
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 100)

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Belgrade - Urban	30	0	30	0	0	0	0
Belgrade - Rural	10	0	10	0	0	0	0
Vojvodina - Urban	28	0	28	0	0	0	0
Vojvodina - Rural	16	0	16	0	0	0	0
Central Serbia - Urban	47	0	45	2	0	0	1
Central Serbia - Rural - Main School	15	0	14	0	1	0	1
Central Serbia - Rural - Branch Department	20	1	16	2	1	0	2
Total	**166**	**1**	**159**	**4**	**2**	**0**	**4**

Singapore

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools, and private schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- No explicit stratification
- No implicit stratification
- Sampled two classrooms per school
- No overlap between Grade 4 and Grade 8 schools.
- Census of all schools. Within schools, two classrooms were sampled with probability proportional to the size of the classroom. Within selected classrooms, 19 students were randomly sampled for eTIMSS. The remaining students were assigned to the bridge sample.
- 100 % of students in the bridge sample were in schools selected for the eTIMSS sample
- Schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
None	187	0	187	0	0	0	0
Total	**187**	**0**	**187**	**0**	**0**	**0**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
None	56	0	56	0	0	0	0
Total	**56**	**0**	**56**	**0**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools, and private schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- No explicit stratification
- No implicit stratification
- Sampled two classrooms per school
- No overlap between Grade 4 and Grade 8 schools.
- Census of all schools. Within schools, two classrooms were sampled with probability proportional to the size of the classroom. Within selected classrooms, 19 students were randomly sampled for eTIMSS. The remaining students were assigned to the bridge sample.
- 100 % of students in the bridge sample were in schools selected for the eTIMSS sample
- Schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
None	153	0	153	0	0	0	0
Total	**153**	**0**	**153**	**0**	**0**	**0**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
None	56	0	55	0	0	1	0
Total	**56**	**0**	**55**	**0**	**0**	**1**	**0**

Slovak Republic

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 3), special needs schools, instruction language other than Slovak or Hungarian, and alternative schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by language (Slovak, Hungarian), national testing score (4), and school size (small, large) within Slovak schools
- No implicit stratification
- Sampled two classrooms per school. In schools sampled for eTIMSS and bridge, one additional classroom sampled for the bridge.
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- The bridge sample was obtained using a combination of strategies. In the large school strata, the bridge sample was selected as a subset of the eTIMSS school sample and classes were randomly assigned to either the eTIMSS or bridge samples. In the small school strata, a distinct sample of schools was selected for the bridge sample. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.
- 34.5 % of students in the bridge sample were in schools selected for the eTIMSS sample

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Slovak - Low Average Math Score - Small	8	0	7	1	0	0	0
Slovak - Low Average Math Score - Large	3	0	2	0	1	0	0
Slovak - Medium Average Math Score - Small	30	0	29	1	0	0	0
Slovak - Medium Average Math Score - Large	10	0	10	0	0	0	0
Slovak - High Average Math Score - Small	42	0	41	1	0	0	1
Slovak - High Average Math Score - Large	43	0	43	0	0	0	0
Slovak - Missing Score - Small	12	0	11	0	0	1	0
Hungarian	10	0	10	0	0	0	0
Total	**158**	**0**	**153**	**3**	**1**	**1**	**1**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Slovak - Low Average Math Score - Small	4	0	4	0	0	0	0
Slovak - Low Average Math Score - Large	2	0	1	0	1	0	0
Slovak - Medium Average Math Score - Small	14	0	14	0	0	0	0
Slovak - Medium Average Math Score - Large	4	0	4	0	0	0	0
Slovak - High Average Math Score - Small	18	0	18	0	0	0	0
Slovak - High Average Math Score - Large	18	0	18	0	0	0	0
Slovak - Missing Score - Small	6	0	6	0	0	0	0
Hungarian	4	0	4	0	0	0	0
Total	**70**	**0**	**69**	**0**	**1**	**0**	**0**

South Africa

Fifth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), and special needs schools
- No within-school exclusions

Sample Design

- Explicit stratification by school type (independent, public) and province (9) within public schools
- Implicit stratification by performance level (1st quintile, 2nd quintile, 3rd quintile, 4th and 5th quintiles, missing) within public schools and province (Gauteng, Western Cape, Eastern Cape, Limpopo, all other provinces) within independent schools
- Sampled one classroom per school
- No overlap between Grade 5 and Grade 9 samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Independent	30	0	26	3	1	0	0
Public - Eastern Cape	30	2	28	0	0	0	0
Public - Free State	30	0	29	1	0	0	0
Public - Kwazulu-Natal	30	0	25	3	1	1	0
Public - Limpopo	30	0	30	0	0	0	0
Public - Mpumalanga	30	0	29	1	0	0	0
Public - North West	30	0	30	0	0	0	0
Public - Northern Cape	30	0	30	0	0	0	0
Public - Gauteng	30	0	30	0	0	0	0
Public - Western Cape	30	0	29	1	0	0	0
Total	**300**	**2**	**286**	**9**	**2**	**1**	**0**

Ninth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 11), and special needs schools
- Within-school exclusions consisted of non-native language speakers

Sample Design

- Explicit stratification by school type (independent, public) and province (9) within public schools. Independent schools were stratified by province (Gauteng, Western Cape, all other provinces).
- Implicit stratification by performance level (1st quintile, 2nd quintile, 3rd quintile, 4th and 5th quintiles, missing within public schools) and province (Eastern Cape, Limpopo) within independent schools from all other provinces stratum
- Sampled one classroom per school
- No overlap between Grade 5 and Grade 9 samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Independent - Western Cape	8	0	7	1	0	0	0
Independent - Gauteng	16	0	14	1	1	0	0
Independent - All Other provinces	10	1	9	0	0	0	0
Public - Eastern Cape	30	2	28	0	0	0	0
Public - Free State	30	0	30	0	0	0	0
Public - Kwazulu-Natal	34	0	34	0	0	0	0
Public - Limpopo	30	0	30	0	0	0	0
Public - Mpumalanga	30	0	30	0	0	0	0
Public - North West	30	0	29	0	0	1	0
Public - Northern Cape	30	0	30	0	0	0	0
Public - Western Cape	142	1	141	0	0	0	0
Public - Gauteng	134	0	134	0	0	0	0
Total	**524**	**4**	**516**	**2**	**1**	**1**	**0**

Spain

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), special needs schools, international schools outside Madrid, and geographically inaccessible or campus schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by region (9) and school type (public, private). Within Madrid, explicit stratification was done by school funding (publicly funded, non publicly funded) within private schools, bilingual status (bilingual, not bilingual) within publicly funded private schools and public schools
- Implicit stratification by region within the other regions strata (12) and school type (public, private) within Ceuta and Melilla strata
- Sampled one classroom per school. In schools sampled for eTIMSS and bridge, one classroom sampled per assessment. in schools selected for the bridge only, two classrooms sampled in large schools (measure of size > 45).
- The Field Test and Main Data Collection TIMSS samples were selected sequentially. The TIMSS Main Data Collection sample was selected by controlling for the overlap with the TIMSS Field Test sample using the Chowdhury approach.
- Oversampled in Asturias, Castile and Leon, Catalonia, La Rioja, Ceuta, Melilla and Madrid in order to get better estimates. A census of schools was taken in the autonomous cities of Ceuta and Melilla.
- The bridge sample was obtained using a combination of strategies. In the large school strata, the bridge sample was selected as a subset of the eTIMSS school sample and classes were randomly assigned to either the eTIMSS or bridge samples. In the small school strata, a distinct sample of schools was selected for the bridge sample. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.

- 53 % of students in the bridge sample were in schools selected for the eTIMSS sample
- In census strata, schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates

School Participation Status

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacements	Participating Schools: 2nd Replacements	Refusal Schools	Excluded Schools
Asturias - Public	30	0	30	0	0	0	0
Asturias - Private	20	0	20	0	0	0	0
Castile and Leon - Public	30	0	28	1	1	0	0
Castile and Leon - Private	20	0	19	1	0	0	0
Catalonia - Public	30	0	30	0	0	0	0
Catalonia - Private	20	0	19	1	0	0	0
La Rioja - Public	30	0	30	0	0	0	0
La Rioja - Private	20	0	20	0	0	0	0
Ceuta	21	0	21	0	0	0	0
Melilla	16	0	16	0	0	0	0
Other larger regions - Public	58	0	55	1	1	1	0
Other larger regions - Private	22	0	22	0	0	0	0
Other smaller regions - Public	12	0	11	1	0	0	0
Other smaller regions - Private	6	0	6	0	0	0	0
Madrid - Public Bilingual	40	0	40	0	0	0	0
Madrid - Public Non Bilingual	40	0	40	0	0	0	0
Madrid - Private Bilingual (Pub. Funded)	39	0	39	0	0	0	0
Madrid - Private Non Bilingual (Pub. Funded)	40	0	40	0	0	0	0
Madrid - Private	8	0	8	0	0	0	0
Total	**502**	**0**	**494**	**5**	**2**	**1**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Asturias - Public	2	0	2	0	0	0	0
Asturias - Private	2	0	2	0	0	0	0
Castile and Leon - Public	2	0	2	0	0	0	0
Castile and Leon - Private	2	0	2	0	0	0	0
Catalonia - Public	6	0	6	0	0	0	0
Catalonia - Private	4	0	3	1	0	0	0
La Rioja - Public	2	0	2	0	0	0	0
La Rioja - Private	2	0	2	0	0	0	0
Ceuta	2	0	2	0	0	0	0
Melilla	2	0	2	0	0	0	0
Other larger regions - Public	20	0	19	0	0	1	0
Other larger regions - Private	8	0	8	0	0	0	0
Other smaller regions - Public	4	0	4	0	0	0	0
Other smaller regions - Private	2	0	2	0	0	0	0
Madrid - Public Bilingual	2	0	2	0	0	0	0
Madrid - Public Non Bilingual	2	0	2	0	0	0	0
Madrid - Private Bilingual (Pub. Funded)	2	0	2	0	0	0	0
Madrid - Private Non Bilingual (Pub. Funded)	2	0	2	0	0	0	0
Madrid - Private	2	0	2	0	0	0	0
Total	**70**	**0**	**68**	**1**	**0**	**1**	**0**

Sweden

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, special schools, and international schools
- Within-school exclusions consisted of students with functional disabilities

Sample Design

- Explicit stratification by average achievement for the grade (low, medium, high, missing) and school type (public, private)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 45)
- The Field Test and Main Data Collection samples were selected sequentially.
- The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample and the Main Data Collection sample at Grade 8 using the Chowdhury approach.
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Achievement not available - Public	97	3	94	0	0	0	0
Low Average Achievement - Public	8	0	7	0	1	0	0
Medium Average Achievement - Public	10	1	9	0	0	0	0
High Average Achievement - Public	16	0	16	0	0	0	0
High Average Achievement - Private	11	1	10	0	0	0	0
Other - Private	8	0	8	0	0	0	0
Total	**150**	**5**	**144**	**0**	**1**	**0**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Achievement not available - Public	32	0	31	1	0	0	0
Low Average Achievement - Public	4	0	4	0	0	0	0
Medium Average Achievement - Public	4	0	4	0	0	0	0
High Average Achievement - Public	4	0	4	0	0	0	0
High Average Achievement - Private	4	0	4	0	0	0	0
Other - Private	4	0	3	1	0	0	0
Total	**52**	**0**	**50**	**2**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, special schools, and international schools
- Within-school exclusions consisted of students with functional disabilities

Sample Design

- Explicit stratification by average achievement for the grade (low, medium, high, missing) and school type (public, private)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 100)
- The Field Test and Main Data Collection samples were selected sequentially.
- The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- Separate samples of schools for eTIMSS and bridge were selected so there was no overlap between the samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Missing - Public	8	0	8	0	0	0	0
Low - Public	20	1	19	0	0	0	0
Medium - Public	48	0	47	0	0	1	0
High - Public	46	1	45	0	0	0	0
High - Private	23	0	22	1	0	0	0
Other - Private	8	0	8	0	0	0	0
Total	**153**	**2**	**149**	**1**	**0**	**1**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Missing - Public	4	0	4	0	0	0	0
Low - Public	8	0	8	0	0	0	0
Medium - Public	16	0	15	0	1	0	0
High - Public	14	1	13	0	0	0	0
High - Private	8	0	8	0	0	0	0
Other - Private	4	1	3	0	0	0	0
Total	**54**	**2**	**51**	**0**	**1**	**0**	**0**

Turkey

Fifth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 10), special needs schools, international schools, and schools abroad
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 5 only'/'Grade 5 and Grade 8', school type (public, private), region (13), and school size (small, large) within Grade 5 and 8 schools
- No implicit stratification
- Sampled one classroom per school. In schools sampled for eTIMSS and bridge, one classroom sampled per assessment
- Grade 5 and Grade 8 school samples were selected simultaneously with maximum overlap
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- No bridge sample required at Grade 5 as they participated in TIMSS 2015 at the 4th Grade

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 5 only	8	0	7	1	0	0	0
Grade 5 & Grade 8 - Private - Large	4	0	4	0	0	0	0
Grade 5 & Grade 8 - Private - Small	6	0	6	0	0	0	0
Grade 5 & Grade 8 - Public - TR1-Istanbul	24	0	24	0	0	0	0
Grade 5 & Grade 8 - Public - TR2-West Marmara - Large	4	0	4	0	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 5 & Grade 8 - Public - TR2-West Marmara - Small	4	0	4	0	0	0	0
Grade 5 & Grade 8 - Public - TR3-Aegean - Large	10	0	10	0	0	0	0
Grade 5 & Grade 8 - Public - TR3-Aegean - Small	4	0	4	0	0	0	0
Grade 5 & Grade 8 - Public - TR4-East Marmara	14	0	14	0	0	0	0
Grade 5 & Grade 8 - Public - TR5-West Anatolia	14	0	14	0	0	0	0
Grade 5 & Grade 8 - Public - TR6-Mediterranean - Large	14	0	14	0	0	0	0
Grade 5 & Grade 8 - Public - TR6-Mediterranean - Small	6	0	6	0	0	0	0
Grade 5 & Grade 8 - Public - TR7-Central Anatolia - Large	4	0	4	0	0	0	0
Grade 5 & Grade 8 - Public - TR7-Central Anatolia - Small	3	0	3	0	0	0	0
Grade 5 & Grade 8 - Public - TR8-West Black Sea - Large	4	0	4	0	0	0	0
Grade 5 & Grade 8 - Public - TR8-West Black Sea - Small	4	0	4	0	0	0	0
Grade 5 & Grade 8 - Public - TR9-East Black Sea - Large	4	0	4	0	0	0	0
Grade 5 & Grade 8 - Public - TR9-East Black Sea - Small	4	0	4	0	0	0	0
Grade 5 & Grade 8 - Public - TRA-Northeast Anatolia - Large	4	0	4	0	0	0	0

School Participation Status (continued)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools: Original Schools	Participating Schools: 1st Replacements	Participating Schools: 2nd Replacements	Refusal Schools	Excluded Schools
Grade 5 & Grade 8 - Public - TRA-Northeast Anatolia - Small	4	0	4	0	0	0	0
Grade 5 & Grade 8 - Public - TRB-Centraleast Anatolia - Large	4	0	4	0	0	0	0
Grade 5 & Grade 8 - Public - TRB-Centraleast Anatolia - Small	4	0	4	0	0	0	0
Grade 5 & Grade 8 - Public - TRC-Southeast Anatolia - Large	16	0	16	0	0	0	0
Grade 5 & Grade 8 - Public - TRC-Southeast Anatolia - Small	6	1	5	0	0	0	0
Grade 5 & Grade 8 - Public - Rural Regions	8	0	8	0	0	0	0
Total	**181**	**1**	**179**	**1**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 10), special needs schools, international schools, and schools abroad
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by ‘Grade 8 only’/’Grade 5 and Grade 8’, school type (public, private), region (13), and school size (small, large) within Grade 5 and 8 schools
- No implicit stratification
- Sampled one classroom per school. In schools sampled for eTIMSS and bridge, one classroom sampled per assessment

- Grade 5 and Grade 8 school samples were selected simultaneously with maximum overlap
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- The bridge sample was obtained using a combination of strategies. In the large school strata, the bridge sample was selected as a subset of the eTIMSS sample and classes were randomly assigned to either the eTIMSS or bridge samples. In the small school strata, a distinct sample of schools was selected for the bridge sample. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.
- 69.7 % of students in the bridge sample were in schools selected for the eTIMSS sample

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8	8	0	8	0	0	0	0
Gr5 and Gr8 - Private - Large	4	0	4	0	0	0	0
Gr5 and Gr8 - Private - Small	6	0	6	0	0	0	0
Gr5 and Gr8 - Public - TR1-İstanbul - Large	24	0	24	0	0	0	0
Gr5 and Gr8 - Public - TR2-West Marmara - Large	4	0	4	0	0	0	0
Gr5 and Gr8 - Public - TR2-West Marmara - Small	4	0	4	0	0	0	0
Gr5 and Gr8 - Public - TR3-Aegean - Large	10	0	10	0	0	0	0
Gr5 and Gr8 - Public - TR3-Aegean - Small	4	0	4	0	0	0	0
Gr5 and Gr8 - Public - TR4-East Marmara	14	0	14	0	0	0	0
Gr5 and Gr8 - Public - TR5-West Anatolia	14	0	14	0	0	0	0
Gr5 and Gr8 - Public - TR6-Mediterranean - Large	14	0	14	0	0	0	0

School Participation Status (continued)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Gr5 and Gr8 - Public - TR6-Mediterranean - Small	6	0	6	0	0	0	0
Gr5 and Gr8 - Public - TR7-Central Anatolia - Large	4	0	4	0	0	0	0
Gr5 and Gr8 - Public - TR7-Central Anatolia - Small	3	0	3	0	0	0	0
Gr5 and Gr8 - Public - TR8-West Black Sea - Large	4	0	4	0	0	0	0
Gr5 and Gr8 - Public - TR8-West Black Sea - Small	4	0	4	0	0	0	0
Gr5 and Gr8 - Public - TR9-East Black Sea - Large	4	0	4	0	0	0	0
Gr5 and Gr8 - Public - TR9-East Black Sea - Small	4	0	4	0	0	0	0
Gr5 and Gr8 - Public - TRA-Northeast Anatolia - Large	4	0	4	0	0	0	0
Gr5 and Gr8 - Public - TRA-Northeast Anatolia - Small	4	0	3	1	0	0	0
Gr5 and Gr8 - Public - TRB-Centraleast Anato-lia - Large	4	0	4	0	0	0	0
Gr5 and Gr8 - Public - TRB-Centraleast Anato-lia - Small	4	0	4	0	0	0	0
Gr5 and Gr8 - Public - TRC-Southeast Anatolia - Large	16	0	16	0	0	0	0
Gr5 and Gr8 - Public - TRC-Southeast Anatolia - Small	6	0	6	0	0	0	0
Gr5 and Gr8 - Public - Rural Regions	8	0	8	0	0	0	0
Total	**181**	**0**	**180**	**1**	**0**	**0**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8	4	0	4	0	0	0	0
Gr5 and Gr8 - Private - Large	2	0	2	0	0	0	0
Gr5 and Gr8 - Private - Small	2	0	2	0	0	0	0
Gr5 and Gr8 - Public - TR1-İstanbul - Large	8	0	8	0	0	0	0
Gr5 and Gr8 - Public - TR2-West Marmara - Large	2	0	2	0	0	0	0
Gr5 and Gr8 - Public - TR2-West Marmara - Small	2	0	2	0	0	0	0
Gr5 and Gr8 - Public - TR3-Aegean - Large	4	0	4	0	0	0	0
Gr5 and Gr8 - Public - TR3-Aegean - Small	2	0	2	0	0	0	0
Gr5 and Gr8 - Public - TR4-East Marmara	4	0	4	0	0	0	0
Gr5 and Gr8 - Public - TR5-West Anatolia	4	0	4	0	0	0	0
Gr5 and Gr8 - Public - TR6-Mediterranean - Large	4	0	4	0	0	0	0
Gr5 and Gr8 - Public - TR6-Mediterranean - Small	2	0	2	0	0	0	0
Gr5 and Gr8 - Public - TR7-Central Anatolia - Large	2	0	2	0	0	0	0
Gr5 and Gr8 - Public - TR7-Central Anatolia - Small	2	0	2	0	0	0	0
Gr5 and Gr8 - Public - TR8-West Black Sea - Large	2	0	2	0	0	0	0
Gr5 and Gr8 - Public - TR8-West Black Sea - Small	2	0	2	0	0	0	0

Bridge Sample School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Gr5 and Gr8 - Public - TR9-East Black Sea - Large	2	0	2	0	0	0	0
Gr5 and Gr8 - Public - TR9-East Black Sea - Small	2	0	2	0	0	0	0
Gr5 and Gr8 - Public - TRA-Northeast Anatolia - Large	2	0	2	0	0	0	0
Gr5 and Gr8 - Public - TRA-Northeast Anatolia - Small	2	0	2	0	0	0	0
Gr5 and Gr8 - Public - TRB-Centraleast Anato-lia - Large	2	0	2	0	0	0	0
Gr5 and Gr8 - Public - TRB-Centraleast Anato-lia - Small	2	0	2	0	0	0	0
Gr5 and Gr8 - Public - TRC-Southeast Anatolia - Large	6	0	6	0	0	0	0
Gr5 and Gr8 - Public - TRC-Southeast Anatolia - Small	2	0	2	0	0	0	0
Gr5 and Gr8 - Public - Rural Regions	4	0	4	0	0	0	0
Total	**72**	**0**	**72**	**0**	**0**	**0**	**0**

United Arab Emirates

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of instruction language other than English or Arabic, and very small schools (measure of size < 10 in Abu Dhabi and < 5 in other emirates but Dubai)
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 4'/'Grade 4 and Grade 8' schools, emirate (Dubai, Abu Dhabi, all other emirates), school type (public, private), as well as by the main curriculum taught (UK/US/CAD/AUS/International, other) within private schools in all emirates with the exception of Dubai
- Implicit stratification by school size (small, large), as well as region (Abu Dhabi, Al Ain, Al Dhafra) within large private schools of Abu Dhabi, language of test (Arabic, English, French) within Dubai, emirate (Sharjah, other emirates) within public schools, and curriculum (Ministry of Education, UK/US/CAD, other) within private schools in the rest of the emirates
- Sampled two classrooms per school. In schools sampled for eTIMSS and bridge, one additional classroom was sampled for the bridge
- Grade 4 and Grade 8 school samples were selected simultaneously with maximum overlap
- The Field Test and Main Data Collection samples were selected sequentially. In areas other than Dubai, the Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- The United Arab Emirates was divided into three areas: Abu Dhabi, Dubai, and the rest of the emirates. All schools were sampled in Dubai. All public schools as well as all private schools with UK/US/CAD/AUS/International/SABIS curriculum were sampled in the other emirates.
- The bridge sample was selected as a subset of the eTIMSS school sample and classes were randomly assigned to either the eTIMSS or bridge samples. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.

- 98.9 % of students in the bridge sample were in schools selected for the eTIMSS sample
- In census strata, schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Dubai - Grade 4 - Public	24	3	21	0	0	0	0
Dubai - Grade 4 - Private	47	0	47	0	0	0	0
Dubai - Grade 4 & Grade 8 - Public	4	1	3	0	0	0	0
Dubai - Grade 4 & Grade 8 - Private	130	2	128	0	0	0	0
Abu Dhabi - Grade 4 - Public - ADEC schools	67	1	66	0	0	0	0
Abu Dhabi - Grade 4 - Private - UK/US/CAD/AUS/Int	23	0	23	0	0	0	0
Abu Dhabi - Grade 4 - Private - Other curriculum	6	1	5	0	0	0	0
Abu Dhabi - Grade 4 & Grade 8 - Public - ADEC schools	22	0	22	0	0	0	0
Abu Dhabi - Grade 4 & Grade 8 - Private - Ministry of Education	20	0	20	0	0	0	0
Abu Dhabi - Grade 4 & Grade 8 - Private - UK/US/CAD/AUS/Int	87	0	87	0	0	0	0
Abu Dhabi - Grade 4 & Grade 8 - Private - Others	24	0	24	0	0	0	0
Other - Grade 4 - Public - Ministry of Education	80	1	79	0	0	0	0
Other - Grade 4 - Private - UK/US/AUS/International/SABIS	13	0	13	0	0	0	0
Other - Grade 4 - Private - Other Curriculum	8	0	8	0	0	0	0

School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Other - Grade 4 & Grade 8 - Public - Ministry of Education	22	0	22	0	0	0	0
Other - Grade 4 & Grade 8 - Private - Ministry of Education	20	0	20	0	0	0	0
Other - Grade 4 & Grade 8 - Private - UK/ US/AUS/International/ SABIS	60	0	60	0	0	0	0
Other - Grade 4 & Grade 8 - Private - Other Curriculum	40	0	40	0	0	0	0
Total	**697**	**9**	**688**	**0**	**0**	**0**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Dubai - Grade 4 - Public	4	0	4	0	0	0	0
Dubai - Grade 4 - Private	5	0	5	0	0	0	0
Dubai - Grade 4 & Grade 8 - Public	2	1	1	0	0	0	0
Dubai - Grade 4 & Grade 8 - Private	14	0	13	0	0	1	0
Abu Dhabi - Grade 4 - Public - ADEC schools	8	0	8	0	0	0	0
Abu Dhabi - Grade 4 - Private - UK/US/CAD/ AUS/Int	4	0	4	0	0	0	0
Abu Dhabi - Grade 4 - Private - Other curriculum	1	0	1	0	0	0	0
Abu Dhabi - Grade 4 & Grade 8 - Public - ADEC schools	4	0	4	0	0	0	0

Bridge Sample School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Abu Dhabi - Grade 4 & Grade 8 - Private - Ministry of Education	6	0	6	0	0	0	0
Abu Dhabi - Grade 4 & Grade 8 - Private - UK/US/CAD/AUS/Int	8	0	8	0	0	0	0
Abu Dhabi - Grade 4 & Grade 8 - Private - Others	4	0	4	0	0	0	0
Other - Grade 4 - Public - Ministry of Education	10	0	10	0	0	0	0
Other - Grade 4 - Private - UK/US/AUS/International/SABIS	4	0	4	0	0	0	0
Other - Grade 4 - Private - Other Curriculum	4	0	4	0	0	0	0
Other - Grade 4 & Grade 8 - Public - Ministry of Education	4	0	4	0	0	0	0
Other - Grade 4 & Grade 8 - Private - Ministry of Education	3	0	3	0	0	0	0
Other - Grade 4 & Grade 8 - Private - UK/US/AUS/International/SABIS	8	0	8	0	0	0	0
Other - Grade 4 & Grade 8 - Private - Other Curriculum	8	0	7	0	0	1	0
Total	**101**	**1**	**98**	**0**	**0**	**2**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of instruction language other than English or Arabic, special needs schools, and very small schools (measure of size < 10 in Abu Dhabi and < 5 in other emirates but Dubai)
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 8'/'Grade 4 and Grade 8' schools, emirate (Dubai, Abu Dhabi, all other emirates), school type (public, private), as well as by the main curriculum taught (UK/US/CAD/AUS/International, other) within private schools in all emirates with the exception of Dubai
- Implicit stratification by school size (small, large), as well as region (Abu Dhabi, Al Ain, Al Dhafra) within large private schools of Abu Dhabi, language of test (Arabic, English, French) within Dubai, emirate (Sharjah, other emirates) within public schools, and curriculum (Ministry of Education, UK/US/CAD, other) within private schools in the rest of the emirates
- Sampled two classrooms per school. In schools sampled for eTIMSS and bridge, one additional classroom was sampled for the bridge
- Grade 4 and Grade 8 school samples were selected simultaneously with maximum overlap
- The Field Test and Main Data Collection samples were selected sequentially. In areas other than Dubai, the Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- The United Arab Emirates was divided into three areas: Abu Dhabi, Dubai, and the rest of the emirates. All schools were sampled in Dubai. All public schools as well as all private schools with UK/US/CAD/AUS/International/SABIS curriculum were sampled in the other emirates.
- The bridge sample was selected as a subset of the eTIMSS school sample and classes were randomly assigned to either the eTIMSS or bridge samples. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.
- 98 % of students in the bridge sample were in schools selected for the eTIMSS sample

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Dubai - Grade 8 - Public	23	3	20	0	0	0	0
Dubai - Grade 8 - Private	7	0	7	0	0	0	0
Dubai - Grade 4 & Grade 8 - Public	4	1	3	0	0	0	0
Dubai - Grade 4 & Grade 8 - Private	137	4	133	0	0	0	0
Abu Dhabi - Grade 8 - Public	72	0	72	0	0	0	0
Abu Dhabi - Grade 8 - Private	4	0	4	0	0	0	0
Abu Dhabi - Grade 4 & Grade 8 - Public - ADEC schools	23	0	23	0	0	0	0
Abu Dhabi - Grade 4 & Grade 8 - Private - Ministry of Education	20	0	20	0	0	0	0
Abu Dhabi - Grade 4 & Grade 8 - Private - UK/US/CAD/AUS/Int	87	0	87	0	0	0	0
Abu Dhabi - Grade 4 & Grade 8 - Private - Others	24	0	24	0	0	0	0
Other - Grade 8 - Public	83	0	83	0	0	0	0
Other - Grade 8 - Private	7	0	7	0	0	0	0
Other - Grade 4 & Grade 8 - Public - Ministry of Education	21	0	21	0	0	0	0
Other - Grade 4 & Grade 8 - Private - Ministry of Education	20	0	20	0	0	0	0
Other - Grade 4 & Grade 8 - Private - UK/US/AUS/Int./SABIS	59	0	59	0	0	0	0
Other - Grade 4 & Grade 8 - Private - Others	40	0	40	0	0	0	0
Total	**631**	**8**	**623**	**0**	**0**	**0**	**0**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Dubai - Grade 8 - Public	4	3	1	0	0	0	0
Dubai - Grade 8 - Private	2	0	2	0	0	0	0
Dubai - Grade 4 & Grade 8 - Public	2	1	1	0	0	0	0
Dubai - Grade 4 & Grade 8 - Private	14	0	13	0	0	1	0
Abu Dhabi - Grade 8 - Public	10	0	10	0	0	0	0
Abu Dhabi - Grade 8 - Private	4	0	4	0	0	0	0
Abu Dhabi - Grade 4 & Grade 8 - Public - ADEC schools	2	0	2	0	0	0	0
Abu Dhabi - Grade 4 & Grade 8 - Private - Ministry of Education	6	0	6	0	0	0	0
Abu Dhabi - Grade 4 & Grade 8 - Private - UK/US/CAD/AUS/Int	8	0	8	0	0	0	0
Abu Dhabi - Grade 4 & Grade 8 - Private - Others	4	0	4	0	0	0	0
Other - Grade 8 - Public	11	0	11	0	0	0	0
Other - Grade 8 - Private	3	0	3	0	0	0	0
Other - Grade 4 & Grade 8 - Public - Ministry of Education	3	0	3	0	0	0	0
Other - Grade 4 & Grade 8 - Private - Ministry of Education	4	0	4	0	0	0	0
Other - Grade 4 & Grade 8 - Private - UK/US/AUS/Int./SABIS	8	0	8	0	0	0	0
Other - Grade 4 & Grade 8 - Private - Others	8	0	8	0	0	0	0
Total	**93**	**4**	**88**	**0**	**0**	**1**	**0**

United States

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- No school-level exclusions
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by poverty level (high, low), school type (public, private), census region of public school (Northeast, Midwest, South, West), and type of private school (Roman Catholic, other private)
- Implicit stratification by urbanization (city, suburb, town, rural), ethnicity status (above 15% non-White students in a school, below 15% non-White students in a school), and state (52)
- Sampled two classrooms per school. In schools sampled for eTIMSS and bridge, one classroom sampled per assessment
- No overlap between Grade 4 and Grade 8 samples.
- The TIMSS Main Data Collection sample was selected by controlling for the overlap with the national NAEP sample using the Chowdhury approach
- The bridge sample was obtained using a combination of strategies. In the large school strata, the bridge sample was selected as a subset of the eTIMSS school sample and classes were randomly assigned to either the eTIMSS or bridge samples. In the small school strata, a distinct sample of schools was selected for the bridge sample. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.
- 90.7% of students in the bridge sample were in schools selected for the eTIMSS sample

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
High poverty - Public - Census Region 1	20	0	13	0	1	6	0
High poverty - Public - Census Region 2	26	0	18	3	0	5	0
High poverty - Public - Census Region 3	65	0	59	5	0	1	0
High poverty - Public - Census Region 4	42	0	29	7	2	4	0
Low poverty - Private - Non Catholic	13	2	5	1	0	5	0
Low poverty - Private - Catholic	8	0	4	1	0	3	0
Low poverty - Public - Census Region 1	29	0	21	5	0	3	0
Low poverty - Public - Census Region 2	36	0	27	4	3	2	0
Low poverty - Public - Census Region 3	56	0	49	3	0	4	1
Low poverty - Public - Census Region 4	32	0	24	2	1	5	1
Total	**327**	**2**	**249**	**31**	**7**	**38**	**2**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
High poverty - Public - Census Region 1	6	0	5	0	0	1	0
High poverty - Public - Census Region 2	6	0	6	0	0	0	0
High poverty - Public - Census Region 3	16	0	13	3	0	0	0
High poverty - Public - Census Region 4	10	0	5	4	0	1	0
Low poverty - Private - Non Catholic	4	1	1	1	1	0	0
Low poverty - Private - Catholic	4	0	2	1	0	1	0

Bridge Sample School Participation Status (continued)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Low poverty - Public - Census Region 1	8	0	5	2	0	1	0
Low poverty - Public - Census Region 2	10	0	8	2	0	0	0
Low poverty - Public - Census Region 3	14	0	12	0	0	2	0
Low poverty - Public - Census Region 4	8	0	8	0	0	0	0
Total	**86**	**1**	**65**	**13**	**1**	**6**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- No school-level exclusions
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by poverty level (high, low), school type (public, private), census region of public school (Northeast, Midwest, South, West), and type of private school (Roman Catholic, other private)
- Implicit stratification by urbanization (city, suburb, town, rural), ethnicity status (above 15% non-White students in a school, below 15% non-White students in a school), and state (52)
- Sampled two classrooms per school. In schools sampled for eTIMSS and bridge, one classroom sampled per assessment
- No overlap between Grade 4 and Grade 8 samples.
- The TIMSS Main Data Collection sample was selected by controlling for the overlap with the national NAEP sample using the Chowdhury approach
- The bridge sample was obtained using a combination of strategies. In the large school strata, the bridge sample was selected as a subset of the eTIMSS school sample and classes were randomly assigned to either the eTIMSS or bridge samples. In the small school strata, a

distinct sample of schools was selected for the bridge sample. During data collection, schools with only one class selected for both the eTIMSS and bridge samples were randomly assigned to administer either the eTIMSS or bridge assessment, and school weights were adjusted accordingly during the weighting process.

- 91.6% of students in the bridge sample were in schools selected for the eTIMSS sample

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8 - High - Public - Census region 1	17	1	6	2	0	8	0
Grade 8 - High - Public - Census region 2	22	0	17	2	0	3	0
Grade 8 - High - Public - Census region 3	59	1	50	4	2	2	0
Grade 8 - High - Public - Census region 4	39	0	27	2	1	9	0
Grade 8 - Low - Private - Non Catholic	13	1	5	1	2	4	0
Grade 8 - Low - Private - Catholic	8	0	5	1	0	2	0
Grade 8 - Low - Public - Census region 1	29	0	17	7	2	3	1
Grade 8 - Low - Public - Census region 2	42	0	25	5	3	9	0
Grade 8 - Low - Public - Census region 3	61	0	53	3	1	4	0
Grade 8 - Low - Public - Census region 4	34	0	26	3	1	4	0
Total	**324**	**3**	**231**	**30**	**12**	**48**	**1**

Bridge Sample School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8 - High - Public - Census region 1	6	0	3	0	0	3	0
Grade 8 - High - Public - Census region 2	6	0	4	2	0	0	0
Grade 8 - High - Public - Census region 3	14	0	13	1	0	0	0

Bridge Sample School Participation Status (continued)

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8 - High - Public - Census region 4	9	0	6	0	0	3	0
Grade 8 - Low - Private - Non Catholic	4	1	1	0	0	2	0
Grade 8 - Low - Private - Catholic	4	0	3	1	0	0	0
Grade 8 - Low - Public - Census region 1	8	0	5	1	0	2	0
Grade 8 - Low - Public - Census region 2	10	0	5	1	1	3	0
Grade 8 - Low - Public - Census region 3	14	0	13	0	0	1	0
Grade 8 - Low - Public - Census region 4	8	0	5	0	0	3	0
Total	**83**	**1**	**58**	**6**	**1**	**17**	**0**

Ontario, Canada

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 6), special needs schools, and remote and hard to access school
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by language (English, French), school type (private, Catholic, public), and by school size (small, large) within Catholic and public schools
- Implicit stratification by regional office (Thunder Bay, Sudbury-North Bay, London, Barrie, Ottawa, Toronto and Area)
- Sampled two classrooms in large schools (measure of size > 40)
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
English - Public - Small	24	1	23	0	0	0	0
English - Public - Large	72	0	70	1	0	1	0
English - Catholic - Small	18	0	18	0	0	0	0
English - Catholic - Large	19	0	19	0	0	0	0
Private	8	0	0	1	1	6	0
French - Catholic & Public - Small	18	0	18	0	0	0	0
French - Catholic & Public - Large	12	0	12	0	0	0	0
Total	**171**	**1**	**160**	**2**	**1**	**7**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 6), and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by language (French, English), school type (public, Catholic, private), and by school size (small, large) within French Catholic and public schools
- Implicit stratification by regional office (Thunder Bay, Sudbury-North Bay, London, Barrie, Ottawa, Toronto and Area)
- Sampled two classrooms in large schools (measure of size > 100)
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
English - Public	94	2	86	0	1	5	0
English - Catholic	40	0	40	0	0	0	0
French - Large - Catholic & Public	16	0	15	0	0	1	0
French - Small - Catholic & Public	14	0	14	0	0	0	0
Private	8	0	2	0	0	6	0
Total	**172**	**2**	**157**	**0**	**1**	**12**	**0**

Quebec, Canada

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 10), special school, First Nation schools/federal schools, international schools, and school boards with special status
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by language (French, English), school type (public, private), and school size (small, large)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 75)
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- Class group option was used within schools with regular and enriched programs

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
English - Public - Small	12	0	8	1	0	3	0
English - Public - Large	14	0	13	1	0	0	0
English - Private	8	0	6	0	0	2	0
French - Public - Small	44	0	35	1	0	8	0
French - Public - Large	84	0	70	4	0	10	0
French - Private - Small	4	0	3	1	0	0	0
French - Private - Large	6	0	5	0	0	1	0
Total	**172**	**0**	**140**	**8**	**0**	**24**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 10), special school, First Nation schools/federal schools, international schools, and school boards with special status
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by language (French, English), school type (public, private), and school size (small, large) within French public schools
- Implicit stratification by Mathematics average score (4) and available programs (regular program, with enriched program)
- Sampled two classrooms in schools with regular and enriched programs
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- Class group option was used within schools with regular and enriched programs

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
English - Public	23	0	17	0	0	6	1
English - Private	8	2	4	0	0	2	0
French - Private	28	1	19	1	0	7	0
French - Large - Public	62	0	46	1	0	15	0
French - Small - Public	44	1	33	3	0	7	0
Total	**165**	**4**	**119**	**5**	**0**	**37**	**1**

Moscow City, Russian Fed.

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 4'/'Grade 4 and Grade 8' schools and school type (state, private) within 'Grade 4 and Grade 8' schools
- Implicit stratification by school size (up to 1,000 students, from 1,000 to 3,000 students, over 3,000 students) within 'Grade 4 and Grade 8' state schools
- Sampled two classrooms in large schools (measure of size > 300)
- Grade 4 and Grade 8 samples were selected simultaneously with maximum overlap

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4	4	0	4	0	0	0	0
Grade 4 & Grade 8 - State	140	1	138	0	1	0	0
Grade 4 & Grade 8 - Private	8	0	6	1	0	1	0
Total	**152**	**1**	**148**	**1**	**1**	**1**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 8'/'Grade 4 and Grade 8' schools and school type (state, private) within 'Grade 4 and Grade 8' schools
- Implicit stratification by school size (up to 1,000 students, from 1,000 to 3,000 students, over 3,000 students) within 'Grade 4 and Grade 8' state schools
- Sampled two classrooms in large schools (measure of size > 260)
- Grade 4 and Grade 8 samples were selected simultaneously with maximum overlap

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8	4	0	3	0	1	0	0
Grade 4 & Grade 8 - State	140	1	138	0	1	0	0
Grade 4 & Grade 8 - Private	8	0	6	1	0	1	0
Total	**152**	**1**	**147**	**1**	**2**	**1**	**0**

Gauteng, RSA

Ninth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 11), and special needs schools
- Within-school exclusions consisted of non-native language speakers

Sample Design

- Explicit stratification by school type (independent, public)
- Implicit stratification by performance level (1st quintile, 2nd quintile, 3rd quintile, 4th quintile, 5th quintiles, missing) within public schools
- Sampled one classroom per school
- No overlap control between Grade 5 and Grade 9 samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Independent	16	0	14	1	1	0	0
Public	134	0	134	0	0	0	0
Total	**150**	**0**	**148**	**1**	**1**	**0**	**0**

Western Cape, RSA

Ninth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 11), and special needs schools
- Within-school exclusions consisted of non-native language speakers

Sample Design

- Explicit stratification by school type (independent, public)
- Implicit stratification by performance level (1st quintile, 2nd quintile, 3rd quintile, 4th quintile, 5th quintiles, missing) within public schools
- Sampled one classroom per school
- No overlap control between Grade 5 and Grade 9 samples

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Independent	8	0	7	1	0	0	0
Public	142	1	141	0	0	0	0
Total	**150**	**1**	**148**	**1**	**0**	**0**	**0**

Madrid, Spain

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private publicly funded, private) and bilingual status (bilingual, not bilingual) within public schools and private publicly funded schools
- No implicit stratification
- Sampled one classroom per school
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Public Bilingual	40	0	40	0	0	0	0
Public Non Bilingual	40	0	40	0	0	0	0
Private Bilingual (Pub. Funded)	39	0	39	0	0	0	0
Private Non Bilingual (Pub. Funded)	40	0	40	0	0	0	0
Private	8	0	8	0	0	0	0
Total	**167**	**0**	**167**	**0**	**0**	**0**	**0**

Abu Dhabi, UAE

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 10), and instruction language other than English or Arabic
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 4'/'Grade 4 and Grade 8' schools and school type (public, private) and main curriculum (UK/US/CAD/AUS/International, Ministry of Education, other) within private schools
- Implicit stratification by school size (small, large) and region (Abu Dhabi, Al Ain, Al Dhafra) within large private schools
- Sampled two classrooms whenever possible
- Grade 4 and Grade 8 samples were selected simultaneously with maximum overlap
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- All public schools and all private schools with UK/US/CAD/AUS/International main curriculum were sampled
- In census strata, schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 - Public - ADEC schools	67	1	66	0	0	0	0
Grade 4 - Private - UK/US/CAD/AUS/International Curriculum	23	0	23	0	0	0	0
Grade 4 - Private - Other Curriculum	6	1	5	0	0	0	0
Grade 4 & Grade 8 - Public - ADEC schools	22	0	22	0	0	0	0
Grade 4 & Grade 8 - Private - Ministry of Education	20	0	20	0	0	0	0
Grade 4 & Grade 8 - Private - UK/US/CAD/AUS/International Curriculum	87	0	87	0	0	0	0
Grade 4 & Grade 8 - Private - Other Curriculum	24	0	24	0	0	0	0
Total	**249**	**2**	**247**	**0**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 10), and instruction language other than English or Arabic
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private) and main curriculum (UK/US/CAD/AUS/International, Ministry of Education, other) within private schools
- Implicit stratification by school size (small, large) and region (Abu Dhabi, Al Ain, Al Dhafra) within large private schools

- Sampled two classrooms whenever possible
- Grade 4 and Grade 8 samples were selected simultaneously with maximum overlap
- The Field Test and Main Data Collection samples were selected sequentially. The Main Data Collection sample was selected by controlling for the overlap with the Field Test sample using the Chowdhury approach.
- All public schools and all private schools with UK/US/CAD/AUS/International main curriculum were sampled

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8 - Public	72	0	72	0	0	0	0
Grade 8 - Private	4	0	4	0	0	0	0
Grade 4 & Grade 8 - Public - ADEC schools	23	0	23	0	0	0	0
Grade 4 & Grade 8 - Private - Ministry of Education	20	0	20	0	0	0	0
Grade 4 & Grade 8 - Private - UK/US/CAD/AUS/Int	87	0	87	0	0	0	0
Grade 4 & Grade 8 - Private - Others	24	0	24	0	0	0	0
Total	**230**	**0**	**230**	**0**	**0**	**0**	**0**

Dubai, UAE

Fourth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of instruction language other than English or Arabic
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 4'/'Grade 4 and Grade 8' schools and school type (public, private)
- Implicit stratification by school size (small, large) and language of test (Arabic, English, French)
- Sampled two classrooms whenever possible
- Grade 4 and Grade 8 samples were selected simultaneously with maximum overlap
- The Field Test and Main Data Collection samples were selected sequentially
- Census of all schools
- Schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 4 - Public	24	3	21	0	0	0	0
Grade 4 - Private	47	0	47	0	0	0	0
Grade 4 & Grade 8 - Public	4	1	3	0	0	0	0
Grade 4 & Grade 8 - Private	130	2	128	0	0	0	0
Total	**205**	**6**	**199**	**0**	**0**	**0**	**0**

Eighth Grade

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of instruction language other than English or Arabic
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by 'Grade 8'/'Grade 4 and Grade 8' schools and school type (public, private)
- Implicit stratification by school size (small, large) and language of test (Arabic, English, French)
- Sampled two classrooms whenever possible
- Grade 4 and Grade 8 samples were selected simultaneously with maximum overlap
- The Field Test and Main Data Collection samples were selected sequentially
- Census of all schools
- Schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates

School Participation Status

			Participating Schools				
Explicit Strata	Total Sampled Schools	Ineligible Schools	Original Schools	1st Replacements	2nd Replacements	Refusal Schools	Excluded Schools
Grade 8 - Public	23	3	20	0	0	0	0
Grade 8 - Private	7	0	7	0	0	0	0
Grade 4 & Grade 8 - Public	4	1	3	0	0	0	0
Grade 4 & Grade 8 - Private	137	4	133	0	0	0	0
Total	**171**	**8**	**163**	**0**	**0**	**0**	**0**

CHAPTER 10

Reviewing the TIMSS 2019 Achievement Item Statistics

Bethany Fishbein
Pierre Foy
Lillian Tyack

Overview

Prior to scaling and reporting the results for an assessment, the TIMSS & PIRLS International Study Center reviews key diagnostic statistics for each achievement item to evaluate its psychometric characteristics across the participating countries. This item-by-item, country-by-country review to detect unusual item properties or anomalous patterns plays a crucial role in the quality assurance of the achievement data. Finding a faulty item this late in the process is rare, but an uncharacteristically difficult item, or one with unusually low discriminating power, could indicate a potential problem with either translation or printing. Similarly, a human-scored constructed response item with low scoring reliability could indicate a problem in the translation of the scoring guide for a particular country. If such an item is found, the country's translation verification documents, printed booklets, and digital item archives can be examined for flaws or inaccuracies and, if necessary, the item can be removed from the international database for that country.

The TIMSS 2019 assessment cycle marked the beginning of the transition to eTIMSS—a digital version of the TIMSS assessment that was administered to students on computers and tablets. eTIMSS also included a novel section consisting of problem solving and inquiry tasks (PSIs), which were designed to capitalize on the digital environment to its fullest. About half the participating countries chose to transition to eTIMSS. eTIMSS trend countries also administered the paper version of their trend items to a sample of schools, providing a "bridge" that helped link the two test-taking modes. For TIMSS 2019, comparing the item statistics for eTIMSS and paperTIMSS was integral in identifying items that were psychometrically invariant (equivalent) under the IRT scaling.

The TIMSS 2019 Item Review

The TIMSS & PIRLS International Study Center computed item statistics for all achievement items in the 2019 fourth and eighth grade assessments, including both eTIMSS and paperTIMSS versions, including the eTIMSS PSIs and the paper "bridge" booklets administered in eTIMSS trend countries. TIMSS fourth grade included 175 mathematics items and 175 science items in both paper and digital formats, 131 less difficult mathematics items in paper only, and 39 mathematics items from three PSIs and 19 science items from two PSIs.[1] The fourth grade paper bridge booklets consisted of only trend items—92 mathematics items and 111 science items. TIMSS eighth grade included 211 mathematics items and 220 science items in both paper and digital formats, plus 25 mathematics items from three PSIs and 29 science items from two PSIs. The bridge booklets consisted of 117 mathematics and 122 science trend items at the eighth grade.

In addition to evaluating the performance of each item, the TIMSS & PIRLS International Study Center looked for any changes in the measurement properties of trend items from the 2015 assessment, and examined differences between items common to eTIMSS and the paper bridge booklets for mode effects. Item position effects were evaluated to ensure student performance remained steady throughout the assessment. Finally, using the item statistics, extensive analyses of each country's data were conducted to detect any anomalous patterns relative to previous cycles or the pool of participating countries on average.

Although reviewing item statistics took place over several months, staff at the TIMSS & PIRLS International Study Center met for four consecutive working days in March 2020 to conduct a formal adjudication of the achievement data in preparation for IRT scaling. During these four days, decisions were made about any modifications needed to the data or if further analyses were required. The review was conducted item-by-item simultaneously for eTIMSS and paperTIMSS. Both versions of an item and its scoring guide were displayed while staff reviewed the item statistics as well as accompanying graphical displays. Country reports about translation errors, printing issues, or other technical problems were also referenced. In addition, graphical displays of item statistics were reviewed to detect any anomalous and systematic patterns in a particular country's data that may warrant further investigation.

Following item review, some National Research Coordinators from the participating countries and benchmarking entities were contacted to inquire about concerns or anomalies detected in the data. Decisions about item deletions or recodes were communicated to IEA Hamburg to make edits to the international data files.

Item Review Statistics

The item statistics for each of the TIMSS 2019 participating countries were computed and combined to produce "item almanacs" for eTIMSS and paperTIMSS respectively. Each item almanac page included unweighted statistics for all countries that administered the particular item. The paperTIMSS item almanacs included data from paperTIMSS countries, countries participating in the fourth grade less

1 Two fourth grade mathematics items involving an on-screen ruler tool were only included in the eTIMSS assessment. Forty-eight fourth grade mathematics items were also included in the less difficult version of the assessment.

difficult mathematics assessment, and the bridge samples from eTIMSS trend countries were also reported with paperTIMSS countries. Exhibits 10.1 and 10.2 show actual samples of the statistics calculated for a selected response item and a constructed response item, respectively.

Exhibit 10.1: Example International Item Statistics for a TIMSS 2019 Selected Response Item

Trends in International Mathematics and Science Study - TIMSS 2019 Assessment Results - 4th Grade
International Item Review Statistics (Unweighted)

Mathematics: Number / Reasoning (MP03_05 - MP61228) - 2 Points
Label: Art teacher cuts paper for her class

				Percentages					Point Biserials						Reliability			
Country	N	DIFF	DISC	P_0	P_1	P_2	P_OM	P_NR	PB_0	PB_1	PB_2	PB_OM	PB_NR	RDIFF	N	Score	Code	Flags
Albania (LD)	637	17.2	0.49	55.9	17.4	8.5	18.2	0.6	-0.22	0.32	0.34	-0.27	-0.02	2.48	218	99.5	99.5	__H_F__
Armenia	759	11.2	0.39	36.0	15.9	3.2	44.8	5.7	0.00	0.30	0.23	-0.30	-0.01	1.79	217	100.0	100.0	__HAF__
*Australia	835	33.5	0.55	48.6	22.6	22.2	6.6	0.7	-0.42	0.21	0.44	-0.24	-0.02	0.91	219	95.4	95.4	__E____
Austria (Br)	494	18.5	0.57	38.9	17.2	9.9	34.0	0.0	-0.29	0.36	0.40	-0.24	.	1.71	218	96.3	95.4	__H_F__
Azerbaijan	743	12.8	0.47	49.1	10.6	7.5	32.7	1.3	-0.04	0.26	0.36	-0.33	-0.20	2.14	191	99.5	99.5	__H_F__
Bahrain	826	22.1	0.54	57.2	16.8	13.7	12.3	0.7	-0.32	0.13	0.49	-0.19	0.02	0.69	231	99.6	98.7	__E___G
Belgium (Flemish)	663	23.1	0.54	56.0	23.9	11.2	8.9	0.3	-0.44	0.37	0.34	-0.17	0.02	1.51	213	88.7	87.8	___A___
Bosnia and Herzegovina (LD)	802	6.7	0.37	47.5	8.3	2.5	41.7	1.9	-0.09	0.27	0.24	-0.14	0.02	2.50	207	98.6	98.1	__HAF__
Bulgaria	607	15.3	0.50	56.7	6.8	11.8	24.7	1.2	-0.17	0.19	0.44	-0.25	-0.12	1.78	205	98.0	98.0	__H_F__
*Canada (Br)	395	25.7	0.63	50.4	26.7	12.3	10.5	1.5	-0.46	0.33	0.45	-0.21	-0.04	1.05	195	89.2	87.7	__E____
Chile (Br)	410	12.6	0.48	56.0	10.9	7.1	26.0	4.1	-0.19	0.26	0.37	-0.19	-0.01	0.81	209	100.0	100.0	__E_F__
Chinese Taipei (Br)	420	48.8	0.60	31.9	19.0	39.3	9.8	0.0	-0.39	0.06	0.53	-0.34	.	1.41	217	94.9	94.9	______G
Croatia (Br)	358	14.7	0.53	57.5	14.2	7.5	20.7	0.0	-0.31	0.31	0.40	-0.14	.	1.59	177	98.9	97.7	____F__
*Cyprus	587	21.5	0.55	54.4	12.0	15.6	18.1	0.3	-0.32	0.24	0.46	-0.23	0.03	1.41	207	89.4	88.9	_______
*Czech Republic (Br)	499	18.7	0.56	42.8	13.1	12.1	31.9	0.8	-0.22	0.29	0.44	-0.28	-0.10	1.37	201	95.0	93.0	_______
Denmark (Br)	345	32.2	0.51	38.9	29.5	17.5	14.2	3.8	-0.35	0.22	0.38	-0.22	-0.03	1.07	181	92.8	92.8	__E____
*England (Br)	305	33.2	0.65	44.6	17.5	24.4	13.5	0.7	-0.40	0.09	0.60	-0.28	-0.06	1.24	165	93.3	93.3	__E____
Finland (Br)	492	23.9	0.61	43.4	23.7	12.1	20.9	0.6	-0.34	0.33	0.44	-0.28	-0.13	1.48	251	96.8	96.8	_______
France (Br)	482	23.0	0.55	45.7	20.5	12.7	21.1	0.6	-0.35	0.29	0.41	-0.20	-0.05	0.62	247	89.1	88.7	__E____
Georgia (Br)	404	24.9	0.63	44.6	12.5	18.7	24.2	0.7	-0.33	0.29	0.52	-0.32	-0.16	0.89	203	99.5	98.5	__E____
Germany (Br)	373	22.2	0.58	42.7	15.3	14.5	27.4	0.3	-0.33	0.32	0.44	-0.24	-0.10	1.21	200	95.0	94.0	__E____
*Hong Kong SAR (Br)	339	39.4	0.62	41.6	18.6	30.1	9.7	0.0	-0.46	0.13	0.54	-0.24	.	1.62	165	98.8	98.8	_______
*Hungary (Br)	447	26.2	0.58	50.8	20.1	16.1	13.0	0.0	-0.38	0.20	0.48	-0.20	.	1.39	198	98.5	98.0	_______
*Iran, Islamic Rep. of	859	9.8	0.50	55.1	13.0	3.3	28.6	4.4	-0.12	0.38	0.30	-0.26	-0.02	1.64	216	95.8	91.7	___AF__
*Ireland	646	32.9	0.61	48.0	22.0	21.9	8.1	0.3	-0.45	0.18	0.51	-0.22	0.01	1.23	230	99.1	99.1	__E____
*Italy (Br)	474	22.3	0.57	47.5	17.3	13.6	21.5	1.1	-0.27	0.30	0.43	-0.31	-0.05	1.18	243	94.2	93.0	__E____
*Japan	604	59.6	0.62	31.2	12.0	53.7	3.2	0.3	-0.55	-0.01	0.57	-0.18	-0.07	0.84	235	91.5	91.5	__E___G
Kazakhstan	692	19.5	0.47	54.8	25.4	6.9	13.0	0.9	-0.33	0.35	0.27	-0.17	0.00	1.32	141	90.1	87.2	__EAF__
*Korea, Rep. of (Br)	391	47.6	0.62	41.9	7.2	44.0	6.9	0.0	-0.51	0.00	0.60	-0.20	.	1.35	215	98.6	98.6	____F__
Kosovo (LD)	642	11.0	0.44	60.3	10.4	5.8	23.5	1.1	-0.12	0.35	0.28	-0.27	-0.09	2.03	224	92.0	90.2	__HAF__
Kuwait (LD)	630	8.4	0.45	61.4	9.8	3.5	25.2	1.3	-0.05	0.34	0.29	-0.30	-0.02	1.92	206	100.0	99.0	__HAF_G
Latvia	635	29.1	0.58	44.1	22.6	17.9	15.5	0.3	-0.38	0.22	0.47	-0.23	-0.03	1.41	204	91.7	91.2	_______
Lithuania (Br)	396	26.5	0.60	45.7	21.1	16.0	17.3	0.5	-0.33	0.28	0.46	-0.31	-0.07	1.50	213	94.8	94.8	_______
Montenegro (LD)	728	6.6	0.33	38.4	4.6	4.3	52.7	1.6	0.07	0.17	0.28	-0.25	0.02	2.40	233	98.3	97.9	__H_F__
Morocco (LD)	1091	5.4	0.38	54.7	8.7	1.0	35.5	1.5	-0.05	0.38	0.10	-0.19	0.10	2.27	197	90.4	87.3	__HAF__
*Netherlands (Br)	324	35.4	0.60	42.1	19.8	25.5	12.6	1.9	-0.45	0.19	0.50	-0.22	-0.07	0.81	166	86.7	86.1	__E____
*New Zealand	718	25.9	0.55	55.5	20.2	15.8	8.4	0.7	-0.43	0.25	0.43	-0.17	-0.07	0.82	242	96.3	95.9	__E____
North Macedonia (LD)	485	12.4	0.45	47.5	8.4	8.2	35.9	2.3	-0.05	0.24	0.36	-0.30	-0.12	2.33	168	97.6	97.6	__H_F__
Northern Ireland	500	42.2	0.65	39.1	21.4	31.5	8.0	0.2	-0.50	0.18	0.53	-0.29	-0.05	1.11	187	100.0	98.9	__E___G
*Norway (5) (Br)	465	37.8	0.57	35.2	28.7	23.5	12.6	1.1	-0.39	0.26	0.41	-0.31	-0.00	0.92	222	93.7	92.8	__E___G
Oman	975	9.9	0.55	59.1	8.4	5.7	26.7	1.3	-0.20	0.36	0.40	-0.22	-0.00	1.08	216	96.3	92.1	__E_F__
Pakistan (LD)	566	0.7	0.17	29.7	1.4	0.0	68.8	2.5	0.19	0.17	.	-0.23	0.03	3.52	244	100.0	100.0	__HAF__
Philippines (LD)	787	2.0	0.34	67.8	1.9	1.0	29.2	1.7	0.12	0.31	0.20	-0.26	-0.08	1.84	219	98.6	96.3	__HAF__
Poland	704	27.5	0.64	37.1	19.3	17.8	25.7	0.1	-0.31	0.28	0.51	-0.35	-0.02	1.11	183	98.9	98.9	__E___G
*Portugal (Br)	397	31.8	0.68	47.4	19.1	22.2	11.2	1.3	-0.46	0.12	0.61	-0.22	-0.03	1.05	196	96.9	96.9	__E____
Qatar (Br)	369	13.1	0.58	65.8	9.8	8.2	16.1	0.8	-0.30	0.30	0.47	-0.21	0.03	1.13	195	95.4	94.4	__E_F__
Russian Federation (Br)	547	33.9	0.64	41.0	15.2	26.3	17.6	0.0	-0.40	0.24	0.53	-0.32	.	1.33	219	99.1	98.2	_______
Saudi Arabia (LD)	781	7.2	0.38	66.4	10.1	2.2	21.3	0.9	-0.06	0.35	0.17	-0.25	0.03	2.26	200	95.0	90.5	__HAF_G
Serbia	629	17.7	0.58	46.0	12.1	11.7	30.2	0.5	-0.22	0.27	0.48	-0.29	-0.07	1.52	206	97.1	96.1	_______
*Singapore (Br)	467	57.4	0.66	35.5	9.9	52.5	2.1	0.0	-0.57	0.00	0.61	-0.23	.	1.65	232	100.0	100.0	__H_F__
Slovak Republic (Br)	402	22.4	0.59	47.0	12.9	15.9	24.1	0.0	-0.27	0.21	0.51	-0.29	.	0.97	211	99.1	99.1	__E____
South Africa (5) (LD)	1692	4.9	0.44	66.8	6.8	1.5	24.9	0.7	0.03	0.40	0.21	-0.32	-0.03	2.38	232	98.7	97.0	__HAF_G
Spain (Br)	420	20.7	0.51	50.1	22.8	9.4	17.7	0.7	-0.33	0.37	0.30	-0.21	-0.06	1.34	202	95.5	94.1	___AF__
Sweden (Br)	425	31.2	0.62	37.0	28.0	17.1	17.9	2.6	-0.39	0.30	0.45	-0.31	-0.14	0.88	209	91.9	90.4	__E____
United Arab Emirates (Br)	568	21.6	0.63	54.4	23.4	9.9	12.2	0.7	-0.45	0.43	0.40	-0.23	-0.05	1.36	248	94.0	93.5	___AF__
*United States (Br)	416	41.4	0.66	44.8	21.2	30.8	3.2	2.4	-0.61	0.22	0.53	-0.14	-0.08	0.82	185	93.0	93.0	__E____
*Reference Avg (18)	9168	33.3	0.60	45.4	17.8	24.4	12.3	1.0	-0.42	0.19	0.50	-0.23	-0.04	1.18	3732	94.8	94.1	_______
International Avg (56)	32647	23.3	0.54	48.0	15.9	15.3	20.8	1.1	-0.29	0.25	0.42	-0.25	-0.04	1.46	11674	95.8	95.0	______G

Keys: DIFF= Percent correct score; DISC= Item discrimination; P_0...P_2= Percentage obtaining score level; P_OM, P_NR= Percentage Omitted, Not Reached;
PB_0...PB_2= Point Biserial for score level; PB_OM, PB_NR= Point Biserial for Omitted, Not Reached; RDIFF= Rasch difficulty;
Reliability: N= Responses double scored; Score= Percentage agreement on score; Code= Percentage agreement on code.
Flags: A= Ability not ordered/Attractive distractor; B= Boys outperform girls; C= Difficulty less than chance; D= Negative/low discrimination; E= Easier than average;
F= Score obtained by less than 10%; G= Girls outperform boys; H= Harder than average; R= Scoring reliability less than 85%; V= Difficulty greater than 95%.

Exhibit 10.2: Example International Item Statistics for a TIMSS 2019 Constructed Response Item

Trends in International Mathematics and Science Study - eTIMSS 2019 Assessment Results - 8th Grade
International Item Review Statistics (Unweighted)

Science: Biology / Knowing (SE06_02 - SE62274) - 2 Points
Label: Raw materials for photosynthesis

				Percentages					Point Biserials						Reliability			
Country	N	DIFF	DISC	P_0	P_1	P_2	P_OM	P_NR	PB_0	PB_1	PB_2	PB_OM	PB_NR	RDIFF	N	Score	Code	Flags
*Chile	588	38.8	0.45	17.2	59.1	9.2	14.5	0.2	-0.19	0.26	0.24	-0.36	-0.02	0.11	197	97.5	97.5	__EAF_B
*Chinese Taipei	700	35.7	0.59	39.6	39.7	15.9	4.9	0.0	-0.45	0.24	0.42	-0.23	.	1.21	195	94.4	94.4	__H___B
*England	492	37.1	0.40	25.8	51.0	11.6	11.6	0.0	-0.11	0.13	0.30	-0.36	.	0.56	191	99.5	99.5	__H____
Finland	677	50.6	0.43	14.5	61.8	19.7	4.0	0.3	-0.20	-0.06	0.37	-0.25	-0.07	0.05	200	99.5	99.5	__E____
France	557	28.1	0.35	25.4	48.9	3.6	22.1	0.2	-0.05	0.31	0.10	-0.37	-0.10	1.02	200	99.5	99.5	__HAF__
Georgia	472	34.3	0.48	17.2	51.6	8.5	22.7	0.2	-0.12	0.32	0.25	-0.43	-0.01	0.03	183	98.9	98.9	__EAF__
*Hong Kong SAR	469	39.5	0.57	30.3	44.4	17.3	7.9	0.2	-0.36	0.12	0.46	-0.25	0.00	0.12	195	100.0	100.0	__E____
*Hungary	652	38.7	0.46	24.5	53.5	12.0	10.0	0.0	-0.21	0.18	0.32	-0.34	.	0.66	199	94.5	94.5	__H___B
*Israel	533	38.6	0.54	21.2	54.8	11.3	12.8	0.0	-0.22	0.24	0.35	-0.42	.	0.50	196	97.4	97.4	_______
*Italy	529	38.1	0.24	20.6	64.8	5.7	8.9	0.0	-0.03	0.17	0.11	-0.32	.	0.60	200	100.0	100.0	__HAF__
*Korea, Rep. of	553	39.2	0.50	27.1	55.7	11.4	5.8	0.0	-0.32	0.24	0.31	-0.31	.	0.92	200	97.0	97.0	__H____
*Lithuania	546	46.4	0.53	14.7	61.4	15.8	8.2	0.0	-0.19	-0.02	0.46	-0.33	.	0.10	200	98.0	98.0	__E____
*Malaysia	1018	38.1	0.50	30.9	58.5	8.8	1.7	0.0	-0.43	0.27	0.29	-0.14	.	0.31	200	99.0	99.0	____F__
Norway (9)	627	43.9	0.48	13.4	62.6	12.6	11.4	0.3	-0.21	0.18	0.31	-0.37	-0.08	0.02	198	96.0	96.0	__E____
Portugal	480	40.7	0.37	19.4	61.9	9.8	9.0	0.0	-0.09	0.15	0.24	-0.39	.	0.41	193	99.0	99.0	____F__
Qatar	546	37.8	0.50	27.9	52.8	11.4	7.9	0.2	-0.28	0.20	0.35	-0.31	-0.02	0.14	200	97.5	97.5	__E____
*Russian Federation	555	47.5	0.41	17.7	58.5	18.2	5.6	0.2	-0.20	0.06	0.30	-0.31	0.00	0.25	200	98.0	98.0	_______
*Singapore	689	54.5	0.54	14.7	59.0	25.0	1.3	0.1	-0.39	-0.04	0.42	-0.21	-0.03	0.41	200	100.0	100.0	_______
*Sweden	548	53.0	0.50	14.5	46.3	29.9	9.3	0.4	-0.21	-0.06	0.43	-0.33	0.06	-0.20	198	99.0	99.0	__E____
*Turkey	582	41.8	0.58	25.8	46.2	18.7	9.3	0.0	-0.30	0.08	0.48	-0.33	.	0.21	197	98.5	98.5	__E____
United Arab Emirates	3185	36.4	0.57	32.5	45.5	13.6	8.3	0.0	-0.39	0.26	0.38	-0.27	0.02	0.18	199	97.5	97.5	__E____
*United States	1249	38.8	0.45	26.9	61.8	7.9	3.4	0.2	-0.31	0.22	0.28	-0.25	-0.03	0.69	198	99.5	99.5	__H_F__
*Reference Avg (15)	9703	41.7	0.48	23.4	54.3	14.6	7.7	0.1	-0.26	0.14	0.35	-0.30	-0.00	0.43	198	98.1	98.1	______B
International Avg (22)	16247	40.8	0.47	22.8	54.5	13.5	9.1	0.1	-0.24	0.16	0.33	-0.31	-0.02	0.38	197	98.2	98.2	______B
Ontario, Canada	535	37.1	0.35	22.6	64.9	4.7	7.9	0.0	-0.16	0.25	0.16	-0.33	.	0.79	117	100.0	100.0	__HAF__
Quebec, Canada	455	42.2	0.42	22.7	61.5	11.5	4.4	0.2	-0.30	0.19	0.25	-0.24	-0.03	0.56	80	96.3	96.3	__H____
Moscow City, Russian Fe	536	45.9	0.37	16.8	63.8	14.0	5.4	0.0	-0.22	0.17	0.20	-0.30	.	0.75	200	99.5	99.5	__H____
Abu Dhabi, UAE	1166	27.6	0.65	43.7	35.2	10.0	11.1	0.0	-0.44	0.36	0.43	-0.27	.	0.26	70	100.0	100.0	__E____
Dubai, UAE	817	46.9	0.48	21.4	53.9	20.0	4.7	0.1	-0.33	0.12	0.33	-0.28	0.02	0.25	67	98.5	98.5	__E____

Keys: DIFF= Percent correct score; DISC= Item discrimination; P_0...P_2= Percentage obtaining score level; P_OM, P_NR= Percentage Omitted, Not Reached;
PB_0...PB_2= Point Biserial for score level; PB_OM, PB_NR= Point Biserial for Omitted, Not Reached; RDIFF= Rasch difficulty;

Reliability: N= Responses double scored; Score= Percentage agreement on score; Code= Percentage agreement on code.
Flags: A= Ability not ordered/Attractive distractor; B= Boys outperform girls; C= Difficulty less than chance; D= Negative/low discrimination; E= Easier than average;
F= Score obtained by less than 10%; G= Girls outperform boys; H= Harder than average; R= Scoring reliability less than 85%; V= Difficulty greater than 95%.

For all items, regardless of format (i.e., selected response or constructed response) or administration mode, statistics included the number of students that responded in each country, the difficulty level (the percentage of students that answered the item correctly), and the discrimination index (the point-biserial correlation between success on the item and total score).[2] Also provided was an estimate of the difficulty of the item using a Rasch one-parameter IRT model. Statistics for each item were displayed alphabetically by country, together with an international average—i.e., based on all participating countries listed above the international average—and a reference average—based on a pool of countries that have participated regularly in the TIMSS assessments—for each statistic. The reference countries are shown with an asterisk next to their names. The international and reference averages of the item difficulties and item discriminations served as guides to the overall statistical properties of the items. The item review outputs also listed the benchmarking participants.

Statistics displayed for selected response items included the percentage of students that chose each response option—as well as the percentage of students that omitted or did not reach the item—and the point-biserial correlations for each response option. Statistics displayed for constructed response items (which could have 1 or 2 score points) included the percent correct and point-biserial of each score level. Constructed response item tables also provided information about the reliability with which each item was scored in each country, showing the total number of double-scored responses, the percentage of score

2 For computing point-biserial correlations, the total score is the percentage of points a student has scored on the items they were administered. In the context of TIMSS, a separate total score is computed for mathematics and for science. Not-reached responses are not included in the total score.

agreement between the scorers, and—because TIMSS has a 2-digit scoring scheme for diagnostic coding—the percentage of code agreement between scorers. Diagnostic codes between 20 and 29 are worth 2 score points, between 10 and 19 worth 1 score points, and between 70 and 79 worth 0 score points.

During item review, "not reached" responses (i.e., items toward the end of the booklet that the student did not attempt)[3] were treated as "not administered" and thus did not contribute to the calculation of the item statistics. However, the percentage of students not reaching each item was reported. Omitted responses, although treated as incorrect, were tabulated separately from incorrect responses for the sake of distinguishing students who provided no form of response from students who attempted a response.

The definitions and detailed descriptions of the statistics that were calculated are given below. The statistics were calculated separately by grade and subject, and are listed in order of their appearance in the item review outputs:

- **N:** This is the number of students to whom the item was administered. Not-reached responses were not included in this count.
- **DIFF:** The item difficulty is the average percent correct on an item. For a 1-point item, including all selected response items, it is the percentage of students providing a fully correct response to the item. For 2-point items, it is the average percentage of points. For example, if 25 percent of students scored 2 points, 50 percent scored 1 point, and the other 25 percent scored 0 points, then the average percent correct would be 50 percent. For this statistic, not-reached responses were not included.
- **DISC:** The item discrimination is computed as the correlation between the response to an item and the total score on all items administered to a student. Items exhibiting good measurement properties should have a moderately positive correlation, indicating that the more able students get the item right, the less able get it wrong. For this statistic, not-reached items were not included.
- **Percentages (P_A, P_B, P_C, P_D, etc.):** Available for selected-response items. Each column indicates the percentage of students choosing the particular response option for the item (e.g., A, B, C, D, etc.). Not-reached responses were excluded from the denominator.
- **Percentages (P_0, P_1, and P_2):** Available for constructed response items. Each column indicates the percentage of students responding at that particular score level, up to and including the maximum score level for the item. Not-reached items were excluded from the denominator.
- **Percentages (P_OM):** Percentage of students who, having reached the item, did not provide a response. Not reached responses were excluded from the denominator.
- **Percentages (P_NR):** Percentage of students who did not reach the item. This statistic is the number of students who did not reach an item as a percentage of all students who were administered that item, including those who omitted or did not reach that item.

3 An item was considered "not reached" if the item itself and the item immediately preceding it were not answered and no subsequent items were answered. The decision as to whether an item was not reached was made separately for part 1 and part 2 of each assessment booklet.

- **Point Biserials (PB_A, PB_B, PB_C, PB_D, etc.):** Available for selected-response items. These columns show the point-biserial correlations between choosing each of the response options (e.g., A, B, C, D, etc.) and the total score on all of the items administered to a student. Items with good psychometric properties have moderately positive correlations for the correct option and negative correlations for the distracters (the incorrect options). Not-reached responses were not included in these calculations.
- **Point Biserials (PB_0, PB_1, and PB_2):** Available for constructed response items. These columns present the point-biserial correlations between the score levels on the item (0, 1, or 2) and the overall score on all of the items the student was administered. For items with good measurement properties, the correlation coefficients should monotonically increase from negative to positive as the score on the item increases. Not-reached responses were not included in these calculations.
- **Point Biserials (PB_OM):** The point-biserial correlation between a binary variable indicating an omitted response to the item, and the total score on all items administered to a student. This correlation should be negative or near zero. Not-reached responses were not included in this statistic.
- **Point Biserials (PB_NR):** The point-biserial correlation between a binary variable indicating a not-reached response to the item, and the total score on all items administered to a student. This correlation should be negative or near zero.
- **RDIFF:** An estimate of the difficulty of an item based on a Rasch one-parameter IRT model applied to the achievement data of a given country. The difficulty estimate is expressed in the logit metric (with a positive logit indicating a difficult item) and was scaled so that the average Rasch item difficulty across all items within each country was zero.
- **Reliability (N):** Available for human-scored constructed response items. To provide a measure of the reliability of the scoring of the constructed response items, approximately 200 responses per item were independently scored by two scorers. This column indicates the number of responses that were double-scored for a given item in a country.
- **Reliability (Score):** Available for human-scored constructed response items. This column contains the percentage of agreement on the score point value (0, 1, or 2) of the two-digit diagnostic codes assigned by the two independent TIMSS scorers.
- **Reliability (Code):** Available for human-scored constructed response items. This column contains the percentage of full agreement on the two-digit diagnostic codes assigned by the two independent TIMSS scorers.

As an aid to the reviewers, the item-review displays included a series of flags signaling the presence of one or more conditions that might indicate a problem with an item. The flags rarely indicate and actual problem, but serve to draw attention to potential sources of concern. The following conditions were flagged:

- The item discrimination (DISC) was less than 0.10 (flag D)
- The item difficulty (DIFF) was less than 0.25 for selected response items (flag C)
- The item difficulty (DIFF) exceeded 0.95 (flag V)
- The Rasch difficulty estimate (RDIFF) for a given country showed the item either easier (flag E) or more difficult (flag H) relative to the international average for that item
- The point-biserial correlation for at least one distracter in a selected response item was positive, or the point-biserial correlations across the score levels of a constructed response item were not ordered (flag A)
- The percentage of students selecting one of the response options for a selected response item, or one of the score values for a constructed response item, was less than 10 percent (flag F)
- Scoring reliability for agreement on the score value of a constructed response item was less than 85 percent (flag R).

Scoring Reliability for Human Scored Items

Constructed-response items comprised about half the assessment score points in TIMSS 2019, with most of them requiring human scoring, especially for paperTIMSS and the paper bridge. To ensure that the items requiring human scoring were scored reliably in all countries, the TIMSS & PIRLS International Study Center developed detailed scoring guides for each constructed response item (that provided descriptions and examples of acceptable responses for each score point value) and provided extensive training in the application of the scoring guides. See Chapter 1 for more information on developing the scoring guides and see Chapter 6 for information on the human-scoring process.

For eTIMSS countries, the new mode of administration allowed for a substantial portion of the digital items to be machine scored, particularly in mathematics. For eTIMSS items suitable for machine scoring, the scoring guides served as the basis for developing machine scoring specifications for student responses that could be accurately applied without human judgment. Developing the machine scoring specifications involved testing each item in the eTIMSS Player, reviewing the output, and writing rules in terms of the output to classify all possible responses to a code in the item's scoring guide. The scoring unit at IEA Hamburg reviewed all specifications and provided feedback on an item-by-item basis, resulting in several rounds of revision until the rules for all items were clarified. The scoring unit at IEA Hamburg then applied the scoring rules for all machine-scored items and the data analysis team at the TIMSS & PIRLS International Study Center independently replicated the results to validate the scoring.

The following sections describe the three ways human-scoring reliability was assessed and documented in 2019: within-country, over-time (trend), and across countries.

Within-Country Scoring Reliability

To gather and document information about the within-country agreement among scorers for TIMSS 2019, a random sample of approximately 200 student responses per item were scored independently by two scorers. The inter-scorer agreement for each item in each country was examined as part of the item review process, with agreement below 75 percent giving cause for deleting the data for a particular country. Appendix 10A shows the average and range of the within-country percentages of score point agreement and diagnostic code agreement across all items. Exact percent agreement across items was high on average across countries—98 percent in mathematics and 95 percent in science across both the fourth and eighth grade countries. In TIMSS 2019 there also was high agreement at the diagnostic score level, where international average percent agreement ranged from 94 percent in eighth grade science to 97 percent in mathematics at both grades.

Trend Item Scoring Reliability Study

The TIMSS & PIRLS International Study Center also took steps to show that the 2019 human-scored constructed response items used in TIMSS 2015 were scored in the same way in both assessments. In anticipation of this, countries that participated in TIMSS 2015 sent samples of scored student booklets from the 2015 data collections to IEA Hamburg, where they were digitally scanned and stored for later use. As a check on scoring consistency from one administration to the next, staff members working in each country on scoring the 2019 data were asked also to score these 2015 responses using the CodingExpert Software developed by IEA Hamburg. Each country scored 200 responses for each of 11 mathematics items (13 items for countries that administered less difficult mathematics) and 11 science items at the fourth grade, and for 14 mathematics items and 13 science items at the eighth grade. The average and range of scoring consistency over time can be found in Appendix 10B.

There was a very high degree of scoring consistency in TIMSS 2019. The exact agreement between the scores awarded in 2015 and those given by the 2019 scorers ranged from 91 percent in science at both grades to 97 percent in mathematics at the fourth grade, on average internationally. There was similarly high agreement in TIMSS at the diagnostic code level.

Cross-Country Scoring Reliability Study

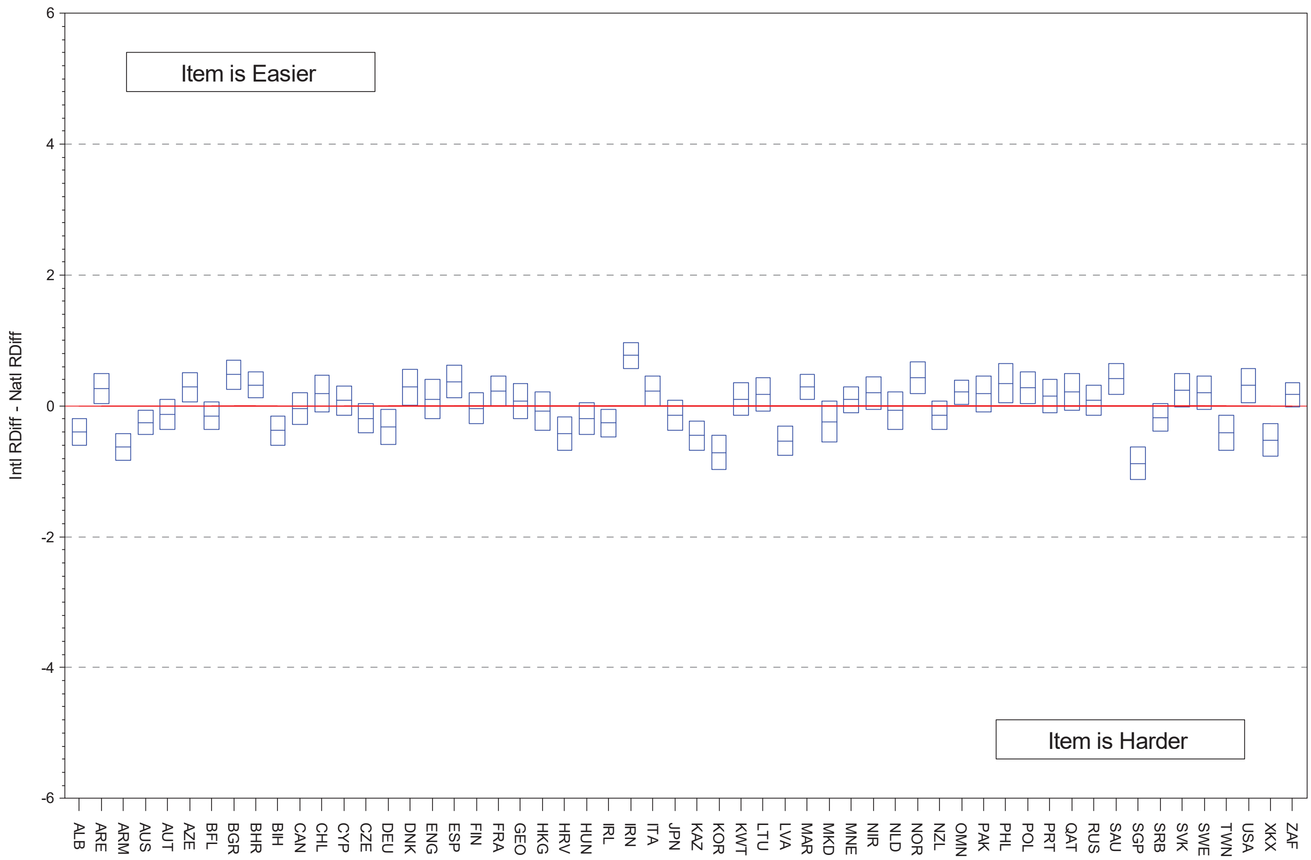
It also was important to document the consistency of scoring across countries. Because of the many different languages in use in TIMSS 2019, establishing the reliability of constructed response scoring across all countries was not feasible. However, the TIMSS & PIRLS International Study Center did conduct a cross-country study of scoring reliability among Northern Hemisphere countries that had scorers who were proficient in English. A sample of student responses was provided by the English-speaking Southern Hemisphere countries. Cross-country scoring included 200 student responses for each of 11 mathematics items (6 items for countries that administered less difficult mathematics) and 11 science items at the fourth grade, and for 14 mathematics items and 13 science items at the eighth grade. This same set of student responses in English was then scored independently in each country that had at least two scorers proficient in English, using IEA's CodingExpert Software. In all, scorers from 54 countries and 1 benchmarking participant at fourth grade and 35 countries and 2 benchmarking participants at the eighth grade participated in the study. Scoring for this study took place shortly after the other scoring reliability activities were completed. Making all possible comparisons among scorers gave 1,485 comparisons at fourth grade and 666 comparisons at eighth grade for each student response to each item. This resulted in more than 290,000 total comparisons at the fourth grade (200,000 for mathematics items not administered to less difficult countries) and more than 120,000 total comparisons at the eighth grade when aggregated across all 200 student responses to that item. Agreement across countries was defined in terms of the percentage of these comparisons that were in exact agreement.

Appendix 10C reports the results of the cross-country scoring reliability study. On average internationally, scorer reliability across countries in TIMSS 2019 was high. The exact agreement between the scores awarded across countries ranged from 89 percent in science to 96 percent in mathematics at the fourth grade and from 89 percent in science to 94 percent in mathematics at the eighth grade, on average internationally. There was similarly high agreement at the diagnostic code level.

Item-by-Country Interactions

Although countries are expected to exhibit some variation in performance across items, in general countries with high average performance on the assessment should perform relatively well on each of the items, and low-scoring countries should do less well on each of the items. When this does not occur (e.g., when a high-performing country has low performance on an item on which other countries are doing well), there is said to be an item-by-country interaction. When large, such item-by-country interactions may be a sign that an item is flawed in some way and that steps should be taken to address the problem. To assist in detecting sizeable item-by-country interactions, the TIMSS & PIRLS International Study Center produced a graphical display for each item showing the difference between each country's Rasch item difficulty and the international average Rasch item difficulty across all countries. An example of the graphical displays is provided in Exhibit 10.3.

Exhibit 10.3: **Example Plot of Item-by-Country Interaction for a TIMSS 2019 Item**



In each of these item-by-country interaction displays, the difference in Rasch item difficulty for each country is presented as a 95 percent confidence interval, which includes a built-in Bonferroni correction for multiple comparisons across the participating countries. The limits for this confidence interval were computed as follows:

$$\text{Upper Limit} = RDIFF_{i.} - RDIFF_{ik} + SE(RDIFF_{ik}) \cdot Z_b \tag{10.1}$$

$$\text{Lower Limit} = RDIFF_{i.} - RDIFF_{ik} - SE(RDIFF_{ik}) \cdot Z_b \tag{10.2}$$

where $RDIFF_{ik}$ is the Rasch difficulty of item i in country k, $RDIFF_{i.}$ is the international average Rasch difficulty of item i, $SE(RDIFF_{ik})$ is the standard error of the Rasch difficulty of item i in country k, and Z_b is the 95 percent critical value from the Z distribution corrected for multiple comparisons using the Bonferroni procedure.

Review of Item Statistics for Measuring Trends

To measure trends, TIMSS assessments include achievement items from previous assessments as well as items developed for use for the first time in 2019. Accordingly, the TIMSS 2019 assessments included items developed in 2011, 2015, and 2019. Therefore, an important review step included checking that these "trend items" had statistical properties in 2019 similar to those they had in the previous assessments (e.g., a TIMSS item that was relatively easy in 2015 should still be relatively easy in 2019).

As shown in the example in Exhibit 10.4, the trend item review focused on statistics for paper trend items from the current and previous assessments (2019 and 2015) for countries that participated in both. This included statistics for the eTIMSS bridge samples. For each country, trend item statistics included the percentage of students in each score category (or response option for selected response items) for each assessment, as well as the difficulty of the item and the percent correct by gender. In reviewing these item statistics, the aim was to detect any unusual changes in item difficulties between administrations, which might indicate a problem in using the item to measure trends.

Exhibit 10.4: Example Item Statistics in 2019 and 2015 for a TIMSS 2019 Trend Item

Trends in International Mathematics and Science Study - TIMSS 2019 Assessment Results - 8th Grade
Trend Achievement Data Almanac for Mathematics Items (Weighted)

MP06_08 (MP62105): Algebra / Reasoning - 2 Points
Label: Area of rectangle with sides x and 2x + 1

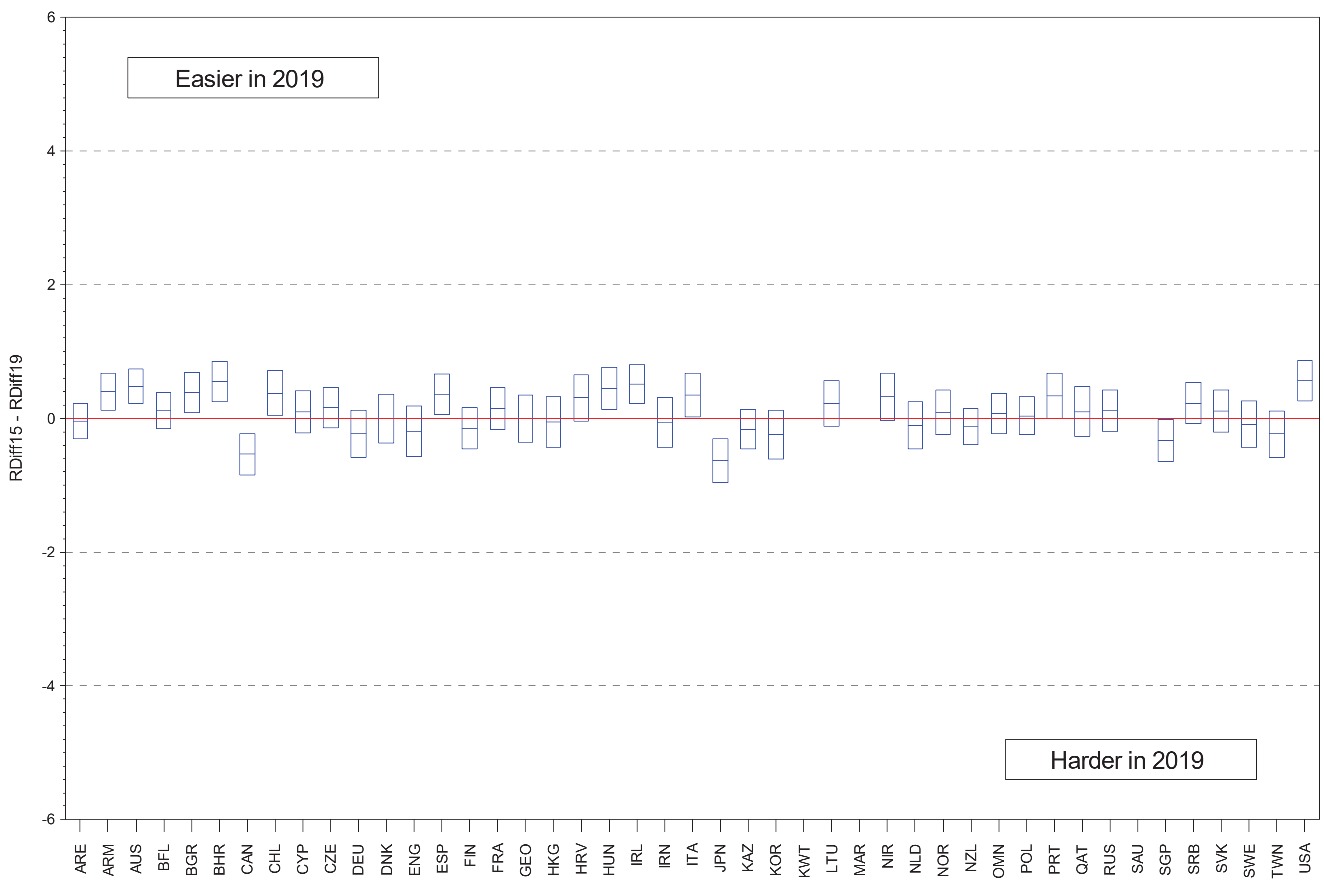
COUNTRY	Year	N	20 %	10 %	79 %	OMITTED %	NOT REACHED %	V1 %	V2 %	GIRL PCT RIGHT	BOY PCT RIGHT
Australia	2015	1476	23.5	1.4	60.1	14.6	0.5	24.8	23.5	23.9	23.0
	2019	1282	26.9	2.3	58.6	11.7	0.5	29.1	26.9	25.0	28.7
Bahrain	2015	710	8.7	5.8	72.2	13.0	0.3	14.5	8.7	8.4	8.9
	2019	814	26.0	3.0	44.2	26.8	0.0	29.0	26.0	22.9	28.8
Egypt	2015	1125	8.9	1.2	76.8	13.0	0.1	10.1	8.9	10.4	7.1
	2019	1036	12.9	2.9	76.3	7.7	0.2	15.8	12.9	12.7	13.1
Iran, Islamic Rep. of	2015	866	13.1	1.9	63.7	20.1	1.2	15.0	13.1	13.7	12.6
	2019	855	14.8	1.8	63.5	18.0	1.9	16.6	14.8	13.9	15.5
Ireland	2015	663	22.7	1.3	62.7	12.9	0.4	24.0	22.7	23.1	22.4
	2019	597	20.8	2.6	61.6	14.4	0.6	23.4	20.8	21.7	20.0
Japan	2015	672	43.4	1.8	46.6	8.1	0.1	45.2	43.4	42.5	44.3
	2019	639	41.4	0.8	49.7	8.1	0.0	42.2	41.4	35.2	47.5
Jordan	2015	1131	3.5	4.0	80.9	10.8	0.8	7.5	3.5	4.2	2.8
	2019	1010	7.7	5.2	78.9	7.5	0.7	12.9	7.7	8.2	7.3
Kuwait	2015	644	5.7	7.0	66.8	19.2	1.4	12.7	5.7	2.8	8.1
	2019	648	5.2	5.7	80.2	8.4	0.5	10.9	5.2	5.5	4.8
Lebanon	2015	547	4.6	4.2	67.6	22.4	1.1	8.8	4.6	5.4	3.5
	2019	671	2.7	3.3	69.9	22.2	1.8	6.0	2.7	2.7	2.7
Morocco	2015	1881	3.7	1.9	72.8	20.8	0.8	5.6	3.7	4.0	3.4
	2019	1213	5.1	0.3	72.0	21.9	0.7	5.4	5.1	3.0	7.4
New Zealand	2015	1122	19.3	1.1	64.0	15.0	0.7	20.3	19.3	19.7	18.8
	2019	862	17.6	2.7	65.4	13.2	1.1	20.3	17.6	20.1	15.4
Oman	2015	1267	6.3	6.5	79.6	6.4	1.3	12.7	6.3	6.2	6.4
	2019	952	7.1	4.3	80.2	7.0	1.3	11.4	7.1	9.5	4.8
Saudi Arabia	2015	538	0.7	4.9	81.7	11.8	0.9	5.6	0.7	0.4	1.1
	2019	814	5.7	4.9	83.0	5.8	0.6	10.6	5.7	7.1	4.3
South Africa (9)	2015	1788	3.3	5.9	85.3	4.7	0.8	9.3	3.3	4.4	2.2
	2019	2964	3.8	4.4	86.2	4.9	0.7	8.2	3.8	3.5	4.0
International Avg (n=14)	2015	14430	11.9	3.5	70.0	13.8	0.7	15.4	11.9	12.1	11.7
	2019	14357	14.1	3.2	69.3	12.7	0.8	17.3	14.1	13.6	14.6

V1 = Percent scoring 1 pt or better; V2 = Percent scoring 2 pts;
Percent right for boys and girls corresponds to percent obtaining full credit.
Because of missing gender information, some totals may appear inconsistent.

The TIMSS & PIRLS International Study Center used two different graphical displays to examine the differences in countries' Rasch item difficulties between 2019 and 2015. While some changes in item difficulties were anticipated, as countries' overall achievement may have improved or declined, items were noted if the difference between the Rasch difficulties across the two assessments for a particular country was greater than 2 logits.

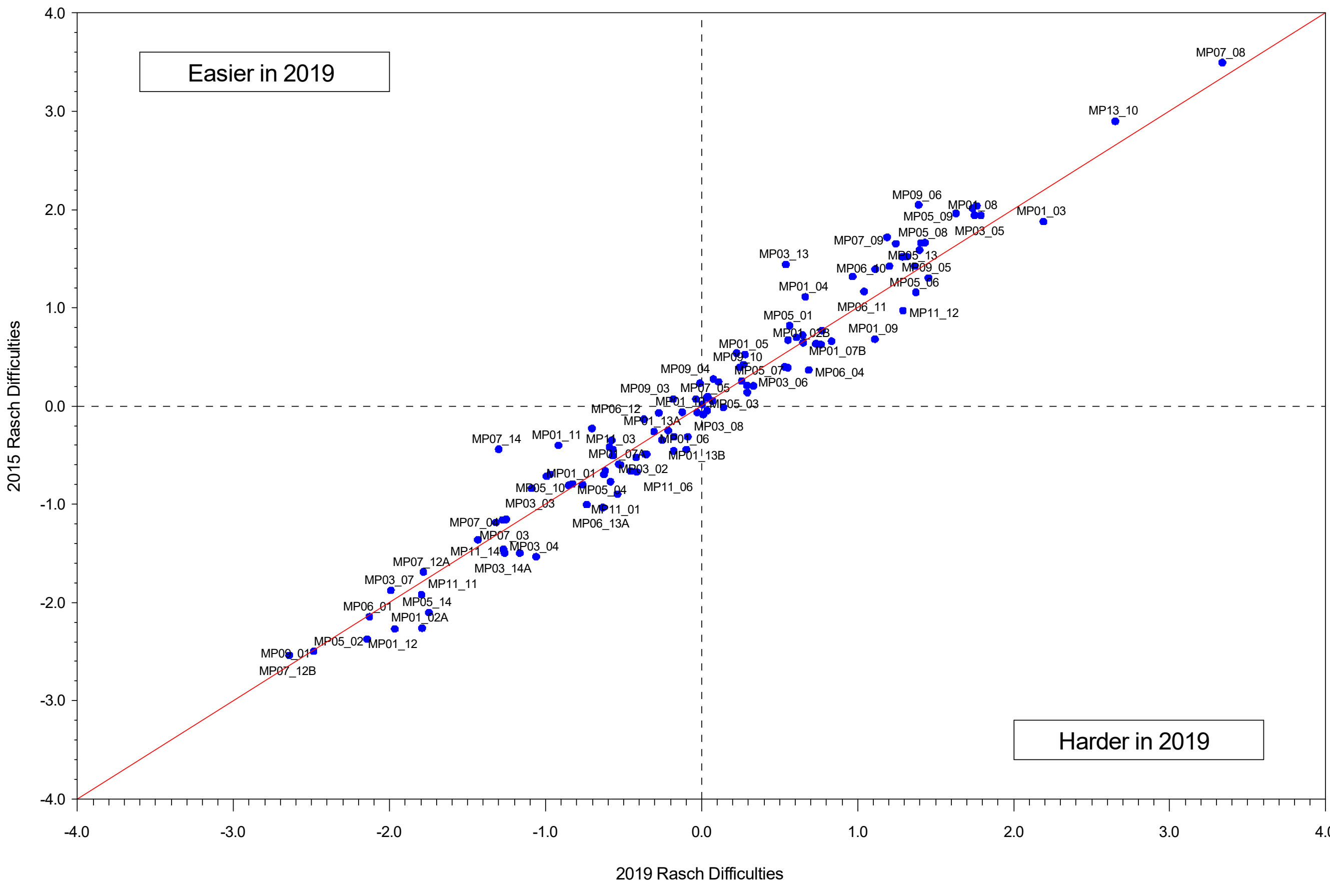
The first of these displays, shown for an example item in Exhibit 10.5, displays the difference in Rasch item difficulty of the item between 2019 and 2015 for each country. The difference in Rasch item difficulty for each country is displayed as a confidence interval, calculated using equations (10.1) and (10.2) but using each country's 2019 and 2015 Rasch difficulties and the standard error of their difference. A positive difference for a country indicates that the item was relatively easier in 2019, and a negative difference indicates that the item was relatively more difficult.

Exhibit 10.5: Example Plot of Differences in Rasch Item Difficulties Between 2019 and 2015 for a TIMSS 2019 Trend Item



The second graphical display, presented in Exhibit 10.6, shows the performance of a given country on all trend items simultaneously. For each country, the graph plots the 2019 Rasch difficulty of every trend item against its Rasch difficulty in 2015. When there were no differences between the difficulties in the two successive administrations, the data points aligned on or near the diagonal. Large deviations from the diagonal were noted for further investigation.

Exhibit 10.6: Example Plot of Rasch Trend Item Difficulties Across TIMSS 2019 and 2015 by Country



Item Position Effects

As described in the TIMSS 2019 Assessment Design, assessment items for each grade and subject are arranged in 14 groups called "item blocks" (for paperTIMSS) or "item block combinations" (for eTIMSS), which were assembled into achievement booklets. Each item block appears in two booklets, with each item block appearing in the first half of one booklet and the second half of another. This counterbalancing helps to control for the impact of item position on the item statistics.

To examine the magnitude of item position effects on item percent correct and the percent of omitted and not-reached responses, block-level item statistics weighted by maximum score points were computed for each of the two positions that each block appears in the booklet design—either position 1 and position 4 or position 2 and position 3. The results are reported in Appendix 10D for each assessment averaged across countries, as well as for each country across item blocks. A summary of results with the average differences in item statistics between the booklet positions is provided in Exhibits 10.7 and 10.8 for the fourth and eighth grade, respectively.

Exhibit 10.7: Summary of International Average Item Block Statistics by Booklet Position (Weighted)—Grade 4

	Average Percent Correct Across Items			Average Percent Omitted Responses Across Items			Average Percent Not Reached Across Items		
	Positions 1&3	Positions 2&4	Difference	Positions 1&3	Positions 2&4	Difference	Positions 1&3	Positions 2&4	Difference
Mathematics									
eTIMSS	49.9	48.3	–1.6	3.3	4.5	1.1	0.1	2.5	2.5
paperTIMSS	49.4	48.1	–1.3	5.8	6.3	0.5	0.2	3.3	3.1
Less Difficult	43.4	42.2	–1.2	7.9	8.3	0.4	0.3	3.7	3.4
Science									
eTIMSS	53.3	52.0	–1.3	3.2	4.4	1.2	0.1	2.5	2.5
paperTIMSS	50.9	49.5	–1.3	5.7	6.7	1.0	0.1	2.5	2.4
Less Difficult	33.9	32.6	–1.3	10.9	12.2	1.3	0.3	4.9	4.6

Exhibit 10.8: Summary of International Average Item Block Statistics by Booklet Position (Weighted)—Grade 8

	Average Percent Correct Across Items			Average Percent Omitted Responses Across Items			Average Percent Not Reached Across Items		
	Positions 1&3	Positions 2&4	Difference	Positions 1&3	Positions 2&4	Difference	Positions 1&3	Positions 2&4	Difference
Mathematics									
eTIMSS	44.0	41.9	–2.1	5.5	7.0	1.5	0.1	1.8	1.7
paperTIMSS	34.9	33.2	–1.6	7.9	8.5	0.7	0.1	2.0	1.9
Science									
eTIMSS	47.8	46.3	–1.5	4.0	5.0	1.1	0.1	0.8	0.7
paperTIMSS	39.6	38.1	–1.5	6.7	7.5	0.8	0.1	1.6	1.6

The results indicate minimal impact of block position on the TIMSS 2019 item statistics. On average, item blocks appearing in the second half of a booklet part (positions 2 and 4) were slightly more difficult and had slightly more missing responses than item blocks appearing the first half of a booklet part (positions 1 and 3). However, the differences were small. Across countries, differences in average item percent correct between positions 2/4 and positions 1/3 ranged from −1.2 at the fourth grade for the less difficult mathematics items to −2.1 in eighth grade eTIMSS mathematics. Differences in average percent omitted ranged between 0.4 in fourth grade less difficult mathematics to 1.5 in eighth grade eTIMSS mathematics; and differences in average percent not reached ranged between 0.7 in eighth grade eTIMSS science and 4.2 in fourth grade science for less difficult countries.

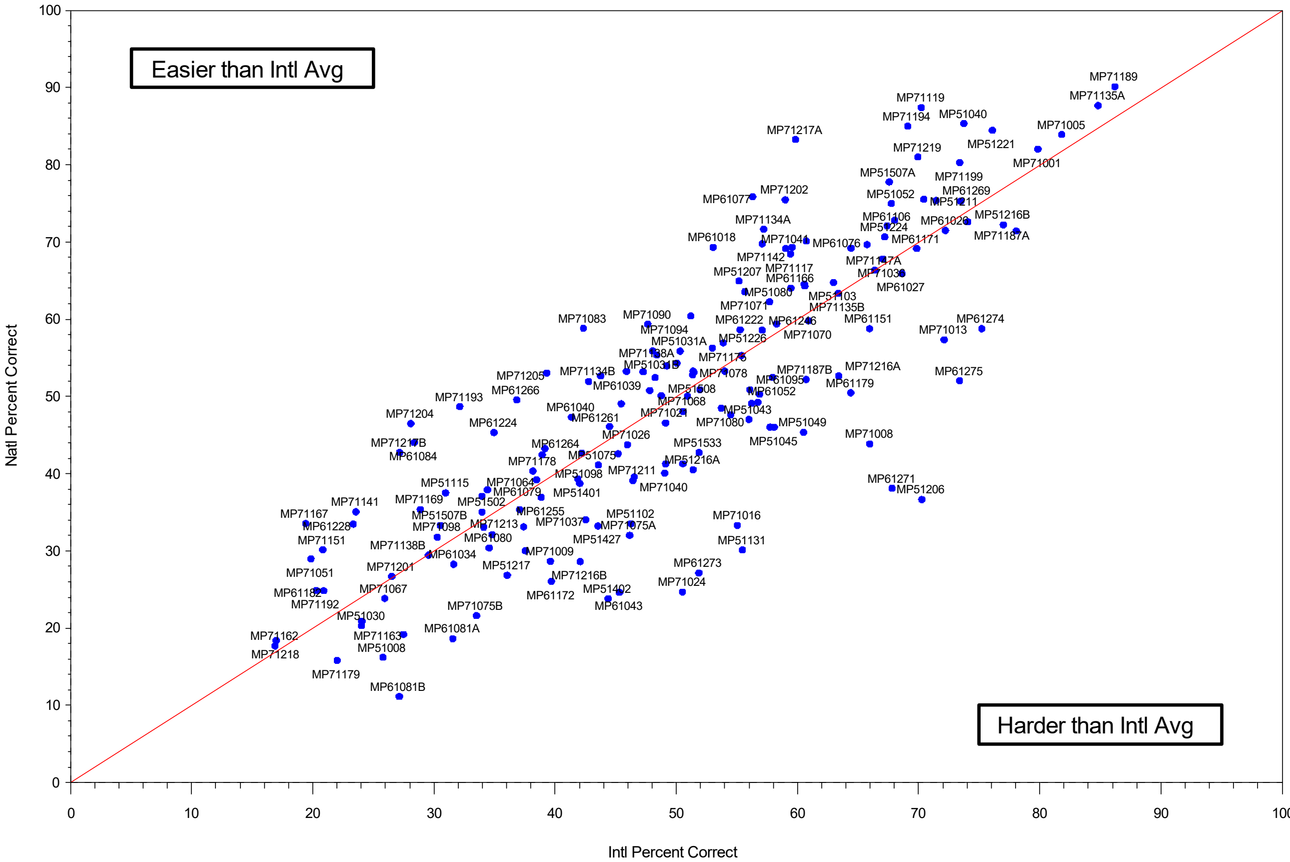
Detecting Anomalies in the TIMSS 2019 Achievement Data

To ensure that each participating country and benchmarking entity had data adhering to TIMSS' quality standards, the TIMSS & PIRLS International Study Center conducted analyses of item statistics at the country level. Several graphical displays were produced for each TIMSS participant for item percent correct, item point-biserials, and percent omitted responses. The graphs were analyzed to detect any anomalous patterns in any particular country's data relative to the international average or to their previous TIMSS performance. Anomalous patterns may be indicative that systematic errors occur in a country's data, which may be due to errors in collecting and processing the data. For any anomalous patterns detected in the item statistics for a particular country, the National Research Coordinator was contacted to discuss how best to address any issues.

The first set of graphical displays compared each country's item performance to the international average for all items simultaneously, where item performance is defined in terms of item percent correct, item discrimination (point-biserial correlation), and item percent omitted. An example is shown in Exhibit 10.9 for item percent correct. For each country, the graph plots the 2019 item percent correct of all items against the 2019 international average. Typical patterns show data points along the range of the *x*- and *y* axis, with random deviations from the diagonal. There may be more points above the diagonal for higher performing countries and more points below for lower performing countries, but otherwise the points should align closely with the diagonal. The best-fit line should be approximately linear and parallel with the diagonal. Any patterns largely deviating from this were noted for further investigation. Plots comparing national and international item discrimination (point-biserial correlation) and percent omitted should have similar patterns, but points more tightly clustered together since there is a smaller range.

These plots of national versus international item statistics were also compared against the same plots produced in TIMSS 2015. If the patterns for both assessments were unusually different, it may indicate a problem in the 2019 data. The plots may also be examined separately for selected response and constructed response formatted items to ensure similar patterns. The relationship between national and international statistics for both item types should also match that from TIMSS 2015.

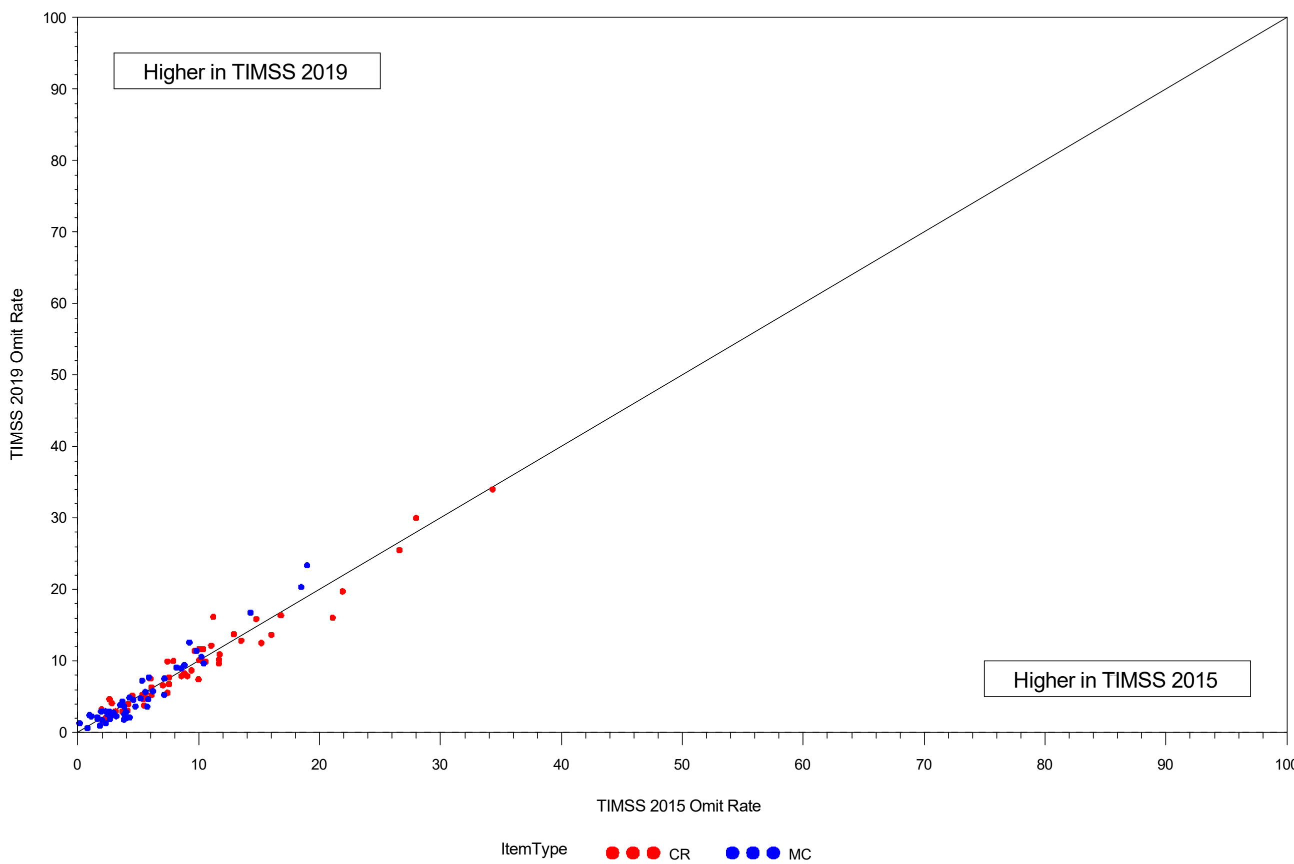
Exhibit 10.9: Example Plot of Item Percent Correct Across National and International by Country



The second set of graphical displays compares each country's TIMSS 2019 trend item performance with their TIMSS 2015 item performance for all items simultaneously, where item performance is defined in terms of percent correct, item discrimination (point-biserial correlation), and item percent omitted. An example is shown in Exhibit 10.10 for item percent omitted, displaying a typical pattern. For each country, the graphs plots the 2019 item percent omitted of every trend item against its item percent omitted in 2015, with points colored according to item type. When there were no differences between the difficulties in the two successive administrations, the data points aligned on or near the diagonal from the graph origin. While some changes were anticipated, as countries' overall achievement may have improved or declined, unusually large deviations from the diagonal were noted for further investigation. For all statistics plotted, comparisons should show similar patterns for both selected response and constructed response item types, and any differences should not relate to the difficulty of the item.

An additional set of plots were produced comparing each country's TIMSS 2019 item performance with their item performance from the field test conducted one year earlier. These plots were similar to the example in Exhibit 10.10 below comparing 2019 and 2015 performance, with the expectation of smaller differences. Large differences in item performance compared to the field test would be considered a implausible change in performance, warranting further review.

Exhibit 10.10: Example Plot of Item Percent Omitted Across TIMSS 2019 and 2015 by Country



Item Review Outcomes

Using all the information from the comprehensive collection of item analyses and reliability data that were computed and summarized for TIMSS 2019, the TIMSS & PIRLS International Study Center thoroughly reviewed all item statistics for every participating country and benchmarking participant to ensure that the items were performing comparably across countries and modes. In particular, items with the following problems were considered for possible deletion from the international database:

- An error was detected for a particular country during translation verification but was not corrected before test administration
- Data checking revealed a selected response item with more or fewer options than in the international version for a particular country
- The item analysis showed the item to have a negative biserial, or, for an item with more than 1 score point, point-biserials that did not increase with each score level
- For selected response items, the item review revealed a faulty distracter influencing the item statistics for all countries
- The item-by-country interaction results showed a very large negative interaction for a particular country
- For constructed response items, the within-country scoring reliability data showed an agreement of less than 75 percent
- For trend items, an item performed substantially differently in 2019 compared to the TIMSS 2015 administration, or an item was not included in the previous assessment for a particular country
- For eTIMSS trend items, a substantially larger than average difference in item difficulty or percent omitted between eTIMSS and bridge for a particular country.

When the item statistics indicated a problem with an item, the documentation from the translation verification was used as an aid in checking the test booklets. If a question remained about potential translation or cultural issues, however, then the National Research Coordinator was consulted before deciding how the item should be treated.

The checking of the TIMSS 2019 achievement data involved review of more than 1,800 items and resulted in the detection of very few items that were inappropriate for international comparisons. Among the few items singled out in the review process were mostly items with differences attributable to either translation or printing problems. A small number of items were identified as having severe differential item functioning after item review during IRT scaling. Diagnostic score codes for some constructed response items may be recoded if the point-biserials did not behave. Decisions about deleting items for all countries were most often implemented for both eTIMSS and paperTIMSS versions, with a few exceptions.

Appendix 10E includes a list of deleted items, as well as a list of recodes made to constructed response items. There also were a number of items in each study that were combined, or derived, for scoring purposes. See Appendix 10F for details about how score points were awarded for each derived item.

Review of Item Statistics Between eTIMSS and paperTIMSS

To establish a link between eTIMSS and paperTIMSS, eTIMSS countries that also participated in TIMSS 2015 administered paper booklets of trend items to randomly equivalent "bridge" samples of students. To strengthen the link, an important review step for TIMSS 2019 included checking the extent that items had similar statistical properties between eTIMSS and paperTIMSS (e.g., an item that was relatively easy on paper should also be easy in digital format).

The review focused on eTIMSS item percent correct statistics for trend items administered in digital format to the regular sample of students compared to those in paper format for the bridge samples (see Exhibit 10.11 for an example of this type of item almanac). For each eTIMSS country, mode difference item statistics included the percentage of students in each score category (or response option for selected response items) for each assessment, as well as the difficulty of the item and the percent correct by gender. In reviewing these item statistics, the aim was to identify items that were likely to be found invariant under IRT, as well as to detect any unusual differences in item difficulties between modes for a particular country that might indicate a problem. Further item equivalence analyses performed during scaling are described in Chapter 12 of this volume.

Exhibit 10.11: Example Item Statistics in eTIMSS and paperTIMSS (Bridge) for a TIMSS 2019 Trend Item

Trends in International Mathematics and Science Study - TIMSS 2019 Assessment Results - 8th Grade
Mode Differences Data Almanac for Science Items (Weighted)

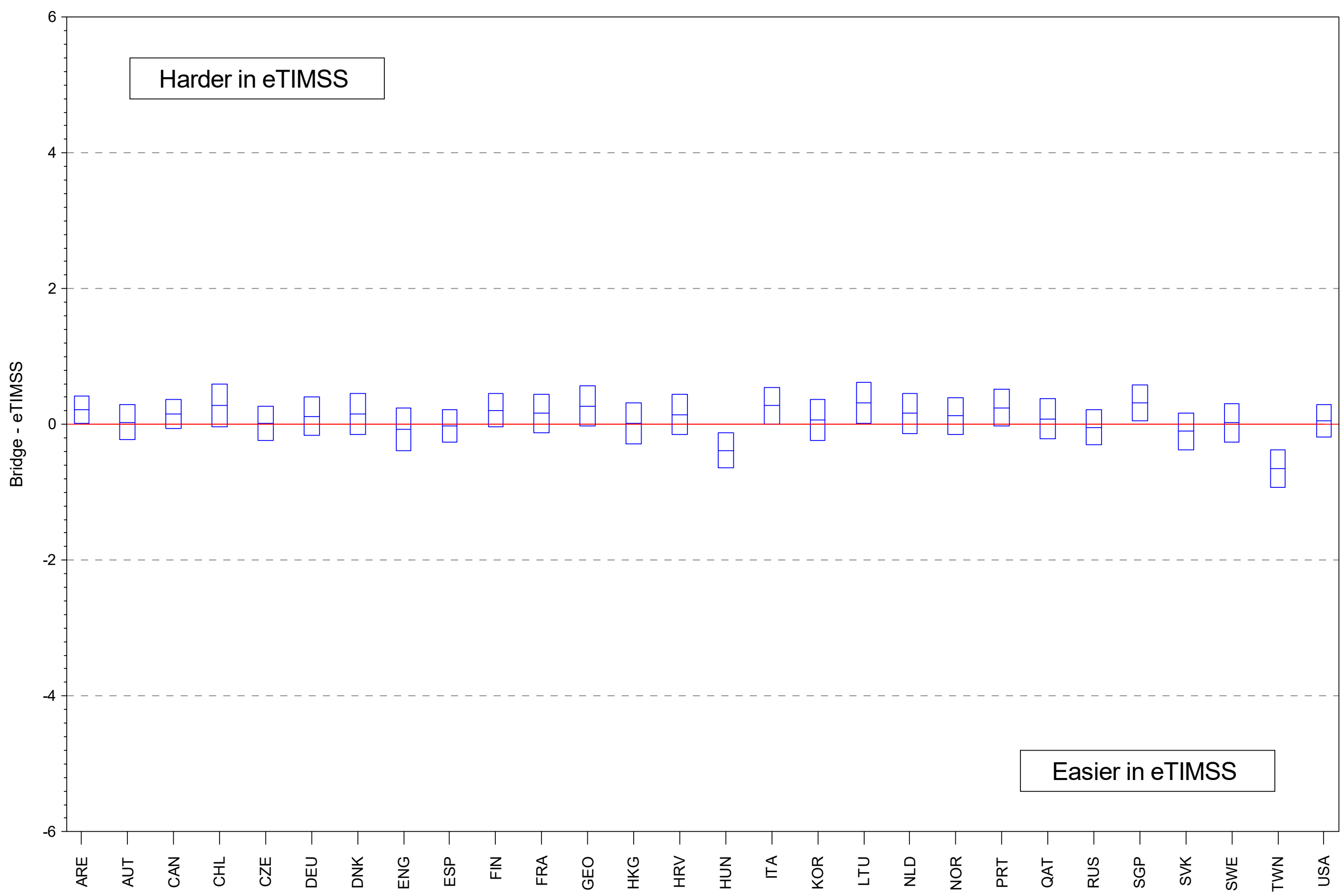
SE06_03 (SE62284): Biology / Applying - 1 Point - Key: B
Label: Hair color of young rabbits

COUNTRY	MODE	N	DIFF %	A %	B %	C %	D %	OMITTED %	NOT REACHED %	GIRL PCT RIGHT	BOY PCT RIGHT
Chile	eTIMSS	588	56.5	14.4	56.5	5.0	22.3	1.6	0.1	62.4	50.6
	pTIMSS	381	58.7	13.6	58.7	4.2	22.0	1.2	0.4	63.5	52.4
Chinese Taipei	eTIMSS	700	81.4	6.2	81.4	1.9	10.4	0.2	0.0	82.8	79.9
	pTIMSS	401	79.1	8.2	79.1	1.2	11.2	0.2	0.0	79.4	78.9
England	eTIMSS	492	60.7	11.6	60.7	5.4	21.8	0.3	0.2	65.9	54.8
	pTIMSS	408	58.1	12.2	58.1	3.0	25.1	1.3	0.2	67.6	50.6
Georgia	eTIMSS	472	46.9	21.0	46.9	7.9	22.5	1.4	0.3	52.5	42.0
	pTIMSS	324	49.9	14.4	49.9	9.3	25.7	0.8	0.0	60.3	40.2
Hong Kong SAR	eTIMSS	469	48.3	12.2	48.3	12.9	25.4	1.0	0.2	52.0	45.2
	pTIMSS	358	54.4	7.6	54.4	3.7	32.9	0.8	0.7	58.6	51.0
Hungary	eTIMSS	652	66.9	18.1	66.9	2.6	11.9	0.4	0.0	71.4	62.7
	pTIMSS	439	69.7	13.4	69.7	2.1	13.5	1.3	0.0	74.2	65.0
Israel	eTIMSS	533	56.9	11.9	56.9	4.3	25.7	1.2	0.1	63.9	49.0
	pTIMSS	452	55.4	9.1	55.4	3.7	29.8	0.9	1.0	59.6	51.1
Italy	eTIMSS	529	71.4	15.4	71.4	3.3	8.7	1.2	0.0	72.2	70.6
	pTIMSS	517	70.0	11.9	70.0	4.3	12.8	0.5	0.6	79.8	59.4
Korea, Rep. of	eTIMSS	553	55.0	6.6	55.0	3.4	35.0	0.0	0.0	61.3	49.7
	pTIMSS	424	51.4	6.5	51.4	2.6	39.2	0.3	0.0	59.4	46.2
Lithuania	eTIMSS	546	64.8	14.3	64.8	2.0	18.7	0.2	0.0	70.9	57.4
	pTIMSS	415	62.8	13.2	62.8	1.3	22.1	0.5	0.1	68.1	57.3
Malaysia	eTIMSS	1018	43.4	13.7	43.4	4.9	37.8	0.2	0.0	46.7	40.0
	pTIMSS	390	43.2	15.4	43.2	5.7	35.1	0.6	0.0	46.7	39.9
Norway (9)	eTIMSS	627	66.9	5.6	66.9	3.3	22.7	1.0	0.6	76.9	57.7
	pTIMSS	499	61.6	8.0	61.6	3.0	24.7	2.4	0.2	69.9	53.6
Qatar	eTIMSS	546	45.7	15.8	45.7	8.6	28.7	0.9	0.3	51.2	40.8
	pTIMSS	378	47.2	17.4	47.2	6.4	28.5	0.6	0.0	49.2	45.2
Russian Federation	eTIMSS	555	60.6	12.6	60.6	4.2	21.5	1.0	0.2	69.0	52.8
	pTIMSS	520	69.6	11.8	69.6	1.6	16.1	0.9	0.0	76.3	62.9
Singapore	eTIMSS	689	66.7	6.7	66.7	2.7	23.7	0.0	0.1	67.7	65.8
	pTIMSS	469	65.0	4.4	65.0	3.8	26.0	0.6	0.2	68.1	62.1
Sweden	eTIMSS	548	66.6	7.3	66.6	4.8	20.2	0.8	0.4	73.4	60.4
	pTIMSS	396	67.8	8.8	67.8	1.4	19.8	1.4	0.7	76.9	60.4
Turkey	eTIMSS	582	57.6	11.5	57.6	2.4	28.5	0.0	0.0	60.6	54.3
	pTIMSS	455	57.8	10.1	57.8	3.5	28.2	0.3	0.1	63.2	52.7
United Arab Emirates	eTIMSS	3186	49.9	15.4	49.9	11.3	22.3	1.0	0.0	53.4	46.6
	pTIMSS	526	50.4	12.9	50.4	7.3	28.0	1.4	0.1	47.9	53.0
United States	eTIMSS	1249	69.1	8.4	69.1	4.4	17.4	0.6	0.1	72.6	65.5
	pTIMSS	370	66.7	11.6	66.7	5.0	15.6	0.6	0.6	75.6	59.9
International Avg (n=19)	eTIMSS	14534	59.8	12.0	59.8	5.0	22.4	0.7	0.1	64.6	55.0
	pTIMSS	8122	59.9	11.1	59.9	3.8	24.0	0.9	0.3	65.5	54.8

DIFF = Percent correct
Because of missing gender information, some totals may appear inconsistent.

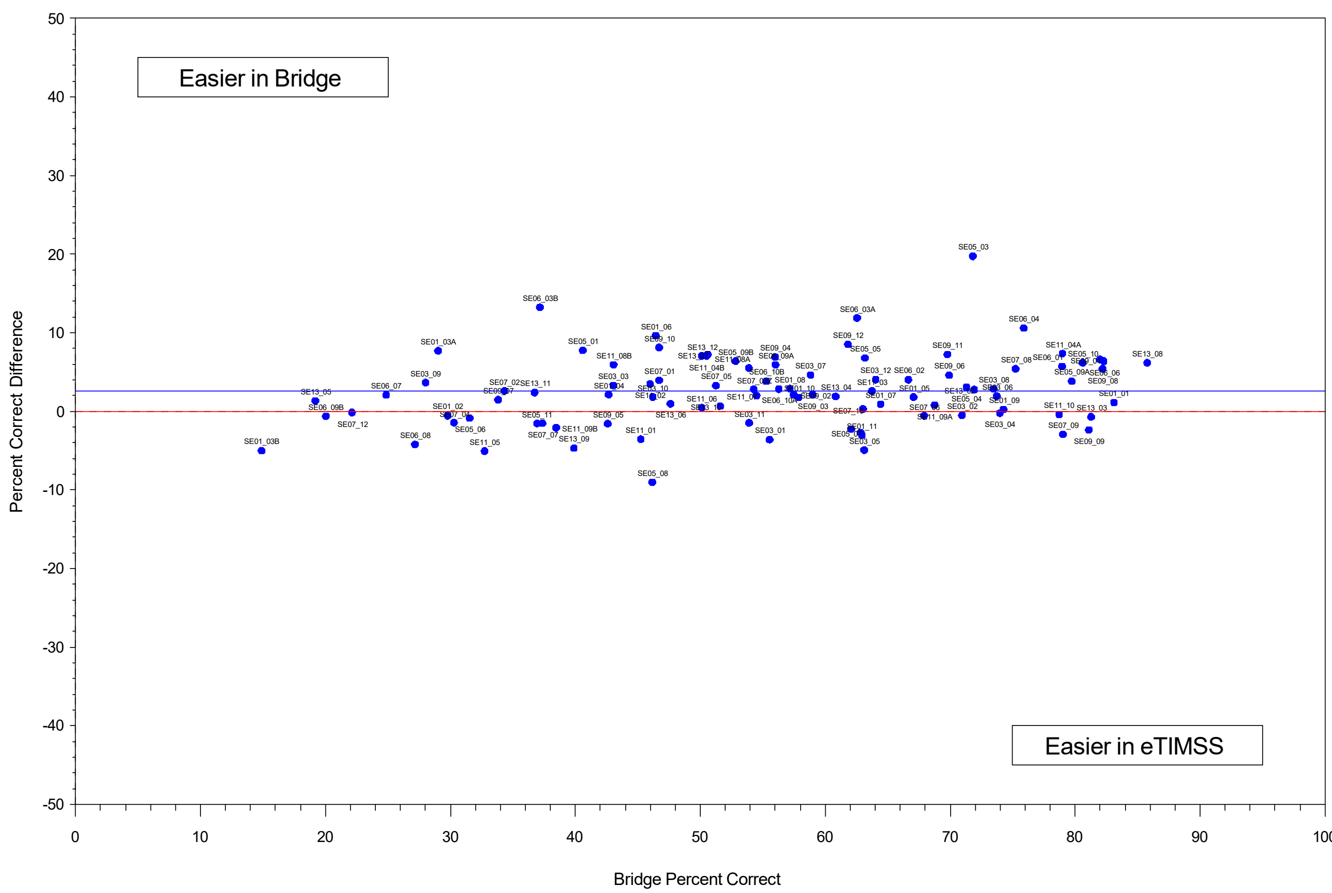
Two different graphical displays were produced for item review to assess the differences in item difficulty by mode of administration. Exhibit 10.12 shows the first of these for an example item. For each country, the difference in item percent correct between eTIMSS and bridge is displayed as a confidence interval. This was calculated using equations (10.1) and (10.2), but using each country's 2019 bridge and eTIMSS percent correct values and the standard error of their difference. A positive difference for a country indicates that the item was relatively harder in eTIMSS, and a negative difference indicates that the item was relatively easier.

Exhibit 10.12: Example Plot of Differences in Item Percent Correct Between eTIMSS and paperTIMSS (Bridge) for a TIMSS 2019 Trend Item



The second graphical display, presented in Exhibit 10.13, shows the mode differences in percent correct for a given country on all items simultaneously. The blue horizontal line represents the country's average difference across all the items. Where there were no differences between the percent correct in the two modes, the data points aligned on or near the horizontal axis. A positive difference for an item indicates that it was relatively easier on paper, and a negative difference indicates that the item was relatively easier in eTIMSS. Any large or systematics deviations were flagged for further review.

Exhibit 10.13: Example Plot of Differences in Trend Item Percent Correct Between eTIMSS and paperTIMSS (Bridge) by Country



Appendix 10A: TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items

TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 4 Mathematics

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Albania	100	98	100	100	97	100
Armenia	100	100	100	100	99	100
Australia	99	95	100	99	94	100
Austria	98	95	100	98	95	100
Azerbaijan	100	98	100	100	98	100
Bahrain	92	78	100	91	75	99
Belgium (Flemish)	97	86	100	97	82	100
Bosnia and Herzegovina	99	91	100	98	91	100
Bulgaria	99	97	100	99	97	100
Canada	97	93	100	96	93	100
Chile	98	94	100	98	94	100
Chinese Taipei	97	89	100	97	89	100
Croatia	99	97	100	99	97	100
Cyprus	98	89	100	97	88	100
Czech Republic	98	94	100	97	92	100
Denmark	97	90	100	96	89	100
England	98	92	100	98	92	100
Finland	100	98	100	100	98	100
France	98	94	100	98	94	100
Georgia	96	89	100	96	88	100
Germany	98	93	100	97	93	100
Hong Kong SAR	100	100	100	100	100	100
Hungary	98	91	99	97	91	99
Iran, Islamic Rep. of	99	94	100	97	92	100
Ireland	100	97	100	99	97	100
Italy	99	94	100	98	94	100
Japan	99	87	100	98	87	100

TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 4 Mathematics (continued)

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Kazakhstan	92	81	99	89	75	97
Korea, Rep. of	99	96	100	99	96	100
Kosovo	96	70	100	94	70	100
Kuwait	100	98	100	99	96	100
Latvia	99	92	100	98	91	100
Lithuania	98	92	100	98	92	100
Malta	95	89	99	94	87	99
Montenegro	98	92	100	97	92	100
Morocco	95	72	100	94	72	100
Netherlands	96	89	100	95	89	100
New Zealand	99	92	100	98	90	100
North Macedonia	99	96	100	99	91	100
Northern Ireland	100	97	100	100	95	100
Norway (5)	98	92	100	97	92	100
Oman	98	92	100	96	88	100
Pakistan	100	100	100	100	100	100
Philippines	99	97	100	99	93	100
Poland	99	93	100	98	88	100
Portugal	98	95	100	97	94	99
Qatar	98	95	100	97	94	100
Russian Federation	98	91	100	97	91	100
Saudi Arabia	96	72	100	94	69	99
Serbia	98	95	100	98	93	100
Singapore	98	95	100	98	95	100
Slovak Republic	99	95	100	98	95	100
South Africa (5)	98	83	100	97	82	100
Spain	97	92	100	96	92	100
Sweden	98	91	100	98	91	100
Turkey (5)	99	96	100	99	96	100
United Arab Emirates	99	96	100	98	96	100
United States	98	96	100	98	96	100
International Average	**98**	**92**	**100**	**97**	**91**	**100**

TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 4 Mathematics (continued)

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Ontario, Canada	97	88	100	96	86	100
Quebec, Canada	96	89	100	95	87	100
Moscow City, Russian Fed.	99	95	100	98	95	100
Madrid, Spain	97	92	100	97	92	100
Abu Dhabi, UAE	99	94	100	99	94	100
Dubai, UAE	98	92	100	98	92	100

TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 4 Science

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Albania	100	99	100	99	98	100
Armenia	100	100	100	100	100	100
Australia	97	90	100	97	90	100
Austria	95	85	100	95	85	100
Azerbaijan	100	98	100	100	98	100
Bahrain	98	93	100	97	92	100
Belgium (Flemish)	93	74	99	92	74	99
Bosnia and Herzegovina	99	94	100	98	90	100
Bulgaria	98	94	100	97	92	100
Canada	92	80	100	92	80	100
Chile	95	90	100	94	90	100
Chinese Taipei	95	83	100	95	81	100
Croatia	94	87	100	94	87	100
Cyprus	93	85	100	92	85	100
Czech Republic	93	84	100	93	84	100
Denmark	93	84	100	93	84	100
England	94	90	100	93	90	100
Finland	97	91	100	96	91	100
France	95	90	99	95	90	99
Georgia	92	76	100	91	76	99
Germany	96	85	100	95	85	100
Hong Kong SAR	100	100	100	100	100	100
Hungary	92	76	99	91	71	99
Iran, Islamic Rep. of	96	84	100	94	75	100
Ireland	98	91	100	98	91	100
Italy	97	94	100	97	93	100
Japan	95	85	100	94	85	100
Kazakhstan	89	79	97	86	75	95
Korea, Rep. of	98	93	100	98	93	100
Kosovo	89	75	97	85	60	96
Kuwait	100	99	100	99	97	100

TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 4 Science (continued)

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Latvia	94	81	100	93	79	100
Lithuania	95	87	99	94	87	99
Malta	92	85	99	91	85	99
Montenegro	98	90	100	98	90	100
Morocco	91	74	99	88	65	99
Netherlands	92	81	100	91	81	100
New Zealand	96	83	100	95	82	100
North Macedonia	98	94	100	98	90	100
Northern Ireland	96	90	100	95	90	100
Norway (5)	93	84	100	92	84	100
Oman	97	93	100	95	87	99
Pakistan	100	100	100	100	100	100
Philippines	99	96	100	98	93	100
Poland	94	72	100	93	71	100
Portugal	96	90	100	95	89	100
Qatar	96	91	100	95	91	100
Russian Federation	94	88	100	94	88	100
Saudi Arabia	94	73	100	92	55	100
Serbia	97	92	100	96	91	100
Singapore	96	86	100	96	86	100
Slovak Republic	97	93	100	97	93	100
South Africa (5)	98	91	100	97	90	100
Spain	93	86	100	92	85	100
Sweden	92	80	100	92	80	100
Turkey (5)	96	90	100	96	85	100
United Arab Emirates	94	90	99	94	90	99
United States	96	87	100	95	87	100
International Average	**95**	**87**	**100**	**95**	**86**	**100**

TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 4 Science (continued)

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Ontario, Canada	92	77	100	91	77	100
Quebec, Canada	91	74	100	91	74	100
Moscow City, Russian Fed.	96	86	100	95	86	100
Madrid, Spain	92	82	100	92	81	100
Abu Dhabi, UAE	95	87	100	94	86	100
Dubai, UAE	92	81	100	92	81	100

TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 8 Mathematics

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Australia	98	85	100	97	85	100
Bahrain	90	61	100	89	61	100
Chile	96	90	100	95	90	100
Chinese Taipei	97	87	100	97	87	100
Cyprus	98	89	100	97	88	100
Egypt	99	95	100	98	92	100
England	97	93	100	97	93	100
Finland	99	97	100	99	97	100
France	97	88	100	97	88	100
Georgia	96	89	100	95	87	100
Hong Kong SAR	100	100	100	100	100	100
Hungary	97	91	100	96	90	100
Iran, Islamic Rep. of	99	89	100	97	89	100
Ireland	99	91	100	98	86	100
Israel	98	90	100	97	90	100
Italy	98	91	100	97	91	100
Japan	99	90	100	98	90	100
Jordan	99	96	100	98	93	100
Kazakhstan	93	78	100	90	69	99
Korea, Rep. of	99	96	100	99	95	100
Kuwait	100	97	100	99	96	100
Lebanon	100	100	100	100	100	100
Lithuania	97	89	100	96	89	100
Malaysia	99	97	100	99	97	100
Morocco	96	68	100	90	30	100
New Zealand	98	87	100	97	85	100
Norway (9)	97	92	100	97	92	100
Oman	99	95	100	98	89	100
Portugal	97	89	100	97	89	100
Qatar	97	93	100	96	93	100
Romania	99	95	100	97	92	100

TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 8 Mathematics (continued)

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Russian Federation	98	90	100	97	90	100
Saudi Arabia	98	76	100	95	29	100
Singapore	99	95	100	98	93	100
South Africa (9)	100	98	100	99	95	100
Sweden	97	85	100	97	85	100
Turkey	99	94	100	98	93	100
United Arab Emirates	97	93	100	97	93	100
United States	98	91	100	98	91	100
International Average	**98**	**90**	**100**	**97**	**87**	**100**
Ontario, Canada	96	83	100	95	83	100
Quebec, Canada	96	84	100	95	80	100
Moscow City, Russian Fed.	98	92	100	98	92	100
Gauteng, RSA (9)	100	95	100	99	91	100
Western Cape, RSA (9)	100	93	100	99	91	100
Abu Dhabi, UAE	97	91	100	97	91	100
Dubai, UAE	97	87	100	96	87	100

TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 8 Science

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Australia	94	76	100	92	76	100
Bahrain	98	89	100	97	83	100
Chile	95	90	100	94	90	100
Chinese Taipei	95	81	100	94	81	100
Cyprus	94	86	100	93	86	100
Egypt	97	89	100	96	86	100
England	95	90	100	94	90	100
Finland	97	91	100	96	91	100
France	96	86	100	96	86	100
Georgia	93	73	100	92	73	100
Hong Kong SAR	100	100	100	100	100	100
Hungary	91	79	99	90	78	99
Iran, Islamic Rep. of	96	85	100	94	83	100
Ireland	96	83	100	95	82	100
Israel	93	81	100	92	81	100
Italy	97	92	100	97	92	100
Japan	95	86	100	95	86	100
Jordan	99	96	100	98	85	100
Kazakhstan	88	72	99	85	67	98
Korea, Rep. of	97	90	100	96	90	100
Kuwait	99	98	100	99	97	100
Lebanon	100	100	100	100	100	100
Lithuania	94	80	100	93	80	100
Malaysia	95	90	100	95	90	100
Morocco	91	70	100	87	38	99
New Zealand	95	78	99	94	77	99
Norway (9)	94	85	100	94	85	100
Oman	97	89	100	96	89	100
Portugal	95	87	100	95	87	100
Qatar	95	90	100	94	90	100
Romania	96	81	100	95	80	100

TIMSS 2019 Within-Country Scoring Reliability for Human Scored Items—Grade 8 Science (continued)

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Russian Federation	94	87	100	94	87	100
Saudi Arabia	92	75	100	90	45	100
Singapore	96	87	100	95	87	100
South Africa (9)	99	96	100	99	96	100
Sweden	93	80	100	93	80	100
Turkey	96	88	100	96	88	100
United Arab Emirates	94	90	100	94	90	100
United States	95	90	100	95	89	100
International Average	**95**	**86**	**100**	**94**	**84**	**100**
Ontario, Canada	93	82	100	92	82	100
Quebec, Canada	92	78	100	90	78	100
Moscow City, Russian Fed.	97	87	100	96	87	100
Gauteng, RSA (9)	98	84	100	98	84	100
Western Cape, RSA (9)	99	94	100	99	94	100
Abu Dhabi, UAE	95	83	100	94	83	100
Dubai, UAE	93	84	100	93	83	100

Appendix 10B: TIMSS 2019 Trend Scoring Reliability for Human Scored Items

TIMSS 2019 Trend Scoring Reliability for Human Scored Items—Grade 4 Mathematics

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Australia	98	90	100	97	89	100
Bahrain	99	96	100	98	91	100
Belgium (Flemish)	96	87	99	96	86	99
Bulgaria	99	95	100	99	95	100
Canada	95	76	99	94	75	99
Chile	96	85	100	96	85	100
Chinese Taipei	98	93	100	97	92	100
Croatia	98	90	100	97	90	100
Cyprus	97	87	100	97	87	100
Czech Republic	97	89	100	97	89	100
Denmark	96	84	100	94	78	100
England	98	86	100	97	86	100
Finland	99	93	100	98	93	100
France	97	87	100	96	87	100
Georgia	95	67	100	95	67	100
Germany	98	88	100	97	88	100
Hong Kong SAR	96	81	100	96	81	100
Hungary	97	93	99	97	92	99
Iran, Islamic Rep. of	98	92	99	97	92	99
Ireland	98	85	100	97	85	100
Italy	97	87	100	97	86	100
Japan	97	89	100	97	89	100
Korea, Rep. of	99	94	100	99	94	100
Lithuania	98	89	100	97	84	100
Netherlands	98	84	100	97	84	100
New Zealand	97	86	100	96	86	100
Northern Ireland	98	87	100	98	87	100

TIMSS 2019 Trend Scoring Reliability for Human Scored Items—Grade 4 Mathematics (continued)

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Norway (5)	98	91	100	97	90	100
Oman	95	87	100	93	72	100
Poland	97	89	100	97	89	100
Portugal	98	88	100	98	87	100
Qatar	97	89	100	95	77	100
Saudi Arabia	93	85	98	93	84	98
Serbia	96	79	100	95	79	100
Singapore	99	95	100	99	94	100
Slovak Republic	97	82	100	96	81	100
South Africa (5)	97	94	99	97	93	99
Spain	96	87	100	96	86	100
Sweden	96	78	100	96	77	100
United Arab Emirates	97	89	100	96	86	100
United States	97	85	100	97	85	100
International Average	**97**	**87**	**100**	**96**	**86**	**100**

TIMSS 2019 Trend Scoring Reliability for Human Scored Items—Grade 4 Science

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Australia	93	85	99	93	85	99
Bahrain	91	82	98	89	77	98
Belgium (Flemish)	92	82	99	91	82	99
Bulgaria	96	85	100	95	85	100
Canada	91	79	99	90	79	99
Chile	87	73	98	87	73	98
Chinese Taipei	93	81	100	93	76	100
Croatia	92	73	99	91	73	99
Cyprus	91	77	99	91	77	99
Czech Republic	91	69	98	90	69	98
Denmark	86	75	97	86	75	97
England	89	70	99	88	70	99
Finland	93	84	100	93	84	100
France	93	73	99	92	73	99
Georgia	88	69	97	87	69	95
Germany	93	82	99	92	82	99
Hong Kong SAR	89	82	96	88	80	96
Hungary	93	86	99	92	83	99
Iran, Islamic Rep. of	93	84	99	92	84	99
Ireland	91	74	99	91	74	99
Italy	93	81	100	93	81	100
Japan	91	83	99	90	83	99
Korea, Rep. of	95	88	100	95	88	100
Lithuania	93	80	99	93	80	99
Netherlands	94	78	99	93	78	99
New Zealand	90	85	97	90	81	97
Northern Ireland	92	82	99	91	82	99
Norway (5)	87	67	99	86	67	99
Oman	83	62	97	82	62	97
Poland	92	76	98	92	76	98
Portugal	95	87	98	95	87	98
Qatar	86	71	98	85	71	98

TIMSS 2019 Trend Scoring Reliability for Human Scored Items—Grade 4 Science (continued)

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Saudi Arabia	81	67	95	80	67	95
Serbia	90	81	99	89	77	99
Singapore	95	90	100	95	90	100
Slovak Republic	91	73	98	91	73	98
Spain	86	66	100	86	66	100
Sweden	90	77	99	90	77	99
United Arab Emirates	91	80	99	90	80	99
United States	94	83	100	93	83	100
International Average	**91**	**78**	**99**	**90**	**77**	**99**

TIMSS 2019 Trend Scoring Reliability for Human Scored Items—Grade 8 Mathematics

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Australia	96	89	100	96	89	100
Bahrain	98	89	100	95	65	100
Canada *	88	74	93	85	74	93
Chile	95	86	100	94	78	100
Chinese Taipei	97	90	100	96	79	100
England	95	63	100	95	63	100
Georgia	94	79	100	91	52	100
Hong Kong SAR	92	69	100	92	63	100
Hungary	97	83	100	95	74	100
Iran, Islamic Rep. of	97	89	100	97	89	100
Ireland	97	84	100	97	84	100
Israel	97	89	100	96	89	100
Italy	97	89	100	97	89	100
Japan	96	68	100	94	68	100
Jordan	99	97	100	98	93	100
Korea, Rep. of	94	49	100	93	49	100
Lithuania	98	92	100	96	74	100
Malaysia	96	84	100	93	60	100
New Zealand	95	81	100	95	81	100
Norway (9)	94	71	100	93	71	100
Oman	96	83	99	94	68	99
Qatar	96	83	100	95	80	100
Russian Federation	97	84	100	96	81	100
Saudi Arabia	97	88	100	97	88	99
Singapore	97	78	100	97	78	100
South Africa (9)	97	83	100	92	27	100
Sweden	96	74	100	95	74	100
Turkey	96	75	100	94	75	100
United Arab Emirates	97	86	100	96	85	100
United States	97	84	100	97	84	100
International Average	**96**	**81**	**100**	**95**	**74**	**100**

* Canada participated in trend scoring reliability for the benchmarking participants Ontario and Quebec.

TIMSS 2019 Trend Scoring Reliability for Human Scored Items—Grade 8 Science

Country	Score Point Agreement			Diagnostic Code Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement Across Items	
		Minimum	Maximum		Minimum	Maximum
Australia	92	78	97	91	78	97
Bahrain	89	74	98	87	65	98
Canada *	90	80	98	89	70	98
Chile	87	70	99	86	66	99
Chinese Taipei	92	81	99	90	74	99
England	91	81	98	90	81	98
Georgia	90	75	99	88	66	99
Hong Kong SAR	90	77	99	89	77	99
Hungary	92	83	97	91	78	97
Iran, Islamic Rep. of	93	81	100	92	81	100
Ireland	93	84	100	92	84	100
Israel	92	85	100	91	81	100
Italy	92	83	100	91	73	100
Japan	92	81	100	91	75	100
Jordan	98	94	100	98	94	100
Korea, Rep. of	95	87	99	95	83	99
Lithuania	95	78	100	94	78	100
Malaysia	90	79	98	89	79	98
New Zealand	91	82	98	90	82	98
Norway (9)	91	82	99	90	79	99
Oman	86	71	97	84	65	97
Qatar	89	80	99	87	73	99
Russian Federation	86	66	98	85	65	98
Saudi Arabia	88	78	99	87	72	99
Singapore	94	82	100	94	82	100
South Africa (9)	95	90	100	94	88	100
Sweden	91	78	100	91	78	100
Turkey	90	77	97	88	66	97
United Arab Emirates	92	88	98	91	84	98
United States	93	80	98	92	76	98
International Average	**91**	**80**	**99**	**90**	**76**	**99**

* Canada participated in trend scoring reliability for the benchmarking participants Ontario and Quebec.

Appendix 10C: TIMSS 2019 Cross-Country Scoring Reliability for Human Scored Items

TIMSS 2019 Cross-Country Scoring Reliability for Human Scored Items—Grade 4 Mathematics

Item	Total Valid Comparisons	Exact Percent Agreement	
		Score Point Agreement	Diagnostic Score Agreement
M03_03 – MP61034	295,299	99	99
M03_05 – MP61228	296,261	81	81
M03_06 – MP61166	297,000	97	97
M03_08 – MP61080	295,515	98	98
M03_10 – MP61076	296,892	100	100
M03_11 – MP61084	293,850	95	95
M05_01 – MP51206	215,204	97	97
M05_04 – MP51045	216,200	99	98
M05_06 – MP51030	215,248	97	97
M05_11 – MP51533	216,108	100	100
M05_12 – MP51080	211,916	95	91
Average Percent Agreement		**96**	**96**

TIMSS 2019 Cross-Country Scoring Reliability for Human Scored Items—Grade 4 Science

Item	Total Valid Comparisons	Exact Percent Agreement	
		Score Point Agreement	Diagnostic Score Agreement
S03_02 – SP61023	295,515	98	98
S03_03 – SP61054	293,491	87	87
S03_05 – SP61006	296,730	91	91
S03_09 – SP61088	295,245	91	91
S03_10 – SP61151	296,334	86	86
S03_11 – SP61150	294,925	82	82
S05_01 – SP51044	296,892	86	86
S05_04 – SP51168	296,152	88	86
S05_05 – SP51010	296,677	91	86
S05_07 – SP51059	295,138	75	75
S05_10 – SP51151	296,946	98	98
Average Percent Agreement		**89**	**88**

TIMSS 2019 Cross-Country Scoring Reliability for Human Scored Items—Grade 8 Mathematics

Item	Total Valid Comparisons	Exact Percent Agreement	
		Score Point Agreement	Diagnostic Score Agreement
M03_02 – MP62139	125,965	100	100
M03_04 – MP62142	125,895	98	98
M03_08 – MP62027	125,701	99	99
M03_10 – MP62244	125,518	97	97
M03_12 – MP62300	125,825	94	93
M03_13 – MP62254	125,350	71	71
M03_14 – MP62132A	126,000	100	100
M05_05 – MP52174A	126,000	98	98
M05_05 – MP52174B	125,965	99	98
M05_08 – MP52110	125,791	100	100
M05_09 – MP52105	124,740	88	88
M05_11 – MP52036	125,755	86	86
M05_12 – MP52502	125,721	96	96
M05_13 – MP52117	125,057	90	75
Average Percent Agreement		**94**	**93**

TIMSS 2019 Cross-Country Scoring Reliability for Human Scored Items—Grade 8 Science

Item	Total Valid Comparisons	Exact Percent Agreement	
		Score Point Agreement	Diagnostic Score Agreement
S03_03 – SP62275	125,965	93	93
S03_05 – SP62111	124,565	93	93
S03_06 – SP62116A	125,790	90	90
S03_06 – SP62116B	125,090	90	90
S03_06 – SP62116C	125,790	76	76
S03_10 – SP62162	125,965	85	85
S05_02 – SP52272	125,930	92	86
S05_03 – SP52085A	125,791	80	72
S05_03 – SP52085B	125,755	83	83
S05_04 – SP52094	126,000	96	96
S05_06 – SP52146	124,775	94	92
S05_10 – SP52214	126,000	98	98
S05_12 – SP52101	125,301	88	88
Average Percent Agreement		**89**	**88**

Appendix 10D: TIMSS 2019 Item Statistics by Booklet Position

TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 4 Mathematics (eTIMSS)

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
ME01 - Positions 1 & 4	11,670	11,603	53.1	48.3	2.7	4.7	0.1	2.8
ME02 - Positions 2 & 3	11,541	11,670	51.4	51.8	4.3	4.5	0.0	1.9
ME03 - Positions 1 & 4	11,596	11,542	50.3	49.4	4.4	5.9	0.1	1.2
ME04 - Positions 2 & 3	11,598	11,596	49.2	48.3	4.6	5.4	0.2	4.2
ME05 - Positions 1 & 4	11,634	11,598	51.2	48.2	2.6	4.5	0.1	2.9
ME06 - Positions 2 & 3	11,584	11,634	48.8	47.7	2.5	3.8	0.1	2.6
ME07 - Positions 1 & 4	11,635	11,585	49.7	48.7	2.5	2.9	0.0	0.8
ME08 - Positions 2 & 3	11,594	11,636	46.0	45.7	2.9	3.3	0.1	3.1
ME09 - Positions 1 & 4	11,614	11,594	48.7	45.7	2.5	4.2	0.1	2.8
ME10 - Positions 2 & 3	11,578	11,614	49.2	48.5	3.7	4.9	0.1	3.4
ME11 - Positions 1 & 4	11,613	11,577	51.0	48.8	4.2	5.5	0.1	2.0
ME12 - Positions 2 & 3	11,634	11,613	48.6	46.3	4.0	5.2	0.1	3.5
ME13 - Positions 1 & 4	11,605	11,633	52.1	50.1	2.5	3.3	0.0	1.8
ME14 - Positions 2 & 3	11,603	11,605	49.3	48.6	3.2	4.3	0.1	2.8
Overall	**162,499**	**162,500**	**49.9**	**48.3**	**3.3**	**4.5**	**0.1**	**2.5**

TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 4 Mathematics (paperTIMSS)

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
MP01 - Positions 1 & 4	5,984	5,977	52.7	48.8	3.6	5.1	0.1	4.1
MP02 - Positions 2 & 3	5,997	5,984	49.7	51.3	7.3	6.3	0.3	1.9
MP03 - Positions 1 & 4	5,990	5,997	49.0	46.6	6.8	8.4	0.1	3.4
MP04 - Positions 2 & 3	5,984	5,990	50.1	49.0	6.3	6.7	0.3	4.6
MP05 - Positions 1 & 4	5,960	5,983	51.7	48.8	4.7	5.9	0.1	3.8
MP06 - Positions 2 & 3	5,974	5,960	49.6	48.7	5.3	5.9	0.2	1.8
MP07 - Positions 1 & 4	5,967	5,974	50.9	49.6	3.8	5.1	0.1	1.4
MP08 - Positions 2 & 3	6,002	5,967	42.2	42.7	8.0	6.6	0.3	3.6
MP09 - Positions 1 & 4	5,994	6,002	50.0	46.4	6.2	7.2	0.1	4.1
MP10 - Positions 2 & 3	5,976	5,994	47.5	47.6	5.9	5.4	0.3	3.5
MP11 - Positions 1 & 4	5,973	5,976	49.4	47.6	6.1	8.0	0.1	3.6
MP12 - Positions 2 & 3	6,005	5,973	47.9	47.0	5.6	5.8	0.3	4.1
MP13 - Positions 1 & 4	5,983	6,005	51.7	49.3	5.3	6.1	0.1	3.9
MP14 - Positions 2 & 3	5,978	5,982	49.2	49.3	6.2	6.0	0.2	2.8
Overall	**83,767**	**83,764**	**49.4**	**48.1**	**5.8**	**6.3**	**0.2**	**3.3**

TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 4 Less Difficult Mathematics (paperTIMSS)

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
MN01 - Positions 1 & 4	4,399	4,375	48.0	47.2	6.1	6.7	0.2	2.2
MN03 - Positions 1 & 4	4,434	4,389	48.6	44.6	5.0	6.6	0.2	4.5
MN04 - Positions 2 & 3	4,419	4,434	49.2	49.9	7.1	7.2	0.3	4.1
MN05 - Positions 1 & 4	4,435	4,419	53.4	50.2	5.3	7.1	0.2	2.8
MN07 - Positions 1 & 4	4,440	4,413	53.6	51.6	7.3	7.9	0.3	2.9
MN09 - Positions 1 & 4	4,397	4,405	53.3	50.0	4.3	7.3	0.2	4.3
MN11 - Positions 1 & 4	4,407	4,373	44.3	42.9	6.5	7.5	0.1	2.9
MN12 - Positions 2 & 3	4,391	4,407	49.3	48.8	6.5	6.5	0.2	2.9
MN13 - Positions 1 & 4	4,420	4,391	48.2	46.8	5.2	5.6	0.2	2.6
MN14 - Positions 2 & 3	4,375	4,420	50.8	50.6	7.5	7.2	0.2	2.7
MP02 - Positions 2 & 3	4,389	4,398	28.3	28.0	13.5	12.8	0.4	4.3
MP03 - Positions 2 & 3	4,412	4,435	27.8	27.9	13.8	13.6	0.3	4.0
MP08 - Positions 2 & 3	4,405	4,440	23.6	23.0	13.4	12.6	0.7	7.5
MP13 - Positions 2 & 3	4,373	4,397	28.5	29.3	9.5	8.2	0.5	3.4
Overall	**61,696**	**61,696**	**43.4**	**42.2**	**7.9**	**8.3**	**0.3**	**3.7**

TIMSS 2019 Country Average Item Statistics by Booklet Position—Grade 4 Mathematics

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
Albania	4,417	4,417	62.3	61.8	4.1	5.0	0.2	2.8
Armenia	5,380	5,380	43.9	42.2	13.4	13.8	0.7	7.4
Australia	5,879	5,879	48.9	47.4	3.0	3.8	0.2	2.6
Austria	4,463	4,463	51.4	50.3	4.7	6.1	0.0	1.2
Azerbaijan	5,220	5,220	50.5	50.0	10.0	9.0	0.3	2.8
Bahrain	5,758	5,758	40.2	38.4	7.9	8.6	0.1	1.9
Belgium (Flemish)	4,646	4,646	52.4	51.7	3.9	3.8	0.0	1.6
Bosnia and Herzegovina	5,612	5,612	50.3	49.3	12.0	11.8	0.1	3.4
Bulgaria	4,267	4,267	52.9	51.2	8.6	8.3	0.1	2.8
Canada	13,576	13,576	43.2	41.6	2.7	4.0	0.2	4.0
Chile	4,161	4,161	32.4	28.8	5.1	8.0	0.4	10.0
Chinese Taipei	3,763	3,763	67.9	67.5	0.8	1.0	0.0	0.1
Croatia	3,783	3,783	43.9	43.4	3.8	4.5	0.0	1.1
Cyprus	4,061	4,061	52.7	51.2	3.6	4.3	0.1	3.8
Czech Republic	4,689	4,689	50.5	48.6	4.4	5.5	0.0	1.5
Denmark	3,213	3,213	50.0	46.9	4.7	6.8	0.2	6.4
England	3,393	3,393	55.1	55.0	2.6	3.3	0.1	1.0
Finland	4,723	4,723	50.3	48.9	4.2	5.1	0.1	1.8
France	4,179	4,179	39.7	38.2	6.9	8.7	0.1	4.1
Georgia	3,765	3,765	37.3	35.5	7.8	9.7	0.1	4.7
Germany	3,434	3,434	47.0	45.4	5.0	6.4	0.1	2.6
Hong Kong SAR	2,964	2,964	68.4	67.9	0.9	1.1	0.0	0.2
Hungary	4,569	4,569	50.0	49.3	2.3	2.8	0.0	1.1
Iran, Islamic Rep. of	5,989	5,989	34.8	32.5	10.6	11.8	0.6	10.8
Ireland	4,566	4,566	56.7	55.7	2.4	2.5	0.0	0.9
Italy	3,738	3,738	45.8	43.2	4.6	6.5	0.1	4.6
Japan	4,193	4,193	68.5	68.2	1.7	2.0	0.0	1.1
Kazakhstan	4,786	4,786	46.3	46.0	4.6	5.3	0.1	2.9
Korea, Rep. of	3,893	3,893	68.6	67.7	1.0	1.3	0.0	0.2
Kosovo	4,490	4,491	50.6	49.9	5.2	5.4	0.2	2.6
Kuwait	4,417	4,417	39.5	38.0	5.6	5.9	0.4	3.4
Latvia	4,474	4,473	57.5	55.9	3.4	4.0	0.1	1.8
Lithuania	3,739	3,739	51.5	50.2	2.8	3.4	0.1	0.7

TIMSS 2019 Country Average Item Statistics by Booklet Position—Grade 4 Mathematics (continued)

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
Malta	3,626	3,626	43.6	42.1	2.2	2.9	0.0	1.1
Montenegro	5,060	5,060	51.4	50.2	13.8	14.0	0.2	4.2
Morocco	7,712	7,712	37.1	34.9	7.0	7.8	0.2	3.9
Netherlands	3,336	3,336	51.1	49.1	3.0	4.0	0.1	1.8
New Zealand	5,002	5,002	41.9	40.7	3.7	4.6	0.3	3.8
North Macedonia	3,264	3,264	56.9	55.6	5.9	6.8	0.3	5.1
Northern Ireland	3,491	3,491	61.5	60.9	2.6	3.0	0.1	0.9
Norway (5)	3,938	3,938	54.9	52.8	3.7	5.0	0.3	4.1
Oman	6,801	6,801	30.6	28.9	4.3	5.7	0.3	4.5
Pakistan	3,942	3,942	27.2	25.9	18.5	17.7	0.6	5.8
Philippines	5,495	5,495	26.2	25.5	5.8	6.8	0.5	3.3
Poland	4,881	4,881	50.0	48.2	7.7	8.3	0.1	2.2
Portugal	4,297	4,297	49.0	46.9	2.7	4.0	0.0	3.2
Qatar	4,929	4,929	32.9	30.6	3.9	5.2	0.1	5.0
Russian Federation	4,022	4,022	59.8	59.0	2.2	2.9	0.0	1.0
Saudi Arabia	5,445	5,445	43.2	42.1	5.0	5.6	0.2	2.8
Serbia	4,373	4,373	48.6	47.6	7.9	8.5	0.2	4.4
Singapore	5,983	5,983	72.5	72.2	0.5	0.7	0.0	0.2
Slovak Republic	4,243	4,243	45.0	43.2	4.4	5.1	0.0	1.6
South Africa (5)	11,842	11,842	36.3	35.0	3.7	4.4	0.1	3.0
Spain	9,543	9,543	44.8	43.1	3.5	4.7	0.1	2.1
Sweden	3,958	3,958	49.3	46.4	4.8	7.2	0.3	5.7
Turkey (5)	4,028	4,028	48.1	47.5	1.9	2.6	0.0	0.9
United Arab Emirates	25,785	25,785	39.3	37.4	2.7	3.9	0.1	3.3
United States	8,769	8,769	50.8	49.2	1.1	1.9	0.1	2.9
International Average	**307,965**	**307,965**	**48.5**	**47.1**	**4.9**	**5.7**	**0.2**	**3.0**
Ontario, Canada	3,810	3,810	45.4	43.6	2.4	3.7	0.1	4.2
Quebec, Canada	3,816	3,816	50.3	48.9	2.6	3.8	0.0	2.7
Moscow City, Russian Fed.	3,842	3,842	66.4	65.6	1.6	1.9	0.0	0.4
Madrid, Spain	3,385	3,385	46.7	44.8	3.3	4.0	0.0	1.5
Abu Dhabi, UAE	9,001	8,998	31.3	29.4	2.9	3.9	0.1	3.0
Dubai, UAE	7,262	7,262	52.6	51.2	1.6	2.5	0.0	2.4

TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 4 Science (eTIMSS)

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
SE01 - Positions 2 & 3	11,666	11,617	51.0	51.1	2.5	3.5	0.0	4.2
SE02 - Positions 1 & 4	11,554	11,667	55.0	52.6	3.5	4.5	0.1	1.9
SE03 - Positions 2 & 3	11,585	11,554	53.0	52.1	2.9	4.1	0.0	2.5
SE04 - Positions 1 & 4	11,603	11,584	58.8	57.3	2.5	4.0	0.1	1.1
SE05 - Positions 2 & 3	11,630	11,603	53.7	53.0	3.6	4.8	0.1	3.4
SE06 - Positions 1 & 4	11,594	11,630	50.6	47.7	5.5	8.4	0.2	2.2
SE07 - Positions 2 & 3	11,624	11,594	46.7	46.0	3.2	4.2	0.1	3.6
SE08 - Positions 1 & 4	11,603	11,625	56.1	54.9	1.9	3.3	0.1	1.5
SE09 - Positions 2 & 3	11,603	11,603	53.2	52.6	4.4	5.0	0.1	2.5
SE10 - Positions 1 & 4	11,596	11,603	53.6	52.2	1.5	2.9	0.1	1.6
SE11 - Positions 2 & 3	11,592	11,596	51.0	50.5	6.1	6.6	0.1	3.5
SE12 - Positions 1 & 4	11,647	11,592	56.0	53.1	2.2	3.7	0.1	1.9
SE13 - Positions 2 & 3	11,596	11,647	49.9	50.1	2.9	3.4	0.1	4.1
SE14 - Positions 1 & 4	11,617	11,596	57.1	54.8	1.8	3.1	0.1	1.4
Overall	**162,510**	**162,511**	**53.3**	**52.0**	**3.2**	**4.4**	**0.1**	**2.5**

TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 4 Science (paperTIMSS)

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
SP01 - Positions 2 & 3	5,977	5,987	48.3	48.5	6.5	7.0	0.2	3.6
SP02 - Positions 1 & 4	6,005	5,977	51.0	49.4	7.5	7.8	0.2	2.1
SP03 - Positions 2 & 3	5,984	6,005	48.9	48.1	6.2	6.0	0.0	2.6
SP04 - Positions 1 & 4	5,994	5,984	55.5	52.9	4.9	6.6	0.1	1.7
SP05 - Positions 2 & 3	5,948	5,994	51.2	51.2	5.5	5.5	0.1	2.5
SP06 - Positions 1 & 4	5,974	5,948	50.1	46.6	6.9	9.7	0.2	2.2
SP07 - Positions 2 & 3	5,952	5,974	43.6	43.5	6.2	6.1	0.1	3.3
SP08 - Positions 1 & 4	6,017	5,952	55.6	53.6	4.3	5.5	0.1	1.7
SP09 - Positions 2 & 3	5,983	6,017	48.6	49.2	6.7	6.4	0.2	2.4
SP10 - Positions 1 & 4	6,010	5,983	52.4	49.5	3.4	5.8	0.1	2.7
SP11 - Positions 2 & 3	5,959	6,010	50.0	49.0	7.4	7.9	0.2	5.0
SP12 - Positions 1 & 4	6,011	5,959	53.6	51.6	5.4	6.0	0.0	2.8
SP13 - Positions 2 & 3	5,979	6,011	47.5	46.8	5.8	8.7	0.0	0.4
SP14 - Positions 1 & 4	5,987	5,979	56.1	53.6	3.4	5.3	0.2	2.6
Overall	**83,780**	**83,780**	**50.9**	**49.5**	**5.7**	**6.7**	**0.1**	**2.5**

TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 4 Science (Less Difficult paperTIMSS)

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
SP01 - Positions 2 & 3	4,381	4,396	32.2	31.3	12.0	12.0	0.4	6.3
SP02 - Positions 1 & 4	4,406	4,382	33.9	32.9	13.9	13.3	0.6	4.8
SP03 - Positions 2 & 3	4,408	4,406	35.9	33.8	9.7	11.4	0.1	5.6
SP04 - Positions 1 & 4	4,429	4,408	38.4	37.5	10.0	10.8	0.4	3.5
SP05 - Positions 2 & 3	4,413	4,428	33.0	32.4	9.3	11.3	0.0	4.6
SP06 - Positions 1 & 4	4,431	4,413	31.3	30.3	15.3	16.9	0.7	4.5
SP07 - Positions 2 & 3	4,420	4,431	30.3	30.4	12.6	12.1	0.3	6.7
SP08 - Positions 1 & 4	4,421	4,420	35.7	34.0	9.4	11.2	0.4	3.6
SP09 - Positions 2 & 3	4,379	4,421	34.1	32.9	9.3	11.2	0.3	4.9
SP10 - Positions 1 & 4	4,409	4,378	33.5	32.4	9.0	10.7	0.5	4.3
SP11 - Positions 2 & 3	4,391	4,408	30.4	29.2	13.7	15.6	0.2	6.9
SP12 - Positions 1 & 4	4,411	4,392	36.1	34.1	10.6	10.5	0.1	4.6
SP13 - Positions 2 & 3	4,398	4,410	31.9	29.8	10.3	14.5	0.0	3.2
SP14 - Positions 1 & 4	4,396	4,399	37.9	35.6	7.8	9.4	0.4	4.6
Overall	**61,693**	**61,692**	**33.9**	**32.6**	**10.9**	**12.2**	**0.3**	**4.9**

TIMSS 2019 Country Average Item Statistics by Booklet Position—Grade 4 Science

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
Albania	4,422	4,422	48.7	47.0	6.0	6.9	0.3	3.6
Armenia	5,391	5,391	43.0	40.2	13.0	15.9	0.3	5.0
Australia	5,878	5,878	56.9	55.7	2.9	3.7	0.1	2.0
Austria	4,464	4,464	52.1	51.2	4.6	5.6	0.0	1.7
Azerbaijan	5,185	5,185	38.0	37.4	13.6	13.2	0.4	3.1
Bahrain	5,760	5,760	49.5	46.9	4.5	6.5	0.1	2.0
Belgium (Flemish)	4,653	4,653	47.9	47.5	4.8	4.6	0.0	1.6
Bosnia and Herzegovina	5,611	5,611	41.1	40.0	12.6	13.8	0.2	4.3
Bulgaria	4,267	4,267	59.1	58.2	6.6	7.1	0.0	1.7
Canada	13,579	13,579	51.9	50.4	2.7	3.8	0.1	3.2
Chile	4,158	4,158	44.9	42.0	3.9	6.5	0.4	8.1
Chinese Taipei	3,764	3,764	60.8	60.0	2.0	2.6	0.0	0.5
Croatia	3,784	3,784	52.3	51.5	2.7	3.5	0.0	0.8
Cyprus	4,062	4,062	52.1	50.3	4.4	6.0	0.1	4.4
Czech Republic	4,688	4,688	54.6	53.6	4.0	5.0	0.0	1.1
Denmark	3,220	3,220	52.6	51.0	3.5	4.8	0.1	3.3
England	3,387	3,387	54.9	54.1	2.3	3.0	0.0	0.7
Finland	4,711	4,711	59.3	58.6	2.9	3.6	0.0	0.7
France	4,184	4,184	47.0	45.3	6.3	8.7	0.1	5.0
Georgia	3,764	3,764	39.1	36.7	8.2	10.8	0.2	6.3
Germany	3,432	3,432	52.8	50.7	4.3	6.3	0.1	4.2
Hong Kong SAR	2,968	2,968	54.9	53.6	2.2	3.4	0.0	0.8
Hungary	4,570	4,570	55.8	54.8	2.2	3.1	0.0	1.7
Iran, Islamic Rep. of	5,994	5,994	41.1	39.7	9.7	11.5	0.4	6.9
Ireland	4,576	4,576	55.1	54.0	2.7	3.2	0.0	0.9
Italy	3,740	3,740	50.4	48.7	4.2	6.4	0.1	4.7
Japan	4,192	4,192	62.6	61.1	2.2	2.5	0.0	1.1
Kazakhstan	4,791	4,791	47.4	46.9	5.0	6.0	0.0	2.3
Korea, Rep. of	3,891	3,891	66.7	67.0	1.1	1.3	0.0	0.1
Kosovo	4,486	4,484	34.8	33.6	9.3	10.0	0.2	3.2
Kuwait	4,412	4,412	34.0	32.8	9.2	10.1	0.3	3.4
Latvia	4,476	4,476	59.0	57.6	2.8	3.3	0.1	1.2
Lithuania	3,738	3,738	54.4	54.1	2.7	3.3	0.0	0.6

TIMSS 2019 Country Average Item Statistics by Booklet Position—Grade 4 Science (continued)

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
Malta	3,625	3,625	47.4	46.2	2.5	3.4	0.0	1.1
Montenegro	5,068	5,068	41.7	39.4	14.3	15.8	0.4	7.7
Morocco	7,714	7,714	29.9	28.1	10.5	12.7	0.2	4.7
Netherlands	3,337	3,337	50.9	49.7	2.8	3.4	0.1	1.1
New Zealand	5,003	5,003	49.8	48.8	3.3	4.8	0.2	2.2
North Macedonia	3,262	3,262	38.7	36.8	10.4	12.8	0.6	8.8
Northern Ireland	3,490	3,490	52.9	51.9	3.3	3.7	0.1	1.0
Norway (5)	3,940	3,940	56.8	55.8	2.9	3.4	0.1	1.7
Oman	6,811	6,811	39.9	38.3	6.0	7.7	0.1	3.2
Pakistan	3,919	3,919	21.9	21.9	23.7	23.7	0.6	4.9
Philippines	5,501	5,501	20.9	20.2	10.3	12.0	0.3	4.2
Poland	4,875	4,875	55.6	55.0	6.5	7.2	0.0	1.2
Portugal	4,297	4,297	49.4	46.9	3.0	5.2	0.0	5.1
Qatar	4,933	4,933	41.4	39.8	4.4	6.3	0.3	6.9
Russian Federation	4,021	4,021	62.9	62.2	2.3	3.2	0.1	1.2
Saudi Arabia	5,448	5,448	35.7	35.0	9.0	9.4	0.1	2.7
Serbia	4,376	4,376	55.4	52.8	6.3	7.7	0.2	4.2
Singapore	5,983	5,983	68.0	67.2	0.8	1.1	0.0	0.3
Slovak Republic	4,246	4,246	53.4	52.6	3.6	5.1	0.0	1.4
South Africa (5)	11,852	11,852	26.1	24.6	5.1	7.2	0.3	6.2
Spain	9,544	9,544	51.5	50.2	3.1	4.1	0.0	1.7
Sweden	3,951	3,951	56.7	55.2	3.4	4.7	0.1	2.1
Turkey (5)	4,028	4,028	54.9	53.4	2.2	3.0	0.0	1.7
United Arab Emirates	25,796	25,796	45.5	43.8	3.3	5.1	0.2	4.1
United States	8,770	8,770	56.5	55.1	1.4	2.4	0.2	3.3
International Average	**307,988**	**307,986**	**48.9**	**47.6**	**5.4**	**6.6**	**0.1**	**3.0**
Ontario, Canada	3,806	3,806	53.0	51.1	2.5	3.7	0.1	3.9
Quebec, Canada	3,828	3,828	52.3	50.8	2.7	3.7	0.1	2.1
Moscow City, Russian Fed.	3,841	3,841	68.8	67.5	1.7	2.3	0.0	0.8
Madrid, Spain	3,388	3,388	52.5	51.3	2.7	3.5	0.0	1.0
Abu Dhabi, UAE	9,004	9,004	36.7	35.1	3.8	5.7	0.2	4.6
Dubai, UAE	7,265	7,265	57.6	55.9	1.7	2.7	0.1	2.3

TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 8 Mathematics (eTIMSS)

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
ME01 - Positions 1 & 4	8,099	8,152	46.0	44.9	3.9	5.4	0.1	1.2
ME02 - Positions 2 & 3	8,140	8,099	43.1	41.8	6.8	6.5	0.1	2.0
ME03 - Positions 1 & 4	8,134	8,140	46.0	43.3	4.1	6.4	0.1	1.9
ME04 - Positions 2 & 3	8,164	8,134	43.0	39.9	6.5	7.2	0.1	2.9
ME05 - Positions 1 & 4	8,153	8,164	47.0	45.1	3.3	5.6	0.0	1.1
ME06 - Positions 2 & 3	8,095	8,153	42.9	41.4	8.0	8.4	0.0	2.0
ME07 - Positions 1 & 4	8,090	8,095	49.6	47.5	3.7	5.3	0.0	0.6
ME08 - Positions 2 & 3	8,120	8,090	40.7	38.7	7.5	8.2	0.1	2.1
ME09 - Positions 1 & 4	8,138	8,120	37.0	35.5	5.3	8.3	0.1	1.0
ME10 - Positions 2 & 3	8,106	8,137	43.6	41.1	8.2	10.5	0.1	2.7
ME11 - Positions 1 & 4	8,101	8,106	46.4	44.4	4.0	5.8	0.1	1.2
ME12 - Positions 2 & 3	8,187	8,101	40.0	37.3	8.3	9.8	0.2	2.7
ME13 - Positions 1 & 4	8,188	8,187	47.2	42.9	3.4	5.6	0.0	1.4
ME14 - Positions 2 & 3	8,152	8,188	43.5	42.7	4.4	4.7	0.1	1.6
Overall	**113,867**	**113,866**	**44.0**	**41.9**	**5.5**	**7.0**	**0.1**	**1.8**

TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 8 Mathematics (paperTIMSS)

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
MP01 - Positions 1 & 4	8,070	8,086	39.0	36.9	5.1	6.2	0.1	2.0
MP02 - Positions 2 & 3	8,026	8,070	32.0	31.9	9.2	8.7	0.1	2.3
MP03 - Positions 1 & 4	8,045	8,026	37.7	33.2	7.7	9.9	0.1	2.9
MP04 - Positions 2 & 3	8,063	8,045	35.0	34.0	9.9	9.3	0.2	3.2
MP05 - Positions 1 & 4	8,053	8,063	41.4	38.5	5.7	7.6	0.1	1.9
MP06 - Positions 2 & 3	8,083	8,053	31.4	31.9	8.8	8.4	0.1	1.7
MP07 - Positions 1 & 4	8,088	8,083	38.4	36.8	5.7	7.2	0.0	1.0
MP08 - Positions 2 & 3	8,052	8,088	31.3	30.5	10.0	9.7	0.1	1.4
MP09 - Positions 1 & 4	8,051	8,052	30.6	28.7	7.6	8.9	0.1	1.5
MP10 - Positions 2 & 3	8,107	8,051	34.7	33.8	11.0	11.7	0.2	1.9
MP11 - Positions 1 & 4	8,118	8,107	35.9	33.6	6.6	8.5	0.1	1.9
MP12 - Positions 2 & 3	8,113	8,118	30.3	29.1	9.9	10.4	0.1	2.9
MP13 - Positions 1 & 4	8,087	8,113	36.7	32.7	6.2	6.5	0.1	1.9
MP14 - Positions 2 & 3	8,086	8,088	33.6	33.5	6.7	6.7	0.2	2.0
Overall	**113,042**	**113,043**	**34.9**	**33.2**	**7.9**	**8.5**	**0.1**	**2.0**

TIMSS 2019 Country Average Item Statistics by Booklet Position—Grade 8 Mathematics

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
Australia	9,002	9,002	49.4	48.3	5.0	5.7	0.2	1.4
Bahrain	5,724	5,724	39.3	37.4	11.5	11.8	0.0	0.8
Chile	4,100	4,100	28.5	25.7	10.8	13.2	0.1	4.0
Chinese Taipei	4,914	4,914	66.3	65.7	1.5	2.0	0.0	0.2
Cyprus	3,515	3,515	42.5	40.6	6.6	7.6	0.0	1.1
Egypt	7,201	7,201	25.7	24.0	7.9	8.0	0.2	2.4
England	3,345	3,345	43.2	41.1	7.1	8.9	0.1	1.5
Finland	4,835	4,835	40.4	38.8	5.5	6.9	0.2	1.6
France	3,869	3,869	35.1	32.9	8.9	11.3	0.0	2.4
Georgia	3,309	3,309	31.5	29.1	11.5	13.5	0.1	2.3
Hong Kong SAR	3,255	3,255	58.9	58.3	2.8	3.2	0.2	0.8
Hungary	4,559	4,559	46.0	44.8	4.7	5.3	0.0	0.2
Iran, Islamic Rep. of	5,975	5,975	32.3	29.5	10.5	12.4	0.1	3.3
Ireland	4,109	4,109	48.2	47.2	5.4	6.1	0.3	1.5
Israel	3,725	3,725	45.2	42.0	5.8	7.4	0.0	2.0
Italy	3,618	3,618	38.8	36.2	7.3	9.3	0.1	2.5
Japan	4,444	4,444	66.0	65.0	2.2	2.8	0.1	0.6
Jordan	7,172	7,172	25.3	23.3	5.5	6.6	0.1	1.6
Kazakhstan	4,447	4,447	38.8	38.0	9.4	10.1	0.1	2.2
Korea, Rep. of	3,858	3,858	65.3	64.4	1.9	2.4	0.1	0.3
Kuwait	4,569	4,569	23.4	21.8	6.0	6.4	0.2	2.5
Lebanon	4,724	4,724	25.9	24.4	18.3	19.1	0.3	4.8
Lithuania	3,823	3,823	42.8	41.2	5.8	6.5	0.0	0.4
Malaysia	7,065	7,065	37.7	35.5	1.7	2.6	0.0	1.8
Morocco	8,431	8,431	19.0	18.2	15.3	14.7	0.1	2.4
New Zealand	6,025	6,025	42.4	41.3	6.0	6.8	0.2	1.7
Norway (9)	4,541	4,541	41.1	38.1	9.6	11.9	0.4	3.7
Oman	6,745	6,745	26.6	24.6	5.6	6.2	0.1	3.0
Portugal	3,369	3,369	40.2	37.2	5.8	7.6	0.0	1.4

TIMSS 2019 Country Average Item Statistics by Booklet Position—Grade 8 Mathematics (continued)

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
Qatar	3,882	3,882	29.7	27.3	4.8	6.5	0.1	2.5
Romania	4,485	4,485	41.9	38.9	9.9	10.8	0.1	2.2
Russian Federation	3,900	3,900	50.6	48.2	6.8	8.4	0.0	2.0
Saudi Arabia	5,680	5,680	23.7	21.5	4.8	5.6	0.0	1.0
Singapore	4,845	4,845	67.0	66.2	1.1	1.4	0.0	0.4
South Africa (9)	20,796	20,796	21.4	20.9	4.2	4.5	0.0	2.6
Sweden	3,970	3,970	41.6	37.7	7.5	10.7	0.2	4.2
Turkey	4,075	4,075	38.4	37.0	5.6	6.8	0.0	0.7
United Arab Emirates	22,327	22,326	34.1	31.2	3.3	4.5	0.0	1.6
United States	8,683	8,683	44.8	42.5	1.9	2.7	0.1	2.5
International Average	**226,911**	**226,910**	**40.0**	**38.1**	**6.6**	**7.6**	**0.1**	**1.9**
Ontario, Canada	3,764	3,764	46.6	43.5	4.0	5.7	0.2	3.5
Quebec, Canada	3,173	3,173	50.7	47.7	3.7	5.4	0.2	2.6
Moscow City, Russian Fed.	3,780	3,780	59.1	56.3	5.6	7.4	0.0	1.2
Gauteng, RSA (9)	5,621	5,621	23.0	22.4	3.4	3.6	0.0	2.1
Western Cape, RSA (9)	5,340	5,340	27.0	26.4	4.4	4.8	0.0	2.3
Abu Dhabi, UAE	8,201	8,201	29.0	26.2	2.9	3.9	0.1	1.5
Dubai, UAE	5,726	5,726	47.4	44.3	3.0	4.6	0.0	1.6

TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 8 Science (eTIMSS)

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
SE01 - Positions 2 & 3	8,074	8,160	52.5	51.4	5.6	6.7	0.1	1.2
SE02 - Positions 1 & 4	8,155	8,074	48.9	46.8	2.9	4.2	0.1	0.7
SE03 - Positions 2 & 3	8,124	8,155	45.0	43.7	5.1	5.4	0.2	0.8
SE04 - Positions 1 & 4	8,176	8,124	47.8	45.0	5.3	7.4	0.1	0.7
SE05 - Positions 2 & 3	8,149	8,176	48.6	48.5	5.1	5.8	0.1	0.8
SE06 - Positions 1 & 4	8,100	8,149	43.1	40.7	8.6	10.4	0.0	0.7
SE07 - Positions 2 & 3	8,080	8,100	45.1	44.6	3.9	4.2	0.1	0.6
SE08 - Positions 1 & 4	8,123	8,080	47.6	45.9	2.6	4.5	0.0	0.8
SE09 - Positions 2 & 3	8,126	8,124	39.8	39.4	4.5	4.8	0.1	1.0
SE10 - Positions 1 & 4	8,124	8,126	48.1	46.9	2.2	3.3	0.1	0.5
SE11 - Positions 2 & 3	8,104	8,124	49.4	48.5	3.8	4.1	0.1	1.1
SE12 - Positions 1 & 4	8,195	8,104	51.3	48.4	1.5	2.9	0.0	0.8
SE13 - Positions 2 & 3	8,179	8,195	52.8	52.5	2.8	3.3	0.1	1.4
SE14 - Positions 1 & 4	8,159	8,179	48.8	45.8	2.0	3.7	0.1	0.6
Overall	**113,868**	**113,870**	**47.8**	**46.3**	**4.0**	**5.0**	**0.1**	**0.8**

TIMSS 2019 International Average Item Block Statistics by Booklet Position—Grade 8 Science (paperTIMSS)

Item Block	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
SP01 - Positions 2 & 3	8,068	8,079	43.2	42.0	7.8	8.4	0.1	1.7
SP02 - Positions 1 & 4	8,039	8,068	40.7	38.4	6.0	6.8	0.1	1.3
SP03 - Positions 2 & 3	8,042	8,039	38.7	37.9	6.1	6.7	0.1	2.1
SP04 - Positions 1 & 4	8,063	8,042	39.1	36.4	8.9	10.4	0.1	1.9
SP05 - Positions 2 & 3	8,055	8,063	39.3	38.8	8.3	9.2	0.1	1.9
SP06 - Positions 1 & 4	8,080	8,055	36.9	34.9	10.9	12.5	0.0	1.7
SP07 - Positions 2 & 3	8,092	8,080	38.6	38.6	6.8	6.3	0.1	1.0
SP08 - Positions 1 & 4	8,064	8,092	39.9	37.3	5.5	7.0	0.1	1.1
SP09 - Positions 2 & 3	8,049	8,064	33.4	33.1	7.0	6.9	0.1	1.5
SP10 - Positions 1 & 4	8,105	8,049	40.9	39.7	5.1	5.7	0.0	1.3
SP11 - Positions 2 & 3	8,105	8,105	38.1	37.8	7.1	7.7	0.1	1.3
SP12 - Positions 1 & 4	8,115	8,104	42.5	39.5	4.0	4.8	0.0	1.3
SP13 - Positions 2 & 3	8,092	8,115	42.7	41.4	6.4	7.1	0.2	2.9
SP14 - Positions 1 & 4	8,080	8,092	40.3	37.0	4.0	5.8	0.0	2.1
Overall	**113,049**	**113,047**	**39.6**	**38.1**	**6.7**	**7.5**	**0.1**	**1.6**

TIMSS 2019 Country Average Item Statistics by Booklet Position—Grade 8 Science

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
Australia	9,002	9,002	53.7	52.2	3.8	4.0	0.2	1.1
Bahrain	5,719	5,719	43.3	41.4	4.9	5.9	0.0	0.7
Chile	4,097	4,097	38.8	36.5	6.5	8.6	0.1	1.6
Chinese Taipei	4,910	4,910	59.1	59.1	1.9	2.1	0.0	0.1
Cyprus	3,520	3,520	42.1	40.2	5.7	6.3	0.0	1.0
Egypt	7,200	7,199	28.9	27.4	9.3	9.6	0.1	2.0
England	3,358	3,358	48.0	46.0	4.0	5.4	0.2	1.0
Finland	4,843	4,843	53.1	52.1	3.2	3.8	0.2	0.7
France	3,870	3,870	42.5	40.3	5.4	7.2	0.1	1.2
Georgia	3,308	3,308	34.2	31.9	9.8	12.6	0.1	1.4
Hong Kong SAR	3,253	3,253	44.7	42.8	3.6	4.8	0.2	0.7
Hungary	4,558	4,558	52.0	51.2	3.8	4.2	0.0	0.0
Iran, Islamic Rep. of	5,976	5,976	37.3	34.9	8.1	9.5	0.1	2.1
Ireland	4,097	4,097	49.9	49.2	3.8	4.1	0.2	1.1
Israel	3,721	3,721	47.2	45.3	4.2	5.3	0.0	0.6
Italy	3,618	3,618	44.5	42.7	5.3	6.2	0.0	0.9
Japan	4,442	4,442	58.9	58.3	2.1	2.6	0.0	0.4
Jordan	7,174	7,174	36.6	34.8	4.8	5.9	0.1	1.1
Kazakhstan	4,453	4,453	40.9	39.7	8.3	9.5	0.1	1.9
Korea, Rep. of	3,858	3,858	55.5	54.7	2.0	2.5	0.0	0.1
Kuwait	4,569	4,569	35.5	34.4	5.3	6.0	0.1	1.5
Lebanon	4,714	4,714	26.3	24.6	17.1	18.6	0.1	3.6
Lithuania	3,823	3,823	48.9	47.9	3.5	4.2	0.0	0.0
Malaysia	7,064	7,064	42.5	41.3	1.1	1.6	0.0	0.8
Morocco	8,444	8,444	26.7	25.7	13.7	14.7	0.0	1.8
New Zealand	6,021	6,021	48.9	47.2	4.3	5.0	0.1	1.2
Norway (9)	4,538	4,538	44.2	42.6	6.3	8.1	0.3	1.7
Oman	6,745	6,745	38.9	37.4	4.4	4.9	0.1	1.9
Portugal	3,362	3,362	48.3	46.9	3.8	4.9	0.0	0.3
Qatar	3,881	3,881	41.2	39.0	3.7	5.0	0.1	1.2
Romania	4,489	4,489	42.0	39.6	8.6	9.9	0.1	1.4
Russian Federation	3,899	3,899	52.6	51.5	4.7	5.6	0.0	0.7

TIMSS 2019 Country Average Item Statistics by Booklet Position—Grade 8 Science (continued)

Country	Sample Sizes		Average Percent Correct Across Items (Weighted)		Average Percent Omitted Responses Across Items (Weighted)		Average Percent Not Reached Across Items (Weighted)	
	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4	Positions 1 & 3	Positions 2 & 4
Saudi Arabia	5,678	5,678	35.2	32.9	5.0	5.8	0.0	0.7
Singapore	4,848	4,848	65.1	64.5	0.8	0.9	0.0	0.1
South Africa (9)	20,807	20,807	27.6	26.5	4.1	5.5	0.1	4.3
Sweden	3,974	3,974	50.2	48.1	5.1	6.9	0.2	1.9
Turkey	4,077	4,077	46.4	45.2	4.2	5.2	0.0	0.5
United Arab Emirates	22,322	22,324	40.5	38.7	3.1	4.2	0.0	0.9
United States	8,686	8,686	50.1	48.3	1.5	2.1	0.2	1.6
International Average	**226,918**	**226,919**	**44.2**	**42.6**	**5.1**	**6.1**	**0.1**	**1.2**
Ontario, Canada	3,767	3,767	47.1	45.5	2.8	3.8	0.1	1.6
Quebec, Canada	3,170	3,170	51.5	49.8	2.2	2.9	0.0	0.5
Moscow City, Russian Fed.	3,783	3,783	57.4	55.9	3.4	4.1	0.0	0.3
Gauteng, RSA (9)	5,629	5,629	30.6	29.4	2.8	3.6	0.1	2.5
Western Cape, RSA (9)	5,339	5,339	33.5	32.2	3.4	4.4	0.0	2.9
Abu Dhabi, UAE	8,197	8,198	34.3	32.1	3.4	4.7	0.0	0.9
Dubai, UAE	5,726	5,726	52.9	51.6	1.9	2.8	0.0	0.8

Appendix 10E: Modifications to the TIMSS 2019 Achievement Data

Grade 4 Mathematics
Items Deleted for All Countries
M02_10B – ME71217B, MP71217B (severe differential item functioning)
M04_10A – ME71135A, MP71135A (severe differential item functioning)
M08_09 – ME71199, M08_08 – MP71199 (severe differential item functioning)
M10_01 – ME71005, MP71005 (severe differential item functioning)
Items Recoded for All Countries
M05_12 – ME51080, MP51080 (20 to 10, 10 to 71, 11 to 72)
M10_11 – ME71189, MP71189 (20 to 10, 10 to 79, 11 to 79)
M12_11 – ME71190, M12_10 – MP71190 (20 to 10, 10 to 70)
M13_02 – ME61254, MP61254 (20 to 10, 10 to 70)
M13_08 – ME61224, MP61224 (70 to 12)
M14_09 – ME71177, MP71177 (20 to 10, 10 to 70)
Items Deleted by Country
Chile M05_12 – ME51080 (poor discrimination)
Croatia M12_03 – ME71062 (negative discrimination)
Hungary M12_03 – ME71062 (negative discrimination)
Japan M08_12 – MP71194 (translation error) M12_04B – MP71216B (translation error) M12_10 – MP71202 (translation error)
Korea M11_01 – ME61178, MP61178 (severe item-by-country interaction)
Netherlands M10_08 – ME71179 (derived item, poor discrimination)

Items beginning with “ME” are eTIMSS items. Items beginning with “MP” are paperTIMSS items, or bridge items. paperTIMSS trend items deleted or recoded for all countries were also modified for eTIMSS bridge samples.

Grade 4 Mathematics – Less Difficult
Items Deleted for All Countries
MP02_10B – MP71217B (severe differential item functioning)
MP08_08 – MP71199 (severe differential item functioning)
Items Recoded for All Countries
MP13_02 – MP61254 (20 to 10, 10 to 70)
MP13_08 – MP61224 (70 to 12)
Items Deleted by Country
Bosnia and Herzegovina (Cyrillic language only) MP13_03, MP61244 (translation error)
Morocco MN11_09 – MN11158 (poor reliability)
Saudi Arabia MP13_01 – MP61240 (derived item, translation error)

Items beginning with “MP” are items shared with the regular fourth grade mathematics assessment. Items beginning with “MN” are items unique to less difficult mathematics.

Grade 4 Science
Items Deleted for All Countries
S05_02 – SE51020, SP51020 (poor discrimination)
S06_05 – SE61166, SP61166 (poor discrimination)
S07_03C – SE51138C, SP51138C (poor discrimination)
S08_01 – SE71091, SP71091 (severe differential item functioning)
S10_11 – SP71921 (poor discrimination)
S12_09 – SE71910, SP71910 (severe differential item functioning)
S13_01 – SE61125, SP61125 (poor discrimination)
S14_03 – SE71021, SP71021 (severe differential item functioning)
Items Recoded for All Countries
S12_01 – SE71031, SP71031 (11 to 70)
S13_02 – SE61014, SP61014 (20 to 10, 10 to 70)
Items Deleted by Country
Azerbaijan (Azerbaijani language only) S10_06 – SP71080 (translation error)
Bosnia and Herzegovina (Serbian language only) S04_08 – SP71102 (translation error)
Chile S10_11 – SE71921 (negative discrimination)
France S07_02 – SE51051 (negative discrimination)
Georgia S13_03 – SE61056 (poor discrimination)
Germany S01_03A – SE51132A, SP51132A (translation error)
Kosovo S03_03 – SP61054 (poor reliability) S13_02 – SP61014 (poor reliability)
Morocco S03_03 – SP61054 (poor reliability)
Netherlands S01_06 – SP51063 (negative discrimination)
Portugal S10_11 – SE71921 (negative discrimination)
Russian Federation S07_07 – SE51200 (severe item-by-country interaction)
Saudi Arabia S11_06 – SP61093 (poor reliability) S13_11 – SP61124 (derived item, poor discrimination)

Items beginning with "SE" are eTIMSS items. Items beginning with "SP" are paperTIMSS items, or bridge items. paperTIMSS trend items deleted or recoded for all countries were also modified for eTIMSS bridge samples.

Grade 8 Mathematics
Items Deleted for All Countries
M06_07 – ME62342, MP62342 (poor discrimination)
M09_12B – ME62345B, MP62345B (derived item, poor discrimination)
M10_03 – ME72038, MP72038 (severe differential item functioning)
M12_14B – ME72211B, MP72211B (severe differential item functioning)
M13_12 – ME62048, MP62048 (derived item, poor discrimination)
Items Recoded for All Countries
M03_13 – ME62254, MP62254 (20 to 10)
M07_08 – ME52087, MP52087 (20 to 10, 10 to 70)
M08_09B – ME72128B, MP72128B (10 to 20, 70 to 10)
Items Deleted by Country
Georgia M02_03 – ME72017 (poor discrimination) M03_06 – MP62351 (negative discrimination)
Kazakhstan (Kazakh language only) M10_15 – MP72206 (translation error) M12_09A – MP72110A (translation error) M12_09B – MP72110B (translation error) M12_12 – MP72229 (translation error)
Lebanon M09_07 – MP62350 (negative discrimination)
Saudi Arabia M03_02 – MP62139 (translation error) M05_12 – MP52502 (translation error)

Items beginning with “ME” are eTIMSS items. Items beginning with “MP” are paperTIMSS items, or bridge items. paperTIMSS trend items deleted or recoded for all countries were also modified for eTIMSS bridge samples.

Grade 8 Science
Items Deleted for All Countries
S01_06 – SE52134 (severe differential item functioning)
S03_12 – SE62272, SP62272 (poor discrimination)
S04_02 – SP72403 (severe differential item functioning)
S05_11 – SE52221, SP52221 (poor discrimination)
S08_09 – SE72133, SP72133 (severe differential item functioning)
S10_07 – SE72048, SP72048 (severe differential item functioning)
S11_12 – SE62036, SP62036 (attractive distracter)
S11_15C – SE62242C, SP62242C (poor discrimination)
S12_04 – SE72906, SP72906 (derived item, severe differential item functioning)
S12_15 – SE72329, SP72329 (severe differential item functioning)
S13_05 – SE62266, SP62266 (attractive distracter)
Items Recoded for All Countries
S12_09 – SE72523, SP72523 (10 to 20, 11 to 10)
S12_13A – SE72280A, SP72280A (20 to 10, 10 to 70)
Items Deleted by Country
England S09_03 – SE62106 (translation error)
Egypt S10_16 – SP72720 (negative discrimination)
Iran, Islamic Rep. of S05_05 – SP52248 (negative discrimination)
Japan S10_09 – SP72116 (translation error) S14_16 – SP72303 (translation error)
Jordan S01_06 – SP52134 (negative discrimination)
Morocco S01_06 – SP52134 (negative discrimination) S10_14 – SP72220 (negative discrimination)
Saudi Arabia S02_06 – SP72103 (printing error) S03_04 – SP62225 (item not administered) S04_08B – SP72141B (low reliability) S06_04A – SP62098A (low reliability)
South Africa, including Gauteng and Western Cape S06_05 – SP62032 (poor discrimination)

Items beginning with “SE” are eTIMSS items. Items beginning with “SP” are paperTIMSS items, or bridge items. paperTIMSS trend items deleted or recoded for all countries were also modified for eTIMSS bridge samples.

Appendix 10F: Derived Items in TIMSS 2019

Grade 4 Mathematics
M01_01 – ME51043: Item parts A, B, C, D, E, F, G, and H are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M01_05 – ME51508: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct
M02_03 – ME71167: Item parts A, B, C, D, E, and F are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M02_05 – ME71162, MP71162: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both parts are correct and 1 score point is awarded if 1 part is correct
M02_06 – ME71078: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M02_08 – ME71151, MP71151: Item parts A, B, and C are combined to create a 2-point item, where 2 score points are awarded if all parts are correct and 1 score point is awarded if 2 parts are correct
M02_11 – ME71142: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct
M02_12 – ME71204, MP71024: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M04_03 – ME71036, MP71036: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct
M04_09 – ME71178, MP71178: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M04_12 – ME71175, MP71175: Item parts A, B, and C are combined to create a 2-point item, where 2 score points are awarded if all parts are correct and 1 score point is awarded if 1 or 2 are correct
M06_01 – ME61018, MP61018: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M06_10 – ME61266: Item parts A, B, C, D, E, and F are combined to create a 2-point item, where 2 score points are awarded if all parts are correct and 1 score point is awarded if 5 parts are correct
M08_11 – ME71141, M08_10 – MP71141: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M08_12 – ME71194: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct
M08_13 – ME71193, M08_12 – MP71193: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both are correct and 1 score point is awarded if 1 part is correct
M10_05 – ME71213: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M10_08 – ME71179, MP71179: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M10_12A – ME71187A: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

Grade 4 Mathematics (continued)
M11_08 – ME61095: Item parts B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct (part A is an example)
M12_04A – ME71216A, MP71216A: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct
M12_05 – ME71117: Item parts A, B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M12_10 – ME71202: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M13_01 – ME61240, MP61240: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M13_02 – ME61254: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M14_11A – ME71138A, MP71128A: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M14_13 – ME71205, MP71205: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

Items beginning with “ME” are eTIMSS items. Items beginning with “MP” are paperTIMSS items, or bridge items. paperTIMSS trend items deleted or recoded for all countries were also derived for eTIMSS bridge samples.

Grade 4 Mathematics – Less Difficult
MN04_14 – MN21003: Item parts A, B, C, and D are combined to create a 2-point item, where 2 score points are awarded if all parts correct and 1 score point is awarded if 3 parts are correct
MN14_10 – MN21057: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
MP02_05 – MP71162: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both parts are correct and 1 score point is awarded if 1 part is correct
MP02_08 – MP71151: Item parts A, B, and C are combined to create a 2-point item, where 2 score points are awarded if all parts are correct and 1 score point is awarded if 2 parts are correct
MP02_12 – MP71024: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
MP08_10 – MP71141: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
MP08_12 – MP71193: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both are correct and 1 score point is awarded if 1 part is correct
MP13_01 – MP61240: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

Items beginning with “MP” are items shared with the regular fourth grade mathematics assessment. Items beginning with “MN” are items unique to less difficult mathematics.

Grade 4 Science
S02_03 – SE71017, SP71017: Item parts A, B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S04_02 – SE71902, SP71902: Item parts B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct (part A is an example)
S04_04 – SE71041, SP71041: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both parts are correct and 1 score point is awarded if 1 part is correct
S04_05 – SE71046, SP71046: Item parts A, B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S05_10 – SE51151: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S06_06 – SE61083, SP61083: Item parts B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct (part A is an example)
S06_09A – SE61142A, SP61142A: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S07_03 – SE51138Z, SP51138Z: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct (part C was deleted)
S09_08 – SE61160: Item parts B, C, D, E, and F are combined to create a 1-point item, where 1 score point is awarded if all parts are correct (part A is an example)
S10_01 – SE71009, SP71009: Item parts A, B, C, D, E, and F are combined to create a 2-point item, where 2 score points are awarded if all parts are correct and 1 score point is awarded if 4 or 5 parts are correct
S10_09 – SE71106, SP71006: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S10_13 – SE71254: Item parts A, B, C, D, E, F, G, and H are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S13_11 – SE61124, SP61124: Item parts B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct (part A is an example)
S13_12 – SE61116, SP61116: Item parts B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct (part A is an example)
S14_01 – SE71063: Item parts A, B, C, D, E, and F are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S14_08 – SE71114: Item parts A, B, C, D, E, and F are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

Items beginning with “SE” are eTIMSS items. Items beginning with “SP” are paperTIMSS items, or bridge items. paperTIMSS trend items deleted or recoded for all countries were also derived for eTIMSS bridge samples.

Grade 8 Mathematics
M02_01 – ME72007, MP72007: Item parts A, B, C, D, and E are combined to create a 2-point item, where 2 score points are awarded if all parts are correct and 1 score point is awarded if 4 parts are correct
M02_11 – ME72180: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M02_12 – ME72198, MP72198: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct
M02_14 – ME72170: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M03_10 – ME62244: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct
M04_01 – ME72178: Item parts A, B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M04_03 – ME72020: Item parts A, B, C, and D are combined to create a 2-point item, where 2 score points are awarded if all parts are correct and 1 score point is awarded if 3 parts are correct
M04_05 – ME72052, MP72052: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct
M04_11 – ME72164, MP72164: Item parts A, B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M05_12 – ME52502: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M06_10 – ME62288: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both parts are correct and 1 score point is awarded if 1 part is correct
M07_08 – ME52087: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both parts are correct and 1 score point is awarded if 1 part is correct
M08_04 – ME72055: Item parts A, B, C, D, E, and F are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M09_06 – ME62317: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M09_12A – ME62345A: Item parts A, B, C, and D are combined to create a 2-point item, where 2 score points are awarded if all parts are correct and 1 score point is awarded if 2 or 3 parts are correct
M10_09 – ME72095, MP72095: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct,
M10_14 – ME72232, MP72232: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M11_03 – ME62215: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both are correct and 1 score point is awarded if 1 part is correct
M12_08 – ME72225, MP72225: Item parts A and B are combined to create a 1-point item, where 1 score point is awarded if both parts are correct
M13_09 – ME62170: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both parts are correct and 1 score point is awarded if 1 part is correct

Grade 8 Mathematics (continued)
M14_09 – ME72081: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
M14_10 – ME72140, MP72140: Item parts A, B, C, D, E, and F are combined to create a 1-point item, where 1 score point is awarded if all parts are correct

Items beginning with "ME" are eTIMSS items. Items beginning with "MP" are paperTIMSS items, or bridge items. paperTIMSS trend items deleted or recoded for all countries were also derived for eTIMSS bridge samples.

Grade 8 Science
S01_05 – SE52095Z, SP52095Z: Item parts B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct (part A is an example)
S02_08 – SE72130, SP72130: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S02_11 – SE72232, SP72232: Item parts A, B, C, D, E, and F are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S04_02 – SE72403: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S04_13 – SE72345, SP72345: Item parts A, B, C, D, E, F, and G are combined to create a 2-point item, where 2 score points are awarded if all parts are correct and 1 score point is awarded if 5 or 6 parts are correct
S06_13A – SE62173A, SP62173A: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S07_05 – SE52015Z, SP52015Z: Item parts A, B, C, D, E, and F are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S08_02 – SE72400: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S08_13 – SE72260: Item parts A, B, C, D, E, F, and G are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S08_14 – SE72265, SP72265: Item parts A, B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S09_08 – SE62018, SP62018: Item parts A, B, C, D, and E are combined to create a 2-point item, where 2 score points are awarded if all parts are correct and 1 score point is awarded if 4 parts are correct
S10_01 – SE72033: Item parts A, B, C, D, and E are combined to create a 2-point item, where 2 score points are awarded if all parts are correct and 1 score point is awarded if 4 parts are correct
S10_05 – SE72086: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S10_13 – SE72261, SP72261: Item parts A, B, C, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S11_06 – SE62006: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S11_15 – SE62242, SP62242: Item parts A, B, D, and E are combined to create a 1-point item, where 1 score point is awarded if all parts are correct (part C was deleted)

Grade 8 Science (continued)
S12_03 – SE72000: Item parts A, B, C, D, and E are combined to create a 2-point item, where 2 score points are awarded if all parts are correct and 1 score point is awarded if 4 parts are correct
S12_08 – SE72143: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S13_04 – SE62101: Item parts A, B, C, and D are combined to create a 2-point item, where 2 score points are awarded if all parts are correct and 1 score point is awarded if 2 or 3 parts are correct
S13_07 – SE62047, SP62047: Item parts A, B, and C are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S13_08 – SE62042: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S13_14 – SE62022, SP62022: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S13_15 – SE62243: Item parts A, B, C, and D are combined to create a 2-point item, where 2 score points are awarded if all parts are correct and 1 score point is awarded if 2 or 3 parts are correct
S14_02 – SE72905: Item parts A, B, C, and D are combined to create a 1-point item, where 1 score point is awarded if all parts are correct
S14_04 – SE72016, SP72016: Item parts A and B are combined to create a 2-point item, where 2 score points are awarded if both are correct and 1 score point is awarded if 1 part is correct

Items beginning with "SE" are eTIMSS items. Items beginning with "SP" are paperTIMSS items, or bridge items. paperTIMSS trend items deleted or recoded for all countries were also derived for eTIMSS bridge samples.

IEA TIMSS 2019

CHAPTER 11

TIMSS 2019 Scaling Methodology: Item Response Theory, Population Models, and Linking Across Modes

Matthias von Davier

Introduction

This chapter[1] describes the statistical and psychometric approaches underlying the analysis of the TIMSS 2019 data. The first part of the chapter reviews Item Response Theory (IRT), a methodology frequently used in educational measurement that is also increasingly common in other applications of quantitative analysis of human response data such as patient reported outcomes, consumer choice, and other domains. Building on these foundations, the challenges introduced by a hybrid assessment database consisting of both computer-based and paper-based country data are addressed. In TIMSS 2019, half of the countries administered the computer-based version of TIMSS (known as eTIMSS) while the other half continued to assess the students using the paper-based version (paperTIMSS).

The second part of the chapter describes an extension of IRT that allows controlling for mode of administration effects on student performance and that produces a latent variable scale representing student proficiency that is comparable across paper- and computer-based assessment.

The third part of this chapter reviews the integration of achievement data from the TIMSS 2019 mathematics and science items with contextual data from student questionnaires (and parent questionnaires at the fourth grade), and describes the statistical imputation model used for this purpose. This model is a combination of IRT approaches and a regression-based approach that utilizes the context data as predictors for the derivation of a prior distribution of proficiency, and is essentially the approach adopted by TIMSS since the first assessment in 1995. All three parts provide references and information for further reading as well as information about where in other chapters of this volume these developments are being described in terms of actual application to TIMSS 2019 data.

1 The writeup of the psychometric methods presented in this chapter has many sources and the models presented here were developed by a variety of authors. The presentation as compiled here is focused on TIMSS 2019 and benefited greatly from conversations with, and reviews and proofreading by Michael O. Martin, Pierre Foy, Bethany Fishbein, and Liqun Yin.

Modern Test Theory: Item Response Theory

Item Response Theory (IRT; Lord & Novick, 1968) has become one of the most important tools of educational measurement as it provides a flexible framework for estimating proficiency scores from students' responses to test items. A Google search for the phrase "Item Response Theory" (IRT) produces 1,740,000 hits as of September 15, 2020.

TIMSS has been using IRT from the first round in 1995, initially in the form of the Rasch IRT model (Rasch, 1960; von Davier, 2016) and started to use more general IRT models (Lord & Novick, 1968) for the production of proficiency scores beginning with the 1999 cycle. An overview of recent applications of IRT in IEA studies was given by von Davier, Gonzalez, and Schulz (2020).

One of the major goals and design principles of TIMSS, but also other large-scale surveys of student achievement, is to provide valid comparisons across student populations based on broad coverage of the achievement domain. In mathematics as well as in science, this translates into several hundred achievement items, only a fraction of which can be administered to any one student given the available testing time (72 minutes at fourth grade, 90 minutes at eighth grade). Therefore, TIMSS uses an assessment design based on multi-matrix sampling or balanced incomplete block designs (e.g., Mislevy, Beaton, Kaplan, & Sheehan, 1992). As described in the TIMSS 2019 Assessment Design (Martin, Mullis, & Foy, 2017), these achievement items are arranged in blocks that are then assembled into student booklets (or booklet equivalents for eTIMSS) that contain different (but systematically overlapping) sets of item blocks. Because each student receives only a fraction of the achievement items, statistical and psychometric methods are required to link these different booklets together so that student proficiency can be reported on a comparable numerical scale even though no student sees and answers all tasks.

IRT is particularly well suited to handle such data collection design in which not all students are tested with all items. The assumptions made for enabling IRT methods to handle these types of designs, commonly known as balanced incomplete block designs (e.g. von Davier, Sinharay, Oranje & Beaton, 2006; von Davier & Sinharay, 2013) can be described and tested formally (e.g. Fischer, 1981; Zermelo, 1929).

In terms of mathematical notation used in this chapter, the item response variables on an assessment are denoted by x_i for items $i = 1, \ldots, I$. The set of responses to these items is $(\boldsymbol{x_v}) = (x_{v1}, \ldots, x_{vi})$ for student v. For simplicity, we assume $x_{vi} = 1$ denotes a correct response and $x_{vi} = 0$ denotes an incorrect response.

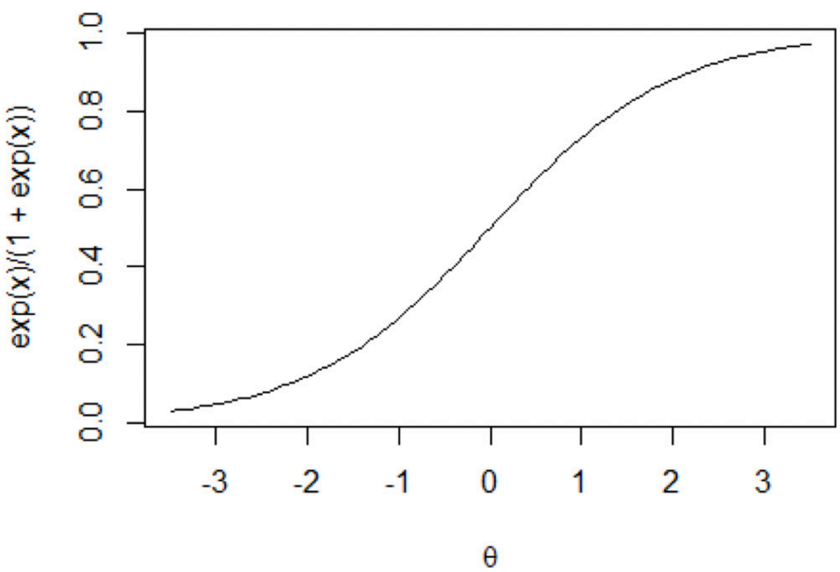
The achievement is assumed to be a function of an underlying latent proficiency variable, often in IRT denoted by θ_v, a real valued variable. Then, we can write

$$P(\boldsymbol{x}_v|\theta_v) = \prod_{i=1}^{I} P(x_{vi}|\theta_v; \boldsymbol{\zeta}_i) \tag{11.1}$$

where $P(x_{vi} \mid \theta_v;\zeta_i)$ represents the probability of an either correct or incorrect response of a respondent with ability θ_v and an item with a certain characteristic $\boldsymbol{\zeta}_i$. In IRT, these item specific effects are referred to as item parameters. Equation (11.1) is a statistical model describing the probability of a set of observed response given ability θ_v. This collective probability is the product of the individual item probabilities.

In TIMSS, the item-level probability model, $P(x_{vi} \mid \theta_v;\boldsymbol{\zeta}_i)$, is given by an IRT model that provides a formal mathematical description, an item function, that describes how the probability of a correct response depends on the ability and the item parameters. One simple approach for an item function is the inverse of the logistic function, also sometimes called the sigmoid function depicted in Exhibit 11.1.

Exhibit 11.1: Sigmoid Function of the Rasch Model

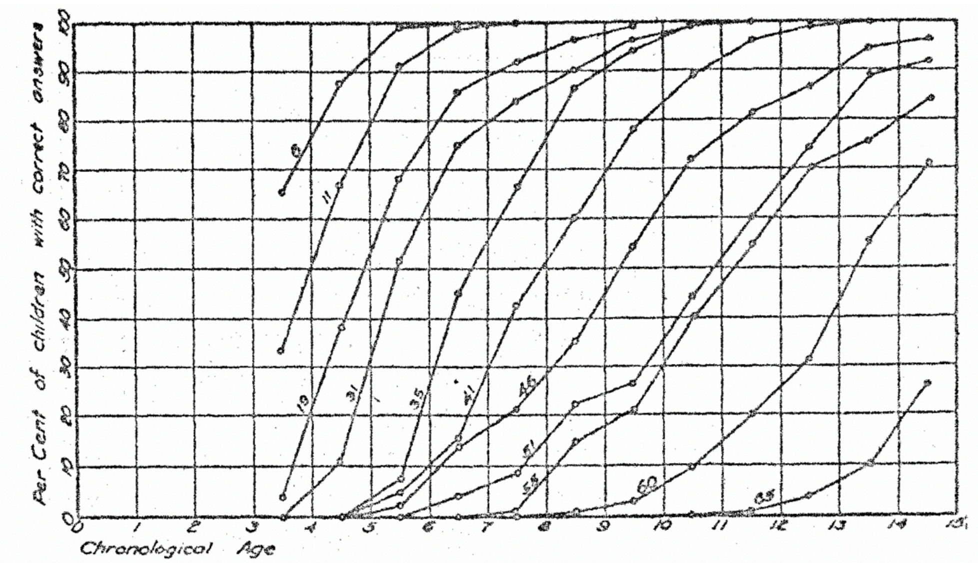


Sigmoid function of the Rasch model $P(x = 1) = exp(T)/[1 + exp(T)]$, *where* $T = a(\theta - b)$ *can be used to linearly adjust for item characteristics.*

Many IRT models used in educational measurement can be understood as relatively straightforward generalizations of the approach shown in Exhibit 11.1. For $a = 1$, where all assessment items contribute equally to the latent construct, this model is called the Rasch model (Rasch, 1960; von Davier, 2016). Why this and other more general approaches of IRT used in TIMSS are suitable choices for modeling assessment data can be seen in the following example.

When looking at test performance by age (a proxy of ability maturation along developmental stages), Thurstone (1925) found that the proportion of respondents who successfully master different tasks is monotonically related to age. Exhibit 11.2 shows this relationship.

Exhibit 11.2: **Relationship between Age and Success on Tasks**



Trace lines obtained by plotting percent correct against age from a series of tasks (Figure 5. in Thurstone, 1925).

The similarity to the sigmoid shown in Exhibit 11.1 is obvious. When, instead of developmental age, the total number of correct responses on a longer test is used, similar graphs are obtained (Lord, 1980). Natural choices for a parametric function that can fit these types of non-linear relationships with a lower and an upper asymptote of zero and one, respectively, are the probit and the logit (e.g., Cramer, 2003).

While the Rasch model specifies a single item parameter b_i in the form of a negative intercept, more general IRT models can be defined that allow for variation of the trace lines in terms of slopes and asymptotes. TIMSS used the Rasch model in 1995, and since 1999 uses the three-parameter logistic (3PL) IRT model (Lord & Novick, 1968) for multiple-choice items, the 2PL IRT model for constructed response items worth 1 score point, and the generalized partial credit model (Muraki, 1992) for constructed response items worth up to 2 score points (Yamamoto & Kulick, 2000).

The 3PL IRT model is given by

$$P(x = 1|\theta_v; \boldsymbol{\zeta}_i) = c_i + (1 - c_i)\frac{exp\left(a_i(\theta_v - b_i)\right)}{1 + exp\left(a_i(\theta_v - b_i)\right)} \tag{11.2}$$

and is a popular choice for binary scored multiple-choice items. In equation (11.2), c_i denotes the pseudo guessing parameter—which, when set to 0.0, yields the 2PL for 1-point constructed response items— b_i denotes the item difficulty parameter, and a_i is the slope parameter.

A model frequently used for binary and polytomous ordinal items (items worth up to 2 points in TIMSS) is the generalized partial credit model (Muraki, 1992), given by

$$P_i(x|\theta_v) = \frac{exp\left(a_i(x\theta_v - b_{ix})\right)}{1 + \sum_{z=1}^{m_i} exp\left(a_i(z\theta_v - b_{iz})\right)} \tag{11.3}$$

assuming a response variable with $m_i + 1$ ordered categories. Very often, the threshold parameters are split into a location and normalized step parameters, $b_{ix} = \delta_i - \tau_{ix}$, with $\sum_x \tau_{ix} = 0$.

The proficiency variable θ_v is sometimes assumed to be normally distributed, that is, $\theta_v \sim N(\mu, \sigma)$. In TIMSS, a normal distribution is used to obtain initial proficiency estimates, as the 3PL model requires constraints of this and other types for identification (Haberman, 2005; San Martín, González, & Tuerlinckx, 2015; von Davier, 2009). Subsequently, this normality constraint can be relaxed and other types of distributions utilized (Haberman, von Davier & Lee, 2008; von Davier & Sinharay, 2013; von Davier et al. 2006; von Davier & Yamamoto, 2004; Xu & von Davier, 2008a).

When there is more than one ability, for example mathematics and science, or content and cognitive process subscales of these, these are represented in a d-dimensional vector $\boldsymbol{\theta}_v = (\theta_{v1}, \ldots, \theta_{vd})$. In this case, one may assume a multivariate normal distribution, $\boldsymbol{\theta}_v \sim N(\boldsymbol{\mu}, \Sigma)$ For the IRT models used in TIMSS, these d-dimensions, examples are main domains or subscales, are assumed to be measured by separate sets of items, so that

$$\boldsymbol{x}_v = \left((x_{v11}, \ldots, x_{vI_1 1}), \ldots, (x_{v1d}, \ldots, x_{vI_d d})\right)$$

represents d sets of I_1 to I_d responses, respectively. A d-dimensional version of the model in (11.1) is given by

$$P(\boldsymbol{x}_v|\boldsymbol{\theta}_v) = \prod_{k=1}^{d}\prod_{i=1}^{I_k} P(x_{vik}|\theta_{vk}; \boldsymbol{\zeta}_{ik}) \tag{11.4}$$

with item-level IRT models (11.2) or (11.3) plugged in for $P(x_{vik} \mid \theta_{vk} ; \boldsymbol{\zeta}_{ik})$ as appropriate. The model given in (11.4) is a multidimensional IRT model for items that show between-item multidimensionality (Adams, Wilson, & Wu, 1997; Adams & Wu, 2007).

Central Assumptions of IRT Models

This section reviews important assumptions of the IRT modeling approach that are central to the types of inferences to be made in TIMSS and other international large-scale assessments. When met, these assumptions allow users of the data to make valid inferences regarding student proficiency in subject domains such as mathematics and science. They ensure that proficiency estimates are comparable across participating countries and over time, and generalizable within the assessment domains described in the framework beyond the limited sample of items each student received.

IRT models describe the probability of a correct response, given examinees proficiency and some item-specific parameters (such as the a_i, b_i described above). This, however, is not how IRT models are actually applied. Not only the item parameters but also the proficiency θ are unknowns that have to be estimated from the data, and all that analysts can rely on is a series of scored answers to a modest number of assessment items. What is needed, and what IRT provides for TIMSS, is a formal model that applies to an assessment domain as a whole, which is delineated in an assessment framework that describes the types of performances on topics viewed as representing the domain. The assumptions underlying IRT facilitate this goal in that they allow inferences about proficiency domains by providing a basis for proficiency estimates that depend on performance on assessment tasks in a well specified and scientifically testable way.

Unidimensionality

TIMSS assesses student achievement on several items students receive. Let I denote the number of items and let the response variables be denoted by $x = (x_1, \ldots, x_I)$. Unidimensionality means that a single quantity is sufficient to describe the probabilities of these responses to each of the items, and that this quantity is the same regardless of the selection of items a student received from within an assessment domain.

Denote P_{iv} and P_{jv} as the probability of person v scoring 1 on items i and j.

$$P_{iv} = P_i(X = 1 \mid \theta_v)$$

and

$$P_{jv} = P_j(X = 1 \mid \theta_v)$$

with the same real valued θ_v in each expression. Unidimensionality ensures that the same underlying proficiency is measured by all the test items in the domain. This of course holds only if the assessment development aims at producing a set of items that are indeed designed to assess the same assessment domain and that test developers diligently refer to the content specifications outlined in the assessment framework. Unidimensionality would (very likely) not hold, for example, if half of the items in a skills test consisted of multiplication problems, and the other half were assessing gross motor skills such as success on a soccer penalty kick practice. As these are two seemingly unrelated skills, one would likely need two proficiency scales: *Multiplication proficiency* and *Penalty kick proficiency*. However, if domains are closely related, requiring for example different mathematical operations such as multiplication and addition, it is typically possible to report these appropriately using only one underlying proficiency variable.

Local Independence and Population Independence

The assumption of population *independence* states that the probabilities of producing a correct response for a given level of proficiency are not dependent on the group to which a test taker belongs. In TIMSS, this independence is important for inferences across countries, but also within countries for inferences across different student groups. Formally population independence holds if

$$P(X_i = x_i \mid \theta, g) = P(X_i = x_i \mid \theta)$$

for any contextual variable g. This also holds for groups defined by performance on x_j on items $j < i$ that precede the current item response x_i. The response to a preceding item can be considered a grouping variable as well, as it splits the sample into those that produced a correct response and those who did not, in the simplest case. Applying the assumption of population independence, this yields

$$P(x_i, x_j \mid \theta) = P(x_i \mid x_j, \theta)P(x_j \mid \theta) = P(x_i \mid \theta)P(x_j \mid \theta). \quad (11.5)$$

The assumption of local independence directly follows. It states that the joint probability of observing a series of responses, given an examinees' proficiency level θ, can be written as the product of the item level probabilities. For a set of responses, local independence takes the form

$$P(\mathbf{X} = x_1, \dots, x_I \mid \theta) = \prod_{i=1}^{I} P_i\,(X = 1 \mid \theta)^{x_i}[1 - P_i(X = 1 \mid \theta)]^{1-x_i}. \quad (11.6)$$

While this assumption appears to be a rather technical one, it can be made more understandable by the following considerations. The proficiency variable intended to be measured is not directly observable, so one can only make inferences about it from observable response behaviors that are assumed to relate to this variable. The assumption of population invariance and local independence facilitates these inferences, in that it is assumed that once a respondent's proficiency level is accounted for, responses become independent from each other, and also from other variables. That is, knowing whether or not a respondent taking a test has answered the previous question correctly does not help predicting the next response, if the respondent's proficiency level θ is known.

According to the assumption of population invariance and local independence, if the model fits the data (and, for example, no learning occurs) and only one single proficiency is 'responsible' for the probability of giving correct responses, then no other variables (including language of the assessment, citizenship, gender, and other contextual variables) are helpful in predicting a respondent's answer to the next item. In this sense, the assumption of local independence and population invariance encapsulate the goal that there is only one variable that needs to be considered, and that estimates of this variable will fully represent the available information about proficiency.

Monotonicity of Item-Proficiency Regressions

One important assumption of IRT models used for achievement data is the (strict) monotonicity of item functions. As seen in Exhibit 11.1, the Rasch model (but also the 2PL and 3PL IRT models) assumes that the probability of a correct response increases with an increasing proficiency. This is represented in the following inequality:

$$P(X_i = 1 \mid \theta_v) < P(X_i = 1 \mid \theta_w) \leftrightarrow \theta_v < \theta_w$$

for all items *i*. This assumption ensures that the proficiency 'orders' the success on the items the students receive, and implies that students with a higher level on the proficiency will also have a higher probability of success on each of the items in the achievement domain. By implication, there is also a strict monotonic relationship between the expected achievement scores and proficiency θ:

$$E(S|\theta_v) = \sum_{i=1}^{I} P(X_i = 1 \mid \theta_v) < E(S|\theta_w) = \sum_{i=1}^{I} P(X_i = 1 \mid \theta_w) \leftrightarrow \theta_v < \theta_w. \quad (11.7)$$

The equation above shows that a person with a greater skill level θ_w compared to a lesser skill level θ_v will in terms of expected score $E(S|\theta_w)$ obtain a larger number of correct responses. This monotonicity ensures that the items and test takers are ordered as one would expect, namely that higher levels on the proficiency are associated with higher expected achievement—a larger expected number of observed correct responses—for any given item or item block measuring the same domain in an assessment booklet.

While the assumptions described above lay the foundation for IRT (and more generally, a large number of latent variable models), each of these assumptions can be relaxed to account for specific attributes of the data collection or assessment design. Models that have been described in this chapter are suitable for achievement data, and the same or variations of these models are used for the analysis of questionnaire data (as described in Chapter 16).

Specialized variants of the IRT models described here are used for reporting on an achievement domain when many different test forms are used, as well as when additional factors have to be accounted for. One such example is the transition from paper- to computer-based assessment. In the context of TIMSS 2019, the move from paper-based to computer-based administration and the need to accommodate both administration modes in estimating student proficiency requires statistically sound extensions of IRT models. The next section describes such psychometric tools that can be applied to enable the transition to computer-based testing.

Accounting for Mode of Administration Effects

The change from paper- to computer-based testing requires careful consideration, as students taking the assessment are faced with different types of response modalities (e.g., a keyboard and mouse or a touchpad or touchscreen, compared to a pencil and a paper sheet to record the answers). This section describes methods for linking the paper-based and the computer-based assessment data, utilizing appropriate extensions of IRT models to establish this link. Chapter 13 of this volume presents country-by-country data based on comparisons of the computer-based eTIMSS 2019 assessments and the paper-based bridge assessments. These comparisons focus on observed item statistics as well as estimates of expected proficiency scores.

Despite the advantages of computer-based assessments, the move from a paper- to a computer-based assessment mode poses challenges for the measurement of trend over time because the results of an assessment administered in different modes may not be directly comparable. One concern is that some assessment items may not function the same across modes and may differ in their difficulty, discrimination, or with respect to the composition of skills they tap into. Mode effects may manifest as differential item functioning (DIF) by (at least) some of the items when comparing equivalent groups across different assessment modes. This, in turn, can affect measurement invariance and may cause undesirable changes in comparability of proficiency scores.

The following section provides an overview of the types of violations of measurement invariance and presents extensions of the IRT models described above that can be used to examine mode effects. The approach presented here was used to select an appropriate adjustment for linking the proficiency scales across modes in TIMSS 2019.

Comparability and Measurement Invariance

There are different levels of measurement invariance (Meredith, 2003; Millsap, 2010) that have to be considered before comparing achievement from different groups or assessments across modes or over time. For valid comparisons, the assessments ideally should exhibit *scalar* or *strong* invariance for all items. This means that the same statistical quantities (IRT item parameters in this context) can be used to fit the items independent of the mode of administration. Weaker forms of invariance are *metric* invariance, where slope parameters are invariant across modes while intercepts are allowed to vary across modes or groups, and finally *configural* invariance, where the same loading pattern can be maintained.

When accounting for mode effects, scalar invariance is the gold standard, while metric invariance is a somewhat less desirable but still a manageable level of invariance as long as proper linking designs can be used to adjust for mode differences (von Davier, Khorramdel, He, Shin, & Chen, 2019). In international assessment, any two cycles are different due to item release and new item development. Therefore, as long as a large proportion of the items reach scalar or metric invariance across modes, it is quite appropriate to have a subset of items with weaker forms of invariance, while most items show strong invariance over time and across modes. Trends measured across modalities are expected to be comparable in order to assess change, and trend measures should provide consistent statistical associations across modes, particularly with external variables central to establishing validity. Ensuring that a large proportion of items show strong invariance properties is crucial for these comparisons. It should be noted that mode effects are just one possible source of violations of measurement invariance. Other sources such as translation errors, technical issues, and language differences are routinely examined and treated as well (e.g. Oliveri & von Davier, 2011; von Davier et al. 2006; von Davier & Sinharay, 2013) in fully paper-based as well as in computer-based assessment.

Assessment Design Requirements for Studying Mode Effects

To deal effectively with mode effects, the assessment design needs to involve items that are *by design comparable*. If only student groups are comparable and take completely different items in paper and computer-based assessments, little can be said about mode differences as items are not comparable. Paper-based assessment items converted for computer delivery so that they can be considered equivalent in terms of content, presentation, and response requirements are referred to here as *by design comparable* items, or *comparable items* for short. About 80 percent of the TIMSS 2019 trend items are in this category and provided a strong link across assessment modes (see Chapter 12: Implementing the TIMSS 2019 Scaling Methodology).

To evaluate the extent to which measurement invariance can be assumed when moving from a paper- to computer-based assessment, an appropriate *data collection design* is needed where the same items are administered in both modes to either the same test takers or equivalent groups of test takers. For operational efficiency, administering the assessment to each student in one mode only is often preferred, while randomly assigning students to modes so that groups taking the assessment in one or other mode are randomly equivalent and results can be compared. In this approach, the two modes of delivery can be understood as treatment assignments in an experiment, while the two randomly assigned (and hence equivalent) groups of students can be assumed to have the same proficiency distribution.

To be able to generalize from such a bridge study, a sufficiently large and representative sample for both modes is needed at the level at which inferences are planned. For example, if the level of inference is how items function in two modes on aggregate at the international level, the two samples must cover the range of abilities that are assessed across countries. On the other hand, if the level of inference is the detection and potential treatment of mode effects at the individual country level, the samples for each country would have to be sufficiently large to enable stable estimates of item parameters at that level.

In TIMSS, this would at least require two large samples of at least the size of the TIMSS national sample (150 schools or more, and 1-2 classrooms of approximately 30 students), one for eTIMSS and the other for paperTIMSS. Because this was not possible due to limited resources being available at the national level, TIMSS 2019 opted for a bridge design to link modes at the international level, with eTIMSS countries selecting a full student sample for eTIMSS together with a smaller, randomly equivalent bridge sample for paperTIMSS. The bridge sample of 1,500 students provided about 375 responses per item per country and was sufficient to evaluate items for mode effects at the international level (i.e., aggregated across all samples). However, these sample sizes are not large enough to provide stable item parameter estimates for individual countries, and hence country-level studies in international contexts require a careful consideration of the limitations of the sample. An example of feasible analyses at the country level is given in Chapter 13 of this volume.

Once the data is collected in both modes, a bridge data set that provides comparable data on the previous mode of assessment, and a new mode data set, statistical analysis and psychometric modeling can commence.

Analysis of Mode Effects Using Graphical Model Checks

As an initial comparison prior to any psychometric modeling approaches or IRT-based analysis, *graphical model checks* (e.g. Khorramdel & von Davier, 2016; Rasch, 1960) can provide important insights. These checks reveal whether the rank order of item parameters and the relations between item parameters agree strongly (as they should) in the eTIMSS and paperTIMSS samples, ensuring that invariance assumptions implemented in subsequent statistical and psychometric modeling are tenable. For this analysis, item parameters for comparable items were estimated separately for the eTIMSS and paperTIMSS samples but pooled across countries to focus on mode comparisons only and to ensure sufficient sample sizes for accurate calibrations.

Exhibits 11.3 through 11.6 show examples of location parameter comparisons between modes using the eTIMSS and paperTIMSS data (including bridge) for each grade and subject assessed by TIMSS. It should be noted that eTIMSS bridge samples responded to trend items only, so comparisons for "new" items are less informative for mode comparisons due to different countries taking the new items in different modes.

Exhibit 11.3: Location Parameter Comparison between eTIMSS and paperTIMSS 2019—Grade 4 Mathematics

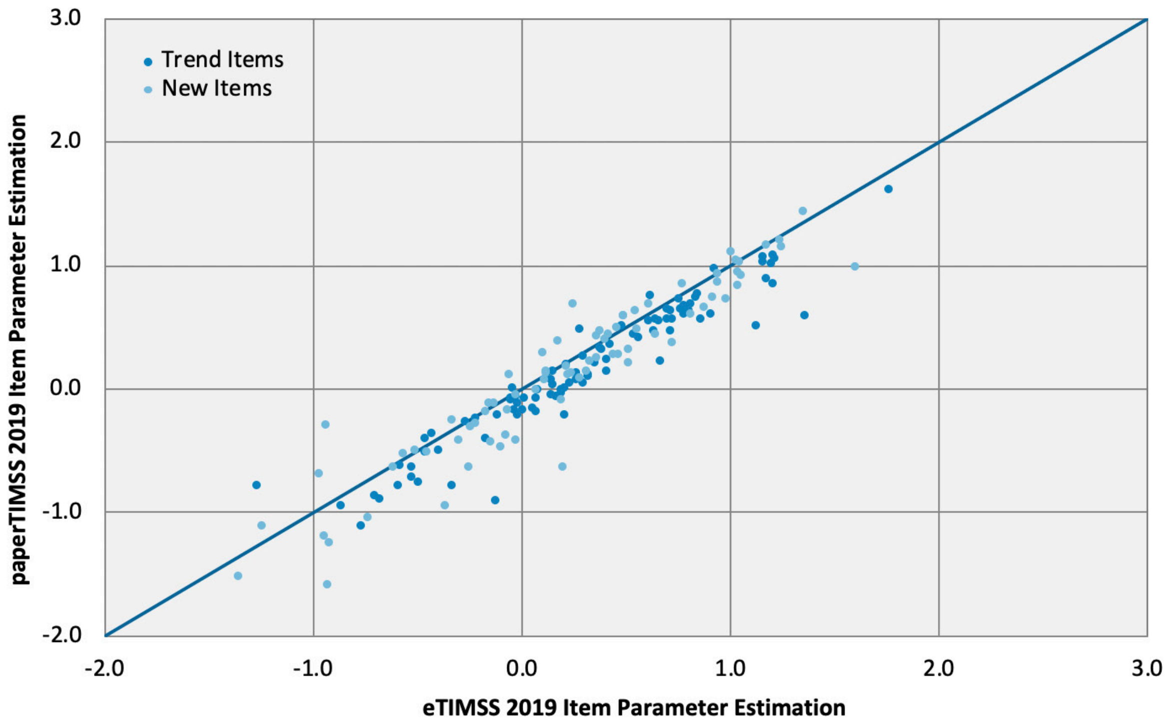


Exhibit 11.4: Location Parameter Comparison between eTIMSS and paperTIMSS 2019—Grade 4 Science

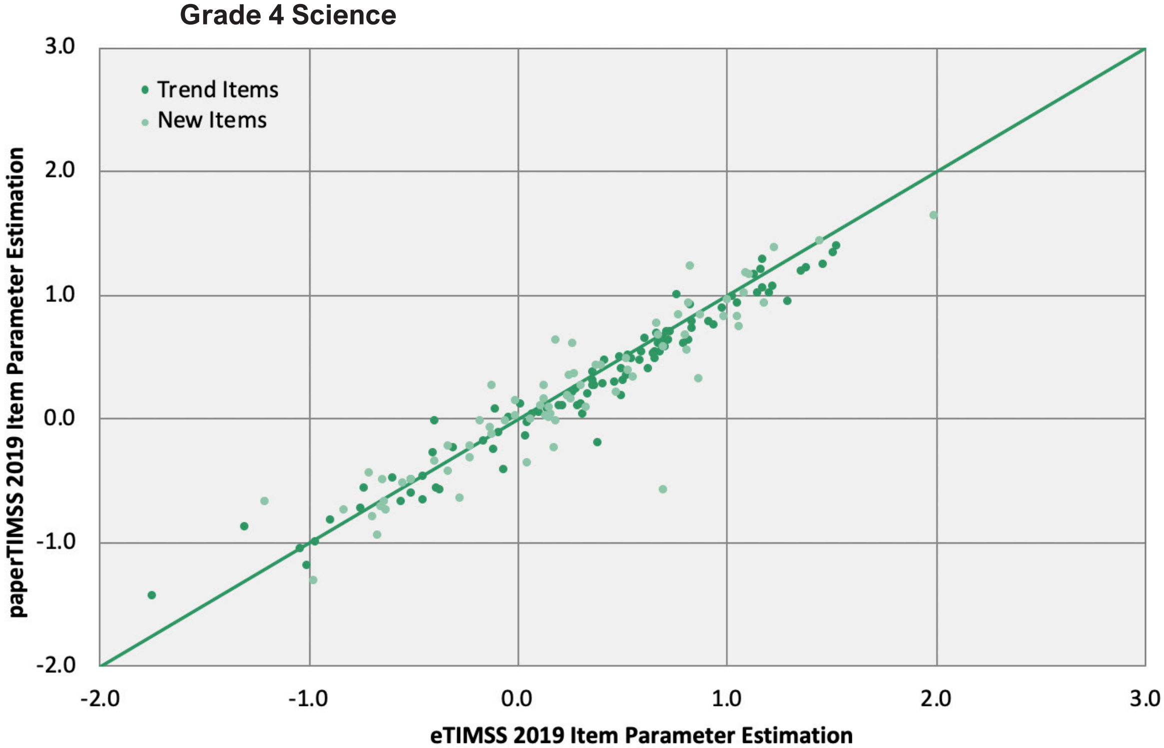


Exhibit 11.5: Location Parameter Comparison between eTIMSS and paperTIMSS 2019—Grade 8 Mathematics

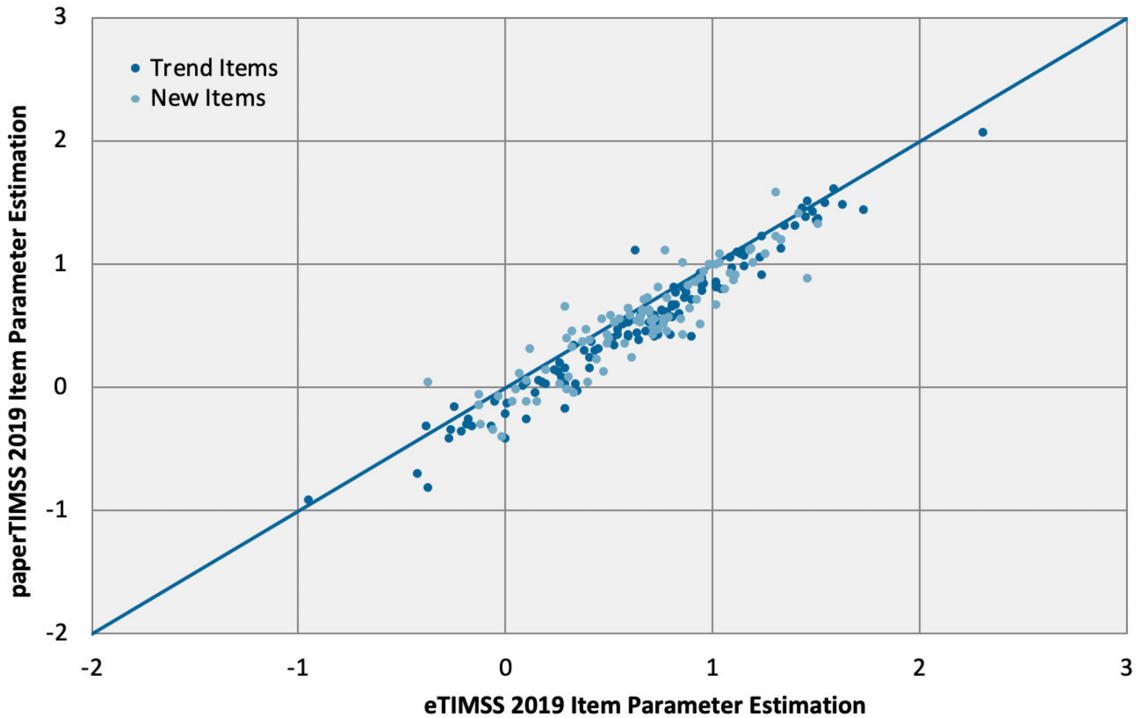
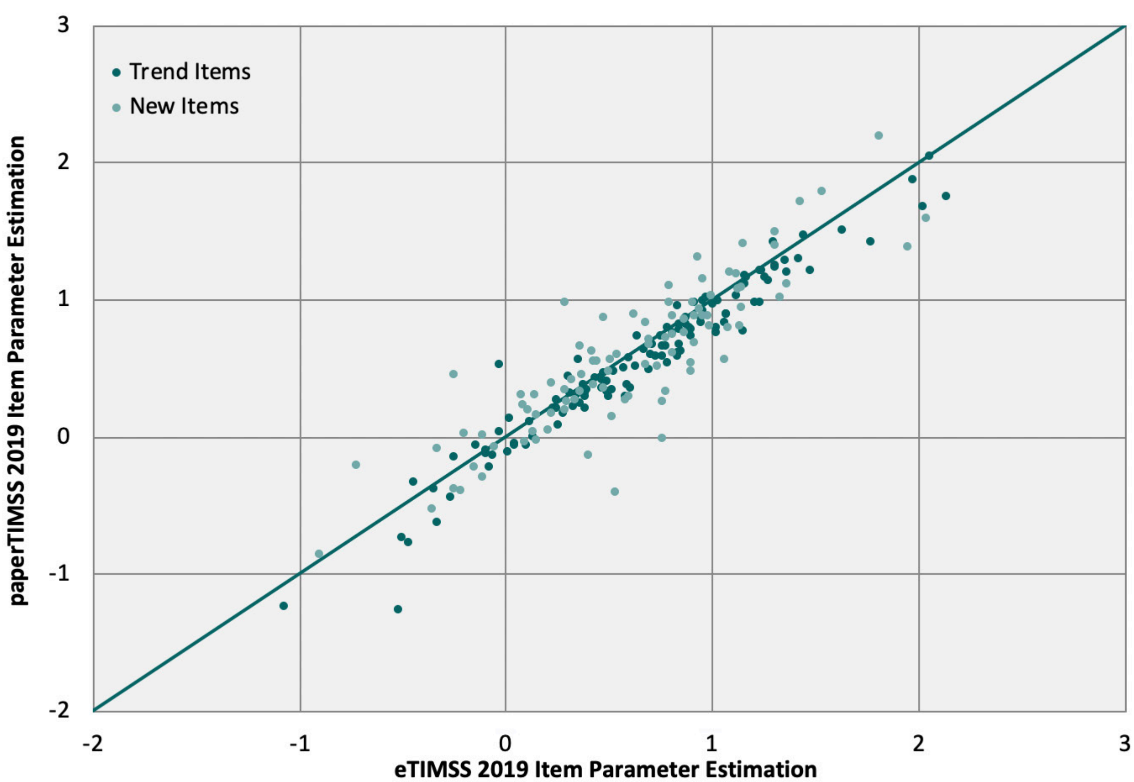


Exhibit 11.6: Location Parameter Comparison between eTIMSS and paperTIMSS 2019—Grade 8 Science



Exhibits 11.3 through 11.6 show that item location parameters (item difficulties) are highly correlated across the eTIMSS and paperTIMSS modes. A similar level of agreement was found for other parameter types. These results suggest there is excellent agreement between paper- and computer-based items for items that were deemed comparable based on design and response similarity across modes. The presence of some outliers, however, suggests that some items differ between modes and may require separate parameter estimates. Note that there is always estimation error in item parameter estimates and, therefore, parameter estimates from two finite samples are never perfectly correlated, even for two independent samples taking the same assessment in the same mode. However, the cross-mode correlations between item parameters of paper- and computer-based items for TIMSS 2019 are very high, suggesting that a strong link can be established so that computer- and paper-based results across countries can be reported on the same scale.

Before such a link can be established, the extent to which some items may exhibit mode effects, and may require separate estimates, has to be carefully examined during IRT scaling. The next section provides an overview of IRT model extensions that facilitate the examination of mode effects and for linking across assessment modes by testing for, and if present, utilizing the invariance of item parameters across modes.

Mode Effect Models

While graphical model checks provide a useful starting point for examining overall agreement between item parameters from different samples and for exploring potential drivers of these differences, they do not provide the most rigorous way to account for mode effects in proficiency estimation (e.g. von Davier & von Davier, 2007). Extensions of IRT models such as the ones described subsequently can be used to analyze mode differences with a high level of statistical rigor in order to obtain unbiased proficiency estimates by utilizing the equivalency of the bridge and eTIMSS samples in the analysis.

IRT models have been extended to include various types of mode effect parameters in order to provide information about whether the mode effect is best described by an overall difference between assessment modes (i.e., the difference between modes is changing the difficulty of all comparable assessment items by a constant), whether it is a person- or group-specific effect that may have an impact differentially on different groups (i.e., some test takers are more affected by mode differences than others), or whether it is an item-specific effect that is only impacting a subset of tasks.

These different hypotheses about mode differences can be checked by formalizing these within a general latent variable model (von Davier, 2008; von Davier, Xu, & Carstensen, 2011) and applying these models to the eTIMSS and bridge data. Taking the two-parameter logistic model (Birnbaum, 1968) as the base model, von Davier et al. (2019) introduced additional model parameters to formalize various assumptions of how mode effects may impact item functioning. Let

$$P(x = 1|\theta, \alpha_i, \beta_i) = \frac{\exp(\alpha_i\theta + \beta_i)}{1 + \exp(\alpha_i\theta + \beta_i)} \tag{11.8}$$

denote the probability of a correct response by a respondent with proficiency θ for an item i with parameters α_i, β_i. The notation used in (11.2) can be transformed to the customary notation by letting $a = \alpha / 1.7$ and $b = -\beta / \alpha$.

Mode Effects on the Item Level

The most parsimonious mode effect assumption is that all items show strong invariance and need to be "shifted" by a certain amount with respect to their difficulty when comparing groups taking the assessment in one mode of administration with another. This could be because, for example, reading any item stem or stimulus is generally harder or easier (by the same amount for all items) on the computer, or responding using the keyboard or a mouse is more tedious or simpler than bubbling in a response on an answer sheet. Here, the mode is a "treatment" that changes the apparent average proficiency between groups, which needs to be corrected for using the equivalency of randomly assigned groups taking the bridge and the eTIMSS assessment, respectively. A mode treatment effect of this type that applies homogeneously to all comparable items can be controlled for by adding the same constant to each of the item difficulty parameters. This general mode effect parameter δ_m quantifies how much more difficult (or easy) all the comparable items appear when presented in a mode other than the reference mode (11.9). In terms of standard IRT linking designs, this general mode effect shift is similar to a non-equivalent groups design with anchor test (NEAT). However, the groups were randomly assigned, so the non-equivalence is really caused by the treatment (mode) that has an overall effect, which can be controlled for through the δ_m that reflects treatment differences.

Formally, for items presented in the "new" mode, we assume that

$$P(X = 1|\theta, \alpha_i, \beta_i, \delta_m) = \frac{\exp\left(\alpha_i\theta + \beta_i - 1_{\{I+1,\ldots,2I\}}(i)\delta_m\right)}{1 + \exp\left(\alpha_i\theta + \beta_i - 1_{\{I+1,\ldots,2I\}}(i)\delta_m\right)}. \tag{11.9}$$

This can be thought of as a model for twice the number of items. The indicator function $1_{\{I+1,\ldots,2I\}}(i)$ equals 1 if the item index is in the second half, that is, the range $I + 1, \ldots, 2I$. The first $1, \ldots, I$ items are the paper-based items without mode effect, and the items in the new mode are indexed by $I + 1, \ldots, 2I$. In this notation it is assumed that item i and item $i + I$ are the same but administered in different modes. This leads to a model with $2I$ items (instead of I items for each delivery mode) in which the difficulty parameters for items presented in one mode (say, paper) are assumed to be β_i for $i = 1, \ldots, I$ and the item parameters for the other mode (say, computer) are appended as parameters β_j for $j = I + 1, \ldots, 2I$ and arranged in the same order and constrained to follow $\beta_j = \beta_i - \delta_m$.

In the bridge design, each test taker receives a subset of items from either the paper-based items, indexed by $i = 1, \ldots, I$, or the computer-based items indexed by $i = I + 1, \ldots, 2I$. The two assignments are based on randomly equivalent respondents that only differ in the treatment they received, the mode of assessment. Note that this form of adjustment is equivalent to assuming the item parameters for comparable items to be strongly invariant and adjusting only for the overall mean differences, between bridge and eTIMSS sample. This model can be estimated by assuming one ability distribution across groups assessed in different modes and adding the mode parameter δ_m as an explanatory effect to the items administered in the new mode.

In contrast to the assumptions of a general mode effect parameter, δ_m, one could argue that not all items are affected when moving from paper to computer: Some could be more difficult, some could be at the same difficulty level, and some could even get easier. This leads to a model with weaker invariance that adds an item-specific effect δ_{mi} to the difficulty parameter. This can be written as a DIF parameter, quantifying item-specific changes from paper presentation, namely

$$P(X = 1|\theta, \alpha_i, \beta_i, \delta_m) = \frac{\exp\left(\alpha_i\theta + \beta_i - 1_{\{I+1,\ldots,2I\}}(i)\delta_{mi}\right)}{1 + \exp\left(\alpha_i\theta + \beta_i - 1_{\{I+1,\ldots,2I\}}(i)\delta_{mi}\right)}. \tag{11.10}$$

The difference in comparison to the model of metric (or "weak") factorial invariance (Meredith, 1993) is that the computer-based item difficulties relative to the paper-based difficulties are decomposed into two components, that is $\beta_{i+I} = \beta_i - \delta_{mi}$, while continuing to assume that $\alpha_{i+I} = \alpha_i$ for the slope parameters. This decomposition indicates that the difficulties are shifted by some (item or item feature)-dependent amount, the shift being applied to one mode on an item-by-item basis—one that is being considered the reference mode with no shift. Assuming ability equivalence, the average treatment (mode) effect can be assessed by calculating $\frac{1}{I}\sum_i \delta_{mi} = \bar{\delta}_m$. This average effect can be compared against the estimated average effect from model (11.9).

The model in equation (11.10) with constraints across both modes on slope parameters, as well as potential constraints on the DIF parameters, establishes weak (also sometimes called metric) invariance (e.g., Meredith, 1993) IRT model, whereas model (11.9), which TIMSS 2019 was able to use, establishes strong invariance. The average mode effect is equivalent to a shift in group means when the item parameters are invariant in model (11.9), whereas model (11.10) allows individual items to deviate from this average mode shift. The larger the number of constraints of the type $\delta_{mi} = c$ for some constant adjustment can be assumed, the more we approach a model with strong factorial invariance, that adjusts only for overall mode treatment differences. Note that an overall adjustment as used in TIMSS 2019 retains the equality of means and variances of the latent variable in both modes as both groups were randomly assigned to modes but selected from a single population.

Mode Effects on the Respondent or Proficiency Level

For completeness of discourse, if it cannot be assumed that the mode effect is a constant (even if item dependent) shift for all respondents, then an additional proficiency may be required to accurately model response probabilities for the new mode. This leads to a multidimensional model with a second latent variable that is added to the item function for items administered in the new mode. The expression $\alpha_{mi}\vartheta$ in the model (11.11) below indicates that there is a second slope parameter α_{mi} for items ($i = I + 1, \ldots, 2I$) administered in computer mode and that the effect of the mode is person dependent and quantified through a second latent variable ϑ. We obtain

$$P(X = 1|\theta, \alpha_i, \alpha_{mi}, \beta_i, \vartheta) = \frac{\exp\,(\alpha_i\theta + \beta_i - \alpha_{mi}\vartheta)}{1 + \exp\,(\alpha_i\theta + \beta_i - \alpha_{mi}\vartheta)}. \tag{11.11}$$

Note that the common slope parameters, α_i, and item difficulties, β_i, are, as before in models (11.9) and (11.10), equal across modes. However, an additional "mode-slope" parameter α_{mi}, for $i = I + 1, \ldots, 2I$, is estimated, with constant $\alpha_{mi} = 0$ for $i \leq I$ for the reference items that are not affected by mode changes. For the joint distribution $f(\theta, \vartheta)$ one assumes uncorrelated latent variables, $\text{cov}(\theta, \vartheta) = 0$, to ensure identifiability in the bridge design.

In equation (11.11) it is assumed that the effect of the person "mode" variable varies across items, which may be the more plausible variant, but a model with item-invariant effects $\alpha_m\vartheta$ (a Rasch variant of a random mode effect) also is feasible. However, an item-specific model is more likely to provide better model data fit. As in model (11.10), the link between modes can be viewed as increasingly more invariant as more slope parameters can be assumed to be $\alpha_{mi} = 0$ for items in the new mode. Each constraint $\alpha_{mi} = 0$ makes the respective item response functions for items i and $i + I$ identical across modes.

Application of Mode Effect Models to TIMSS 2019

The models presented above were available to accommodate a range of mode effects and item invariances across the two TIMSS assessment modes. However, based on the very good agreement between bridge and eTIMSS sample estimates (see Exhibits 11.3–11.6) of item parameters for the TIMSS 2019 comparable items, it was concluded that only a small overall mode adjustment constant was necessary (see Chapter 12). This adjustment was estimated separately for mathematics and science at the fourth and eighth grades. Additional analyses with standard IRT linking methods (Haebera, 1980; Marco, 1977; von Davier & von Davier 2007; Xu & von Davier, 2008b) were in agreement with the results obtained from model (11.9) as well as with the graphical model checks, so that this convergence of results supported the use of an overall mode adjustment.

Using a single adjustment of parameter for each subject/grade combination based on the randomly equivalent samples from the bridge and eTIMSS samples keeps the scaling methods in line with prior

TIMSS trend scaling methods, and enables country-level mode effect analyses as presented in Chapter 13. The eTIMSS sample and bridge sample were of central importance for linking through model (11.9) because, as randomly equivalent groups with a large set of comparable items as anchors, they form the basis for estimating the adjustment using the proficiency distribution estimates in the two modes.

After establishing the size of the item parameter adjustment required by model (11.9), this adjustment was applied to each of the comparable items in scaling the eTIMSS data, resulting in eTIMSS data on the same scale as the bridge data. The effect of the adjustment was verified in terms of item fit and scaling outcomes using country adjustment compared to the separate scaling of items in equivalent groups designs (see Chapter 12).

The major outcome of the foregoing procedure was that the eTIMSS 2019 proficiency data were successfully linked to the existing TIMSS proficiency scales so that results from the paper- and the computer-based assessments can be directly compared without any further adjustment. Very high levels of comparability of item parameters across the two administration modes were established, so that the mode-adjusted item parameters can be used in the population model described in the following section. This population model is used to generate plausible values for estimation of group level results and to examine the relation between student proficiency and other contextual variables. The strong link of paperTIMSS and eTIMSS across modes based on comparable items and equivalent groups design enabled reporting TIMSS 2019 on the same scale for all participating countries. It also formed the basis of an important and final step that provides the proficiency database by means of a country specific population modeling approach as described in the next section.

Population Models Integrating Achievement Data and Context Information

TIMSS uses a latent regression (or population) model to estimate distributions of proficiencies based on the likelihood function of an IRT model, as introduced in the first section of this chapter, and a latent regression of the proficiency on contextual data (von Davier, Gonzalez, & Mislevy, 2009; von Davier et al., 2006). This approach can be viewed as an imputation model for the unobserved proficiency distribution that aims at obtaining unbiased group-level proficiency distributions. The approach requires the estimation of an IRT measurement model, which provides information about how responses to the assessment items depend on the latent proficiency variable. In addition, the latent regression, which provides information about the extent to which background information is related to achievement, is used to improve estimates by borrowing information through similarities of test takers with respect to context variables and the way these relate to achievement. The population model is estimated separately for each country and in TIMSS 2019 five plausible values (PVs) representing the proficiency variable are drawn from the resulting posterior distribution for each respondent in each cognitive domain. It is important to note that PVs are not individual test scores and should only be used for analyses at the group-level using the procedures described in this report and available, for example, through the IDB analyzer.

Population models are examples of high dimensional imputation models, and utilize a large number of context variables in the latent regression to avoid omission of any useful information collected in the questionnaires (von Davier et al., 2006; von Davier et al., 2009; von Davier & Sinharay, 2013). Prior to estimating the latent regression model, a principal component analysis (PCA) of the student context variables is used to eliminate collinearity by identifying a smaller number of orthogonal predictors that account for most of the variation in the background variables (90% in the case of TIMSS 2019).

In order to fully describe the proficiency estimation procedure, the data from the context questionnaires are combined with the responses obtained from the achievement items. The complete observed data for a person n can be expressed as $d_n = (x_{n1}, \ldots, x_{nI}, g_n, z_{n1}, \ldots, z_{nB})$, where $z_{n1}, \ldots, z_{nB}$ represent the context information; $x_{n1}, \ldots, x_{nI}$ represent the answers to the achievement items, and g_n represents the country or population the respondent was sampled from.

The estimation of student proficiency with IRT models can utilize distributions of proficiency in the population of interest. A population model that incorporates contextual data utilizes this information by specifying a second level model that predicts the distribution of proficiency as a function of contextual variables. The conditional expectation in this model is given by

$$\mu_n = \sum_{b=1}^{B} \beta_{g(n)b} \, z_{nb} + \beta_{g(n)0}. \tag{11.12}$$

This expectation utilizes the available information on how context variables relate to the proficiency. The distribution of proficiency is assumed to be normally distributed around this conditional expectation, namely $\theta_n \sim N(\mu_n, \sigma)$.

Together with the likelihood of the responses expressed by the IRT model, this provides a model for the expected distribution of proficiency given the context data $z_{n1}, \ldots, z_{nB}$ and the responses to the TIMSS items. In other words, the model implements the assumption that the posterior distribution of proficiency depends on the context data as well as on the observed achievement. Given the amount of contextual data is much larger than the number of countries typically participating in an assessment, the added value of using a model that includes contextual information for every test taker is considerable. Therefore, if background variables are selected so that correlations with proficiency are likely, one obtains a distribution around the expected value given in (11.12) that is noticeably more accurate than a country-level distribution of proficiency.

Formally, this approach can be described as a multiple (latent) regression model that regresses the latent proficiency variable on background data collected in context questionnaires. The estimation of the regression is addressed separately within countries. The regression is country specific since it cannot be assumed that context information has the same regression effects across different participating countries. Mothers' highest level of education, for example, is well known as a strong predictor of student performance, but this association can be moderated by other factors at the level of educational systems, so that in some countries it may be stronger than in others.

There are several ways to address the estimation of the latent regression parameters. In TIMSS and other large-scale assessments, the latent trait (proficiency) is determined by the IRT model estimated across countries in a previous step. Then the (latent) regression model is estimated treating the item parameters from the previous IRT estimation as fixed quantities. This ensures that the invariance properties that were determined through IRT estimation and potential mode effect adjustments across countries are applied equally to each national dataset (see for example, Mislevy & Sheehan, 1992; Thomas, 1993; von Davier et al., 2006; von Davier & Sinharay, 2013).

Group-Level Proficiency Distributions and Plausible Values

The goal of the psychometric methods described above is to produce a useful database that contains comparable, valid, and reliable information for reporting student proficiency and for secondary users of the TIMSS assessment data. This information comes in the form of likely proficiency estimates for all respondents given their responses to the assessment items and their answers to the context questionnaires. Integrating the IRT model described in the first part of this chapter with the regression model introduced in the previous section, we can estimate the probability of the responses, conditional on context information, as

$$P_g(\mathbf{x}_n \mid \mathbf{z}_n) = \int_\theta \prod_{i=1}^{I} P_{ig}(x_{ni} \mid \theta)\, \phi\left(\theta; \sum_{b=1}^{B} \beta_{gb}\, z_{nb} + \beta_{g0}, \sigma\right) d\theta. \qquad (11.13)$$

This equation provides the basis for the imputation of proficiency estimates that are commonly known as plausible values (Mislevy, 1991). To allow a more compact notation, we use

$$P_{ig}(x_{ni} \mid \theta) = P_{ig}(X = 1 \mid \theta)^{x_{ni}} \left[1 - P_{ig}(X = 1 \mid \theta)\right]^{1-x_{ni}}.$$

This model enables inferences about the posterior distribution of the proficiency θ, given both the TIMSS assessment items $x_1, \ldots, x_I$ and the context information $z_1, \ldots, z_B$. The posterior distribution of the proficiency given the observed data can be written as

$$P_g(\theta \mid \mathbf{x}_n, \mathbf{z}_n) = \frac{\prod_{i=1}^{I} P_{ig}\,(x_{ni} \mid \theta)\phi\left(\theta; \sum_{b=1}^{B} \beta_{gb}\, z_{nb} + \beta_{g0}, \sigma\right)}{\int_\theta \prod_{i=1}^{I} P_{ig}\,(x_{ni} \mid \theta)\,\phi\left(\theta; \sum_{b=1}^{B} \beta_{gb}\, z_{nb} + \beta_{g0}, \sigma\right) d\theta}. \quad (11.14)$$

An estimate of where a respondent n is most likely located on the proficiency dimension can be obtained by

$$E_g(\theta \mid \mathbf{x}_n, \mathbf{z}_n) = \int_\theta \theta\, P_g(\theta \mid \mathbf{x}_n, \mathbf{z}_n) d\theta. \quad (11.15)$$

The posterior variance, which provides a measure of uncertainty around this expectation, is calculated as follows:

$$V_g(\theta \mid \mathbf{x}_n, \mathbf{z}_n) = E_g(\theta^2 \mid \mathbf{x}_n, \mathbf{z}_n) - \left[E_g(\theta \mid \mathbf{x}_n, \mathbf{z}_n)\right]^2. \quad (11.16)$$

Using these two estimates (the mean and variance) to define the posterior proficiency distribution, it is possible to draw a set of plausible values (Mislevy, 1991) from this distribution for each student. Plausible values are the basis for all reporting of proficiency data in TIMSS, allowing reliable group level comparisons because they are based not only on students' answers to the TIMSS items but also reflect how contextual information is related to achievement.

Note that the correlations between context and proficiency are estimated separately in each country, so that there is no bias or inaccurate attribution that could affect the results. Although the expected value of the country level proficiency is unchanged whether context information is used or not, the advantage of including context information plays out when making group-level comparisons. It can be shown analytically and by simulation (von Davier et al., 2009) that including context information in a population model eliminates bias in group level comparisons using this information, and using country specific population models with context variables ensures there is no bias in country level average proficiency data.

In summary, the plausible values used in TIMSS and other large-scale assessments are random draws from a conditional normal distribution

$$\tilde{\theta}_{ng} \sim N\left(E_g(\theta \mid \mathbf{x}_n, \mathbf{z}_n), \sqrt{V_g(\theta \mid \mathbf{x}_n, \mathbf{z}_n)}\right) \tag{11.17}$$

that depend on response data x_n as well as context information z_n estimated using a group-specific model for each country g. That means two respondents with the same item responses, but different context information will receive a different predicted distribution of their corresponding latent trait. Although this may seem incoherent—and would not be adequate to assign test scores to individual students—it is important to remember that TIMSS and similar assessments are population surveys, not individual assessments, and that it is necessary to include context information in order to achieve unbiased comparisons of population distributions (e.g. Little & Rubin, 1987; Mislevy, 1991; Mislevy & Sheehan, 1992; von Davier et al., 2009). Consequently, plausible values are not and should never be used or treated as individual test scores.

In order to provide a more detailed picture of the analytic methods, this chapter focused on the rationale behind the methodologies used in TIMSS 2019, ranging from IRT, to mode effects, to population modeling for unbiased reporting of group level proficiency distributions. Additional information is available in the chapter on scaling outcomes (Foy, Fishbein, von Davier, & Yin, 2020) and the chapter on examining country-level mode related quantities (von Davier, Foy, Martin, & Mullis, 2020).

References

Adams, R. J., Wilson, M., & Wu, M. (1997). Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics, 22*(1), 47-76.

Adams, R. J., & Wu, M. L. (2007). The mixed-coefficients multinomial logic model: A generalized form the Rasch model. In M. von Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models* (pp. 57-75). New York, NY: Springer Science + Business Media, LLC.

Birnbaum, A. (1968). *On the estimation of mental ability* (Series Report No. 15). Randolph Air Force Base, TX: USAF School of Aviation Medicine.

Cramer, C. (2003). *Advanced quantitative data analysis*. New York, NY: McGraw-Hill.

Fischer, G. H. (1981). On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika, 46*(1), 59-77. Retreived from http://dx.doi.org/10.1007/BF02293919

Foy, P., Fishbein, B., von Davier, M., & Yin, L. (2020). Implementing the TIMSS 2019 scaling methodology. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report.* Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/methods/chapter-12.html

Haberman, S. J. (2005). *Identifiability of parameters in item response models with unconstrained ability distribution* (ETS Research Report Series RR-05-24). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/j.2333-8504.2005.tb02001.x

Haberman, S. J., von Davier, M., & Lee, Y.-H. (2008). *Comparison of multidimensional item response models: Multivariate normal ability distributions versus multivariate polytomous ability distributions.* Educational Testing Service RR-08-45. Princeton, NJ: Educational Testing Service.

Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research, 22,* 144–149.

Khorramdel, L., & von Davier, M. (2016). Item response theory as a framework for test construction. In K. Schweizer & C. Distefano (Eds.), *Principles and methods of test construction: standards and recent advancements* (pp. 52-80). Göttingen, Germany: Hogrefe.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: J. Wiley & Sons.

Lord, F. M. (1980). *Applications of items response theory to practical testing problems*. Hillsdales, NJ: Lawrence Erlbaum Associates.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Marco, G. L. (1977). Item characteristic curves solutions to three intractable testing problems. *Journal of Educational Measurement, 14*, 139–160.

Martin, M. O., Mullis, I. V. S., and Foy, P. (2017). TIMSS 2019 assessment design. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2019 assessment frameworks*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525-543.

Millsap, R. E. (2010). Testing measurement invariance using item response theory in longitudinal data: An introduction. *Child Development Perspectives, 4*, 5–9. doi:10.1111/j.1750-8606.2009.00109

Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika, 56*, 177-196.

Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement, 29*, 133-162.

Mislevy, R. J., & Sheehan, K. M. (1987). Marginal Estimation Procedures. In A. E. Beaton (Ed.), *Implementing the new design: The NAEP 1983–84 technical report* (No. 15-TR-20, pp. 293–360). Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. Applied *Psychological Measurement, 16*(2), 159–176.

Oliveri, M. E., & von Davier, M. (2011). Investigation of model fit and score scale comparability in international assessments. *Psychological Test and Assessment Modeling*, 53(3) 315-333.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Nielsen & Lydiche (Expanded edition, Chicago, University of Chicago Press, 1980).

San Martín, E., González, J., & Tuerlinckx, F. (2015). On the unidentifiability of the fixed-effects 3PL model. *Psychometrika, 80*(2), 450-467.

Thomas, N. (1993). Asymptotic corrections for multivariate posterior moments with factored likelihood functions. *Journal of Computational and Graphical Statistics, 2*, 309–322.

Thurstone, L. L. (1925). A method of psychological and educational tests. *Journal of Educational Psychology, 16*(7), 433-451. https://doi.org/10.1037/h0073357

von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, Vol. 61, No. 2. (November), pp. 287-307. https://doi.org/10.1348/000711007X193957

von Davier, M. (2009). Is there need for the 3PL model? Guess what? *Measurement: Interdisciplinary Research and Perspectives*, 7(2), 110-114, doi:10.1080/15366360903117079

von Davier, M. (2016). The Rasch model. In W. J. van der Linden (Ed.), *Handbook of item response theory* (2nd ed., Vol. 1, pp. 31-48). Boca Raton, FL: CRC Press.

von Davier, M., Foy, P., Martin, M. O., & Mullis, I. V. S. (2020). Examining eTIMSS country differences between eTIMSS data and bridge data: A look at country-level mode of administration effects. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report* (pp. 13.1-13.24). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/methods/chapter-13.html

von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? In M. von Davier & D. Hastedt (Eds.), *IERI Monograph Series: Issues and Methodologies in Large Scale Assessments* (Vol. 2, pp. 9-36). Retrieved from https://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_02_Chapter_01.pdf

von Davier, M., Gonzalez, E., & Schulz, W. (2020). Ensuring validity in international comparisons using state-of-the-art pschometric methodologies. In H. Wagemaker (Ed.), *Reliability and validity of international large-scale assessments*

(Vol. 10, pp. 187-219). International Association for the evaluation of Educational Achievement. https://doi.org/10.1007/978-3-030-53081-5_11

von Davier, M., Khorramdel, L., He, Q., Shin, H. J., & Chen, H. (2019). Developments in psychometric population models for technology-based large-scale assessments: An overview of challenges and opportunities. *Journal of Educational and Behavioral Statistics, 44*(6), 671–705.

von Davier, M., & Sinharay, S. (2013). Analytics in international large-scale assessments: Item Response Theory and population models. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 155-174). Boca Raton, FL: CRC Press.

von Davier, M., Sinharay, S., Oranje, A., & Beaton, A. (2006). Statistical procedures used in the National Assessment of Educational Progress (NAEP): Recent developments and future directions. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics* (Vol. 26: Psychometrics). Amsterdam, Netherlands: Elsevier.

von Davier, M., & von Davier, A. A. (2007). A unified approach to IRT scale linking and scale transformations. Methodology*: European Journal of Research Methods for the Behavioral and Social Sciences*, *3*(3), 115-124.

von Davier, M., Xu, X., & Carstensen, C. H. (2011). Measuring growth in a longitudinal large-scale assessment with a general latent variable model. *Psychometrika, 76,* 318-336.

von Davier, M., & Yamamoto, K. (2004). *A class of models for cognitive diagnosis*. Paper presented at the Fourth Spearman Conference, Philadelphia, PA. Retrieved from https://www.researchgate.net/publication/257822207_A_class_of_models_for_cognitive_diagnosis

Xu, X., & von Davier, M. (2008a). *Fitting the structured general diagnostic model to NAEP data* (ETS Research Report, RR-08-27). Princeton, NJ: Educational Testing Service.

Xu, X., & von Davier, M. (2008b). *Linking with the general diagnostic model* (ETS Research Report, RR-08-08). Princteton, NJ: Educational Testing Service. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.2008.tb02094.x/full

Yamamoto, K., & Kulick, E. (2000). Scaling methodology and procedures for the TIMSS mathematics and science scales. In M. O. Martin, K. D. Gregory, & S. E. Stemler (Eds.), *TIMSS 1999 technical report*. Chestnut Hill, MA: Boston College.

Zermelo, E. (1929). The calculation of tournament results as a maximum-likelihood problem [German]. *Mathematische Zeitschrift, 29*, 436-460.

CHAPTER 12

Implementing the TIMSS 2019 Scaling Methodology

Pierre Foy
Bethany Fishbein
Matthias von Davier
Liqun Yin

Introduction

The TIMSS assessments cover a wide range of topics in mathematics and science at two grade levels. Given this broad coverage, a matrix-sampling booklet design was used such that each student was administered only a subset of the entire TIMSS mathematics and science item pools. Given the complexities of the data collection and the need to describe student achievement on a scale that represents the entirety of the assessment framework, TIMSS relied on Item Response Theory (IRT) scaling to provide accurate measures of student proficiency distributions and trends. In order to provide unbiased estimates of student achievement and its relationship to contextual variables, the TIMSS scaling approach used a latent regression population model (also called a conditioning model) with subsequent multiple imputation to obtain plausible values representing proficiency in mathematics and science for all students. To enhance the reliability of the imputed student scores, the TIMSS latent regression scaling approach used the available student context data in the process. A detailed description of the TIMSS scaling methodology can be found in Chapter 11.

The TIMSS & PIRLS International Study Center, responsible for the development and management of the TIMSS assessments, undertook the psychometric scaling and population modeling of the TIMSS achievement data. The scaling was based largely on a concurrent calibration of the TIMSS 2019 data together with data from the previous TIMSS 2015 cycle for measuring trends from cycle to cycle, which has been implemented successfully in the past. However, with the dual administration mode—paperTIMSS and eTIMSS, the TIMSS scaling approach for 2019 involved additional psychometric analyses so that the 2019 computer-based data were linked to be reported on the same scales as the 2019 paper-based results and trend measurements were maintained from past assessments.

In 2019, TIMSS began the transition to computer-based assessment by introducing a computerized version known as eTIMSS. Half the participating countries in 2019 chose to administer the eTIMSS version, with the other half retaining the traditional paper-based administration—referred to as paperTIMSS. The major challenge in scaling the 2019 data was linking both the eTIMSS and paperTIMSS data on the same scale while maintaining comparability to the previously established TIMSS achievement trend scales.

In 2017, the TIMSS & PIRLS International Study Center conducted an item equivalence study, using a counterbalanced experimental design, to examine whether switching from paper and pencil to computer based administration would likely affect the psychometric properties of the TIMSS mathematics and science achievement items at the fourth and eighth grades (Fishbein, Martin, Mullis, & Foy, 2018). The study showed that, on average, performance was higher on paperTIMSS than on eTIMSS items at both grades, especially for mathematics. Consequently, in expectation of this mode of administration effect in the 2019 main data collection, it was considered prudent to include a bridge component in the data collection plan. This involved eTIMSS countries administering a version of the paperTIMSS assessment (a subset of the achievement items—the "trend" items previously administered in 2015) in addition to the main eTIMSS assessment to randomly selected, equivalent groups of students. This bridge allowed to directly compare and link the psychometric properties of items available in both modes on the basis of equivalent student samples in countries that chose eTIMSS.

Altogether, 64 countries and 8 benchmarking entities participated in TIMSS 2019. Countries participating in either paperTIMSS or eTIMSS had national samples of approximately 150 schools and 4,000 students per grade. To provide bridging data between the paperTIMSS and eTIMSS assessments, eTIMSS trend countries—eTIMSS 2019 countries that also participated in 2015—administered paperTIMSS booklets consisting of trend items to an additional sample of 1,500 students, sampling from the same schools as the full eTIMSS samples to the extent possible. Selected results on country level comparisons of the bridge and the eTIMSS samples are provided in Chapter 13 of this volume.

As an additional option, countries participating at the fourth grade in paperTIMSS that were concerned the regular TIMSS mathematics assessment would be too difficult for their students could choose to administer a "less difficult" mathematics assessment.

Developing eTIMSS and paperTIMSS

As described in the TIMSS 2019 assessment design (Martin, Mullis, & Foy, 2017), each of the four TIMSS 2019 assessments (mathematics and science at fourth and eighth grades) consisted of 14 blocks of achievement items, six of which were developed for first time use in 2019 and eight of which were administered previously in 2015 and re-administered in 2019 (the trend items). Of the eight trend blocks administered in 2015, three also were administered as part of the 2011 assessment.

The development of the six blocks of new items for each TIMSS 2019 assessment followed the content and cognitive domain specifications described in the *TIMSS 2019 Assessment Frameworks* (Mullis & Martin, 2017). While adhering to the framework specifications, the approach was first to develop an eTIMSS version of the items, capitalizing as much as possible on the eTIMSS computer-based environment by including new item types such as drag-and-drop and drop-down menus, and automated scoring through number pad entry. The eTIMSS version of an item was then adapted to the paper-and-pencil environment for its paperTIMSS version, making the paperTIMSS version as similar as possible to the eTIMSS version. The goal was to maximize the comparability of eTIMSS and paperTIMSS by having the two versions of the assessment measure the same mathematics and science constructs using the same items as much as possible, while also capitalizing on the benefits of the computer based environment for eTIMSS.

The eight blocks of trend items for each assessment were developed at a time when paper and pencil was the only mode of administration, and so these existed only in paper format. An eTIMSS version was developed for each of these items, retaining the look and feel of the paper versions as much as possible. This work was conducted as part of the item equivalence study, where it was estimated that about 87 percent of the items appeared fairly equivalent in both versions (Fishbein et al., 2018).

The less difficult mathematics assessment at the fourth grade also consisted of 14 item blocks, eight of which were trend blocks from 2015 and six newly developed. All the less difficult item blocks existed only in paper format. Four of the regular fourth grade mathematics item blocks were shared with the less difficult mathematics assessment, the basis for linking the two assessments. As there was not a less difficult science assessment, countries administering the less difficult mathematics assessment also administered the regular fourth grade science assessment.

Exhibit 12.1 reports the numbers of items from the TIMSS 2019 assessments included for achievement scaling. In addition to newly developed eTIMSS items, the eTIMSS 2019 assessment also included extended mathematics and science assessment tasks called Problem Solving and Inquiry Tasks, or "PSIs," as in Exhibit 12.1. These items were not part of the IRT-based scaling reported here, but will be added to the TIMSS 2019 International Database at a later date. Countries' achievement on the PSIs will be described in a special analysis report to be released in 2021.

Exhibit 12.1: Number of Items in the TIMSS 2019 Assessments

Assessments		Grade 4		Grade 8	
		Mathematics	Science	Mathematics	Science
paperTIMSS		169	168	206	211
Less Difficult TIMSS Mathematics*		177	168	—	—
Bridge		92	95	114	118
eTIMSS	Regular Items	171	169	206	211
	PSI Items	29	19	25	27

* The less difficult TIMSS mathematics assessment shared 46 mathematics items and all 168 science items with paperTIMSS.

Exhibit 12.2 shows the number of participating countries and benchmarking participants in TIMSS 2019, across the various assessments offered. It also indicates the number of trend countries (countries that also participated in 2015) for the concurrent calibration models.

Exhibit 12.2: Number of Countries Participating in TIMSS 2019

		All Countries	Trend Countries	Benchmarking Participants
Grade 4				
paperTIMSS		17	14	0
Less Difficult TIMSS Mathematics		11	3	0
eTIMSS	with Bridge*	28	27	0
	without Bridge	2	0	6
Total		**58**	**44**	**6**
Grade 8				
paperTIMSS		17	14	2
eTIMSS	with Bridge	19	19	0
	without Bridge	3	0	5
Total		**39**	**33**	**7**

* Austria, although not a trend country, opted to administer the Bridge booklets.

The paperTIMSS Assessment

At the eighth grade and for most countries at the fourth grade, the paperTIMSS assessment design replicated the assessment design from the more recent previous TIMSS assessment cycles. The 14 mathematics and 14 science item blocks at each grade were assembled into 14 assessment booklets, with each booklet having two mathematics and two science item blocks and each item block appearing in two booklets.

Countries participating at the fourth grade had the option of administering the less difficult mathematics assessment instead of the regular fourth grade mathematics assessment. The less difficult assessment consisted of four mathematics item blocks shared with regular TIMSS and 10 mathematics item blocks that were developed to be less difficult than the regular TIMSS fourth grade mathematics assessment. These were combined with the regular fourth grade science item blocks, with booklets mimicking the regular paperTIMSS booklets, having two blocks of mathematics items (either less difficult or regular) and two blocks of science items.

At the fourth grade, 17 countries participated in paperTIMSS, with 14 of them having participated in the TIMSS 2015 assessment and considered as trend countries for the concurrent calibration. Additionally, 11 countries participated in the less difficult mathematics assessment, three of which were trend countries. At the eighth grade, 17 countries and two benchmarking participants participated in paperTIMSS, with 14 countries being trend.

The eTIMSS Assessment

The eTIMSS assessment design emulated the paperTIMSS design in the way the item blocks were assembled into student booklets ("block combinations" in eTIMSS parlance) and spiraled across computer-based assessment sessions. There was, however, no less difficult fourth grade mathematics option in eTIMSS. A further difference was that eTIMSS included two blocks of Problem Solving and Inquiry (PSI) tasks and items in both mathematics and science at each grade, with one or two tasks in each block. The PSIs were a new initiative, introduced for the first time in 2019. For data collection, the PSI blocks were assembled in two extra block combinations for each assessment, with two mathematics and two science PSI blocks in each block combination. The two PSI block combinations were included in the normal rotation of eTIMSS block combinations and, while there was no overlap between regular eTIMSS and PSI items (i.e., no student got both eTIMSS and PSI items), the PSI blocks were administered to a randomly equivalent sub-sample of students within selected classes and schools.

Although the newly developed PSI tasks were designed to form an integral part of the mathematics and science assessments, they had no counterpart in paperTIMSS and were not included in the main reporting scales for 2019. Rather, they were included with the eTIMSS 2019 items in a second, separate scaling model and the results reported separately. This allowed for examining how the PSI items fit in with the regular items in psychometric terms and prepared for their full inclusion in TIMSS in the 2023 assessment cycle.

Thirty countries and six benchmarking participants participated in eTIMSS at the fourth grade, and 22 countries and five benchmarking participants at the eighth grade.

The eTIMSS Bridge Booklets

Trend countries participating in eTIMSS 2019—countries that also participated in TIMSS 2015—were required to administer a set of eight assessment booklets consisting entirely of the eight mathematics and eight science paperTIMSS trend blocks at each grade. Six of these booklets were exactly the same as those administered in 2015 and two contained blocks also administered in 2015 but in a different combination. The data from these paper bridge booklets were used to link the eTIMSS assessment to the paperTIMSS assessment and the TIMSS trend scales, relying on equivalent populations between the eTIMSS and bridge samples. They also served to provide countries with valuable data on the behavior of mode effects in their countries (see Chapter 13).

The paper bridge booklets were administered to national samples of 1,500 students drawn to be randomly equivalent to the national eTIMSS samples. Of the 30 eTIMSS countries participating at the fourth grade, 27 trend countries administered the paper bridge booklets. In addition, one non-trend country (Austria) administered bridge booklets to fourth grade students for their own research purposes. At the eighth grade, 19 of the 22 eTIMSS countries administered the paper bridge booklets.

Overview of Scaling the TIMSS 2019 Achievement Data

Scaling and linking the TIMSS 2019 data needed to address two major objectives. First, 2019 results from either paperTIMSS or eTIMSS should measure the same mathematics and science constructs and be reported on the same scales. Second, these 2019 results should maintain trends with past TIMSS assessments. With these two goals in mind, the scaling for each subject and grade was conducted in four major phases.

1. **Scaling the paperTIMSS and Bridge Data**: Relying on the usual TIMSS concurrent calibration approach, data from the paperTIMSS trend countries and bridge data from the eTIMSS trend countries were scaled together with their data from TIMSS 2015 to estimate item parameters for the paperTIMSS and bridge data and to establish the scale transformation required to place these results on the TIMSS trend scales.

2. **Scaling the Fourth Grade Less Difficult Mathematics Data**: Special scale linking approaches were implemented to scale the fourth grade mathematics and science data from the 11 countries that opted to administer this assessment. These methods relied on linkages with the TIMSS 2019 fourth grade assessment and the TIMSS Numeracy 2015 assessment to place these results on the TIMSS trend scales.

3. **Scaling the eTIMSS Data**: Scaling the eTIMSS data was based in large part on its linkage to the bridge data that were collected from equivalent samples. The random assignment of equivalent student groups to the bridge and the eTIMSS assessment was utilized in linking by using an anchor test design. The common set of items was based on a substantial subset of eTIMSS items found to be psychometrically equivalent to their paperTIMSS counterparts.
4. **Scaling the PSI Items**: Although the main reporting results were based on the paperTIMSS and eTIMSS items, it was of great interest to evaluate the introduction of items from the Problem Solving and Inquiry tasks into the TIMSS assessments. This item calibration relied on scaling the PSI items along with the eTIMSS items. This last phase of the TIMSS 2019 achievement scaling will be described in a forthcoming publication to be released in 2021, along with the results.

Each of these phases involved four major tasks: calibrating the achievement items (estimating model parameters for each item), creating principal components from the student questionnaire data for use in conditioning, generating plausible values (proficiency estimates) for mathematics and science, and placing these plausible values on the metrics used to report trend results from previous assessments. The scaling procedures produced plausible values for the mathematics and science scales at both the fourth and eighth grades. In addition, plausible values were produced for the content and cognitive domains of mathematics and science. The IRT models and population models used are described in Chapter 11 of this volume.

Before scaling the achievement data, TIMSS conducted an extensive item-by-item review of descriptive item statistics for all countries to evaluate the quality of the assessment items and to identify any unexpected or problematic item properties based on a review of classical test theory item statistics. This review included analyses of change over time with respect to percent correct and partial credit proportions, omit rates, item discrimination and other classical item statistics for trend items from the 2015 assessment, as well as differences between items common to eTIMSS and the paper bridge booklets. These item review activities are described in Chapter 10.

Treatment of Omitted and Not-Reached Responses

Given the matrix-sampling design used by TIMSS, whereby a student is administered only a sample of the 14 assessment blocks (two mathematics and two science blocks), most item responses are missing by design for each student. However, missing data can also result from a student not answering an item, which can occur when the student does not know the answer, omits the item by mistake, or does not have sufficient time to attempt the item. An item is considered “not reached” when—within part 1 or part 2 of a booklet[1]—the item itself and the item immediately preceding it are not answered, and there are no other items completed in the remainder of that part of the booklet.

1 The TIMSS 2019 assessment booklets, including paperTIMSS, eTIMSS, and bridge, consisted of two parts with a break in between.

Not-reached items were treated differently in estimating item parameters and in generating student plausible values. In estimating the item parameters, items in the assessment booklets that were considered not to have been reached by students were treated as if they were not administered. However, not-reached items always were recoded and treated as incorrect when student plausible values were generated.

This treatment of not-reached items was applied to all scaling procedures. Omitted responses always were treated as incorrect.

Scaling the paperTIMSS and Bridge Data

This first phase of scaling constituted a first and fundamental step in the TIMSS 2019 concurrent calibration with TIMSS 2015 trend data to estimate the item parameters for all paperTIMSS items and determined the scale transformations that placed the TIMSS 2019 paperTIMSS achievement results on the TIMSS trend scales. These same scale transformations also were used when transforming the eTIMSS data based on subsequent calibrations.

The metric of the TIMSS reporting scales for overall mathematics and science at each grade level were originally established in TIMSS 1995 by setting the mean of the national average scores for all countries that participated in TIMSS 1995 to 500 and the standard deviation to 100. To enable measurement of trends over time, achievement data from successive TIMSS assessments were transformed to these same metrics. This was done by concurrently scaling the data from each successive assessment with the data from the previous assessment—a process known as concurrent calibration—and applying linear transformations to place the results from each successive assessment on the same scale as the results from the previous assessment. This procedure enabled TIMSS to measure trends across all seven assessment cycles: 1995, 1999, 2003, 2007, 2011, 2015, and 2019.

The first step in linking the assessments for trend scaling is to estimate (calibrate) the item parameters for the items in the current assessment through a concurrent calibration of the data from the current and previous assessments. In 2019, the TIMSS concurrent calibration consisted of combining TIMSS 2015 and TIMSS 2019 data from the trend countries; the 2019 data included the paperTIMSS data of the paperTIMSS trend countries and the bridge data of the eTIMSS trend countries.

In linking successive assessments, concurrent calibration relies on having a large proportion of trend items—items that are retained from one assessment to the next. The TIMSS assessment consists of 14 mathematics item blocks and 14 science item blocks at each grade. In TIMSS 2019, 6 of the mathematics blocks and 6 of the science blocks consisted of newly developed items. The remaining 8 mathematics blocks and 8 science blocks were carried forward from the TIMSS 2015 assessment and are the basis for linking TIMSS 2019 to the TIMSS achievement scales and maintaining trends over time. Exhibits 12.3 through 12.6 show the number of items present for the paperTIMSS 2019 concurrent calibration by item type and content and cognitive domain for both grades and subjects, respectively.

Exhibit 12.3: Mathematics Items for the paperTIMSS 2019 Concurrent Calibration—Grade 4

Item Type	Points	Items Released in 2015		Items Common in 2015 and 2019		Items Introduced in 2019		Total	
		Items	Points	Items	Points	Items	Points	Items	Points
Multiple Choice	1	45	45	42	42	25	25	112	112
Constructed Response	1	28	28	45	45	45	45	118	118
	2	4	8	5	10	7	14	16	32
Total		**77**	**81**	**92**	**97**	**77**	**84**	**246**	**262**

Items by Content and Cognitive Domains

Mathematics Content Domains	Items Released in 2015		Items Common in 2015 and 2019		Items Introduced in 2019		Total	
	Items	Points	Items	Points	Items	Points	Items	Points
Number	36	38	55	59	28	29	119	126
Measurement and Geometry	28	29	26	27	24	27	78	83
Data	13	14	11	11	25	28	49	53
Mathematics Cognitive Domains	**Items Released in 2015**		**Items Common in 2015 and 2019**		**Items Introduced in 2019**		**Total**	
	Items	**Points**	**Items**	**Points**	**Items**	**Points**	**Items**	**Points**
Knowing	30	31	34	34	24	24	88	89
Applying	32	34	40	42	33	37	105	113
Reasoning	15	16	18	21	20	23	53	60
Total	**77**	**81**	**92**	**97**	**77**	**84**	**246**	**262**

Exhibit 12.4: Science Items for the paperTIMSS 2019 Concurrent Calibration—Grade 4

Item Type	Points	Items Released in 2015		Items Common in 2015 and 2019		Items Introduced in 2019		Total	
		Items	Points	Items	Points	Items	Points	Items	Points
Multiple Choice	1	35	35	47	47	40	40	122	122
Constructed Response	1	30	30	45	45	31	31	106	106
	2	8	16	3	6	2	4	13	26
Total		**73**	**81**	**95**	**98**	**73**	**75**	**241**	**254**

Items by Content and Cognitive Domains

Science Content Domains	Items Released in 2015		Items Common in 2015 and 2019		Items Introduced in 2019		Total	
	Items	Points	Items	Points	Items	Points	Items	Points
Life Science	33	38	41	43	32	34	106	115
Physical Science	26	26	36	37	25	25	87	88
Earth Science	14	17	18	18	16	16	48	51
Science Cognitive Domains	**Items Released in 2015**		**Items Common in 2015 and 2019**		**Items Introduced in 2019**		**Total**	
	Items	**Points**	**Items**	**Points**	**Items**	**Points**	**Items**	**Points**
Knowing	31	34	40	43	29	30	100	107
Applying	28	31	34	34	30	30	92	95
Reasoning	14	16	21	21	14	15	49	52
Total	**73**	**81**	**95**	**98**	**73**	**75**	**241**	**254**

Exhibit 12.5: Mathematics Items for the paperTIMSS 2019 Concurrent Calibration—Grade 8

Item Type	Points	Items Released in 2015		Items Common in 2015 and 2019		Items Introduced in 2019		Total	
		Items	Points	Items	Points	Items	Points	Items	Points
Multiple Choice	1	49	49	62	62	28	28	139	139
Constructed Response	1	40	40	46	46	59	59	145	145
	2	6	12	6	12	5	10	17	34
Total		**95**	**101**	**114**	**120**	**92**	**97**	**301**	**318**

Items by Content and Cognitive Domains

Mathematics Content Domains	Items Released in 2015		Items Common in 2015 and 2019		Items Introduced in 2019		Total	
	Items	Points	Items	Points	Items	Points	Items	Points
Number	28	32	36	37	27	29	91	98
Algebra	31	32	30	31	31	31	92	94
Geometry	18	19	25	28	18	21	61	68
Data and Probability	18	18	23	24	16	16	57	58
Mathematics Cognitive Domains	**Items Released in 2015**		**Items Common in 2015 and 2019**		**Items Introduced in 2019**		**Total**	
	Items	**Points**	**Items**	**Points**	**Items**	**Points**	**Items**	**Points**
Knowing	34	34	35	35	29	31	98	100
Applying	37	41	57	60	39	39	133	140
Reasoning	24	26	22	25	24	27	70	78
Total	**95**	**101**	**114**	**120**	**92**	**97**	**301**	**318**

Exhibit 12.6: Science Items for the paperTIMSS 2019 Concurrent Calibration—Grade 8

Item Type	Points	Items Released in 2015		Items Common in 2015 and 2019		Items Introduced in 2019		Total	
		Items	Points	Items	Points	Items	Points	Items	Points
Multiple Choice	1	47	47	59	59	49	49	155	155
Constructed Response	1	43	43	48	48	33	33	124	124
	2	7	14	11	22	11	22	29	58
Total		**97**	**104**	**118**	**129**	**93**	**104**	**308**	**337**

Items by Content and Cognitive Domains

Science Content Domains	Items Released in 2015		Items Common in 2015 and 2019		Items Introduced in 2019		Total	
	Items	Points	Items	Points	Items	Points	Items	Points
Biology	35	38	39	48	35	40	109	126
Chemistry	18	19	22	23	21	24	61	66
Physics	26	26	30	30	22	24	78	80
Earth Science	18	21	27	28	15	16	60	65
Science Cognitive Domains	**Items Released in 2015**		**Items Common in 2015 and 2019**		**Items Introduced in 2019**		**Total**	
	Items	**Points**	**Items**	**Points**	**Items**	**Points**	**Items**	**Points**
Knowing	31	34	43	48	32	33	106	115
Applying	45	47	44	48	36	44	125	139
Reasoning	21	23	31	33	25	27	77	83
Total	**97**	**104**	**118**	**129**	**93**	**104**	**308**	**337**

In concurrent calibration, item parameters for the current assessment are estimated based on the data from both the current and previous assessments, recognizing that some items (the trend items) are common to both. It is then possible to estimate the latent ability distributions of students in both assessments using the item parameters from the concurrent calibration. The difference between these two distributions is the trend measure between the previous and current assessments, although not yet on the TIMSS scale metric.

After the item calibration and estimation of student proficiency, the next step is to find the linear transformation that transforms the student ability distribution of the previous assessment data under the concurrent calibration to match the student ability distribution of these same data under the calibration that was done in the previous assessment. The final step entails applying this linear transformation to the current assessment data scaled using the concurrent calibration. This places the current assessment data on the TIMSS trend scale.

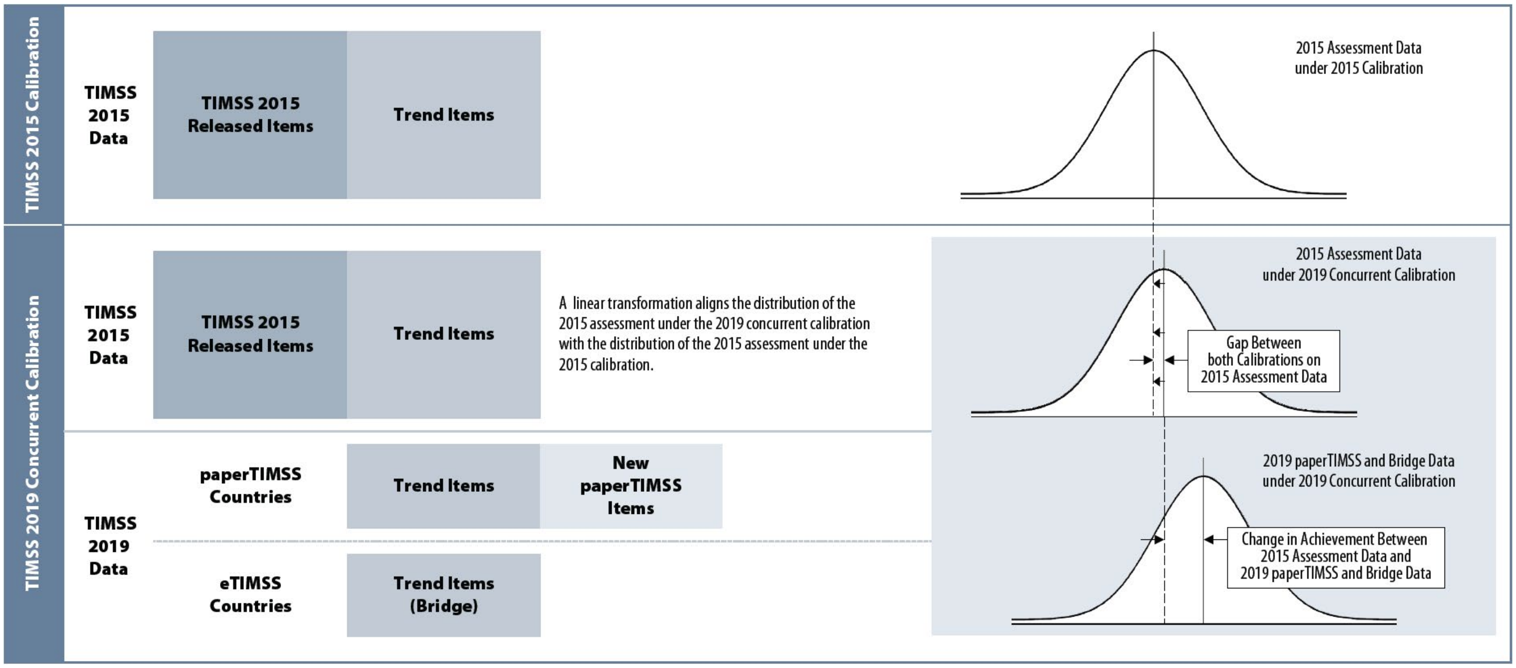
The paperTIMSS 2019 concurrent calibration model provided item parameter estimates for all paperTIMSS items. Using these 2019 item parameters, the TIMSS 2015 ability distribution was re-estimated across all trend countries to find the linear transformation that aligns that re-estimated 2015 student ability distribution with the original student ability distribution that was estimated in 2015. This linear transformation, applied to the 2019 paperTIMSS data and the bridge data, produced achievement results on the TIMSS trend scales. For the paperTIMSS countries, the resulting TIMSS 2019 achievement scores were used for reporting and publishing. The eTIMSS countries with bridge data obtained TIMSS 2019 achievement scores based on their bridge data, which served as a baseline for comparisons with their TIMSS 2019 results based on their eTIMSS data (see Chapter 13).

Calibrating the paperTIMSS and Bridge Data

Item calibration was conducted by the TIMSS & PIRLS International Study Center using the commercially available Parscale software (Muraki & Bock, 1997) and included data from the TIMSS 2015 assessment and data from the TIMSS 2019 assessment, including bridge data, for countries that participated in both assessment cycles. The calibration used all available item response data from each country's student samples and from both the 2019 assessment and the 2015 assessment.

Exhibit 12.7 illustrates the general structure of the paperTIMSS 2019 concurrent calibration model to estimate the paperTIMSS item parameters. The upper panel of the exhibit, labelled "TIMSS 2015 calibration," represents the TIMSS 2015 data from the TIMSS 2019 trend countries and the student ability distribution, shown on the right, which was estimated for this population in the TIMSS 2015 scaling. The lower panel of Exhibit 12.7 is labelled "TIMSS 2019 Concurrent Calibration" and illustrates the full array of paperTIMSS data included in the TIMSS 2019 concurrent calibration model. This included the TIMSS 2015 data from all the TIMSS 2019 trend countries, as well as all 2019 data from paperTIMSS trend countries and bridge data from eTIMSS trend countries.

Exhibit 12.7: **The paperTIMSS 2019 Concurrent Calibration Model**



The 2019 concurrent calibration model included data from three trend countries that participated in the TIMSS 2019 fourth grade less difficult mathematics assessment. Their data contributed to the item parameter estimation for all fourth grade science items, as well as the 46 fourth grade mathematics items shared with the regular fourth grade mathematics assessment.

Exhibits 12.8 and 12.9 show the sample sizes for scaling the paperTIMSS 2019 data, both for item calibration and for proficiency estimation. Countries are shown as being either paperTIMSS (pT), eTIMSS bridge (Br), or less difficult (LD). All student samples were weighted so that each country contributed equally to the item calibration. This was particularly important for the smaller bridge samples of eTIMSS trend countries to ensure their equal contribution in the item calibration.

Exhibit 12.8: Sample Sizes for Scaling the paperTIMSS 2019 Grade 4 Data

Country	Item Calibration		Proficiency Estimation	
	2019	2015	2019	2015
Armenia - pT	5,399	5,384	5,399	5,384
Australia - pT	5,890	6,057	5,890	6,057
Austria - Br	—	—	1,964	—
Azerbaijan - pT	—	—	5,245	—
Bahrain - pT	5,762	4,146	5,762	4,146
Belgium (Flemish) - pT	4,655	5,404	4,655	5,404
Bulgaria - pT	4,268	4,228	4,268	4,228
Canada - Br	1,604	12,283	1,604	12,283
Chile - Br	1,612	4,756	1,612	4,756
Chinese Taipei - Br	1,663	4,291	1,663	4,291
Croatia - Br	1,472	3,985	1,472	3,985
Cyprus - pT	4,062	4,125	4,062	4,125
Czech Republic - Br	2,030	5,202	2,030	5,202
Denmark - Br	1,432	3,710	1,432	3,710
England - Br	1,242	4,006	1,242	4,006
Finland - Br	1,983	5,015	1,983	5,015
France - Br	1,948	4,873	1,948	4,873
Georgia - Br	1,632	3,919	1,632	3,919
Germany - Br	1,505	3,948	1,505	3,948
Hong Kong SAR - Br	1,329	3,600	1,329	3,600
Hungary - Br	1,778	5,036	1,778	5,036
Iran, Islamic Rep. of - pT	6,010	3,823	6,010	3,823
Ireland - pT	4,582	4,344	4,582	4,344
Italy - Br	1,921	4,373	1,921	4,373
Japan - pT	4,196	4,383	4,196	4,383
Kazakhstan - pT	—	—	4,791	—
Korea, Rep. of - Br	1,541	4,669	1,541	4,669
Kuwait - LD	4,437	3,593	—	3,593
Latvia - pT	—	—	4,481	—
Lithuania - Br	1,587	4,529	1,587	4,529
Morocco - LD	7,723	5,068	—	5,068
Netherlands - Br	1,295	4,515	1,295	4,515
New Zealand - pT	5,019	6,322	5,019	6,322
Northern Ireland - pT	3,497	3,116	3,497	3,116
Norway (5) - Br	1,899	4,329	1,899	4,329
Oman - pT	6,814	9,105	6,814	9,105
Poland - pT	4,882	4,747	4,882	4,747
Portugal - Br	1,612	4,693	1,612	4,693
Qatar - Br	1,486	5,194	1,486	5,194
Russian Federation - Br	2,128	4,921	2,128	4,921

Countries are shown as being either paperTIMSS (pT), eTIMSS bridge (Br), or less difficult (LD).

Exhibit 12.8: Sample Sizes for Scaling the paperTIMSS 2019 Grade 4 Data (continued)

Country	Item Calibration		Proficiency Estimation	
	2019	2015	2019	2015
Saudi Arabia - LD	5,453	4,337	—	4,337
Serbia - pT	4,380	4,036	4,380	4,036
Singapore - Br	1,881	6,517	1,881	6,517
Slovak Republic - Br	1,610	5,773	1,610	5,773
Spain - Br	1,670	7,764	1,670	7,764
Sweden - Br	1,697	4,142	1,697	4,142
United Arab Emirates - Br	2,243	21,177	2,243	21,177
United States - Br	1,652	10,029	1,652	10,029
TOTAL	**132,481**	**239,467**	**131,349**	**239,467**

Countries are shown as being either paperTIMSS (pT), eTIMSS bridge (Br), or less difficult (LD).

Exhibit 12.9: **Sample Sizes for Scaling the paperTIMSS 2019 Grade 8 Data**

Country	Item Calibration		Proficiency Estimation	
	2019	2015	2019	2015
Australia - pT	9,060	10,338	9,060	10,338
Bahrain - pT	5,725	4,918	5,725	4,918
Chile - Br	1,526	4,849	1,526	4,849
Chinese Taipei - Br	1,578	5,711	1,578	5,711
Cyprus - pT	—	—	3,521	—
Egypt - pT	7,210	7,822	7,210	7,822
England - Br	1,592	4,814	1,592	4,814
Georgia - Br	1,314	4,035	1,314	4,035
Hong Kong SAR - Br	1,423	4,155	1,423	4,155
Hungary - Br	1,751	4,893	1,751	4,893
Iran, Islamic Rep. of - pT	5,980	6,130	5,980	6,130
Ireland - pT	4,118	4,704	4,118	4,704
Israel - Br	1,863	5,512	1,863	5,512
Italy - Br	2,032	4,481	2,032	4,481
Japan - pT	4,446	4,745	4,446	4,745
Jordan - pT	7,176	7,865	7,176	7,865
Kazakhstan - pT	—	—	4,453	—
Korea, Rep. of - Br	1,693	5,309	1,693	5,309
Kuwait - pT	4,574	4,503	4,574	4,503
Lebanon - pT	4,730	3,873	4,730	3,873
Lithuania - Br	1,687	4,347	1,687	4,347
Malaysia - Br	1,560	9,726	1,560	9,726
Morocco - pT	8,458	13,035	8,458	13,035
New Zealand - pT	6,051	8,142	6,051	8,142
Norway (9) - Br	2,018	4,697	2,018	4,697
Oman - pT	6,751	8,883	6,751	8,883
Qatar - Br	1,490	5,403	1,490	5,403
Romania - pT	—	—	4,494	—
Russian Federation - Br	2,083	4,780	2,083	4,780
Saudi Arabia - pT	5,680	3,759	5,680	3,759
Singapore - Br	1,871	6,116	1,871	6,116
South Africa (9) - pT	20,829	12,514	20,829	12,514
Sweden - Br	1,582	4,090	1,582	4,090
Turkey - Br	1,819	6,079	1,819	6,079
United Arab Emirates - Br	2,089	18,012	2,089	18,012
United States - Br	1,484	10,221	1,484	10,221
Benchmarking Participants				
Gauteng, RSA (9) - pT	—	—	5,633	—
Western Cape, RSA (9) - pT	—	—	5,351	—
TOTAL	**133,243**	**218,461**	**156,695**	**218,461**

Countries are shown as being either paperTIMSS (pT) or eTIMSS bridge (Br).

At the fourth grade, 44 countries contributed to the concurrent calibration, including three countries that participated in the fourth grade less difficult mathematics assessment. These 44 trend countries provided 239,467 students from the 2015 assessment and 132,481 students from the 2019 assessment, including bridge data from eTIMSS trend countries. At the eighth grade, 33 countries contributed to the concurrent calibration, 218,461 students from the 2015 assessment and 133,243 students from the 2019 assessment, including bridge data from eTIMSS trend countries.

The item parameters estimated from these concurrent calibrations, based on the countries that participated in both the previous and current assessments, were used to estimate student proficiency for all countries and benchmarking entities participating in the paperTIMSS 2019 and bridge assessments. These item parameters also were used to estimate student proficiency in the mathematics and science content and cognitive domains. Estimating student proficiency for all eTIMSS countries and benchmarking participants based on their eTIMSS data also relied, to a large extent, on these estimated paperTIMSS item parameters.

At the fourth grade, paperTIMSS and bridge student proficiency was estimated for a total of 45 countries, as shown in Exhibit 12.8. At the eighth grade, student proficiency was estimated for 36 countries and 2 benchmarking participants, as shown in Exhibit 12.9. The item parameters estimated from the paperTIMSS concurrent calibration at the fourth and eighth grades and for mathematics and science are presented in Appendices 12A through 12D.

Variables for Conditioning the paperTIMSS and Bridge Data

Conditioning refers to utilizing a latent regression model that involves all available students' contextual information to improve statistical properties of the estimated student proficiency values. Ideally, all student-level contextual data would be included in the conditioning model, but because TIMSS has so many student context variables that could be used in conditioning, the TIMSS & PIRLS International Study Center follows the practice established by NAEP and followed by other large-scale studies of using principal component analysis to reduce the number of variables while explaining most of their common variance. Principal components for the TIMSS student context variables (including parent context variables at the fourth grade) were constructed as follows:

- For categorical variables (questions with a small number of fixed response options), a dummy coded variable was created for each response option, with a value of one if the option is chosen and zero otherwise. If a student omitted or was not administered a particular question, all dummy coded variables associated with that question were assigned the value zero.
- Context variables with numerous response options (such as year of birth) were recoded using criterion scaling.[2] This was done by replacing the response option with the mean interim achievement score of all students choosing that option. Criterion scaling maximizes

2 The process of generating criterion-scaled variables is described in Beaton (1969).

the correlation between the scaled variable and achievement. For TIMSS, the interim achievement score was the student-level average of the mathematics and science EAP scores produced from the item calibrations.

- Separately for each country, all the dummy-coded and criterion-scaled variables were included in a principal component analysis. Those principal components accounting for 90 percent of the variance of all context variables were retained for use as conditioning variables.[3] Because the principal component analysis was performed separately for each country and benchmarking participant, different numbers of principal components were required to account for 90 percent of the common variance in each country's context variables.

In addition to the principal components, students' gender (dummy coded), the language of the test (dummy coded), an indicator of the classroom in the school to which a student belongs (criterion scaled), and an optional country-specific variable (dummy coded) were included as primary conditioning variables, thereby accounting for most of the variance between students and preserving the between-classroom and within-classroom variance structure in the latent regression conditioning model. Exhibits 12.10 and 12.11 provide details on the conditioning models used for proficiency estimation of the paperTIMSS and bridge data at the fourth and eighth grades, respectively.

3 The number of principal components retained is limited to no more than 5% of a country's student sample size, thereby possibly reducing the percentage of variance accounted for to avoid over-specification of the conditioning model. This constraint played a major role with the eTIMSS bridge samples due to their smaller size.

Exhibit 12.10: Conditioning Models for the paperTIMSS 2019 Grade 4 Data

Country	2019				2015			
	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
Armenia - pT	2	532	269	88	2	615	269	84
Australia - pT	2	323	176	90	2	633	302	89
Austria - Br	2	558	98	61	—	—	—	—
Azerbaijan - pT	3	563	262	85	—	—	—	—
Bahrain - pT	3	562	288	89	3	637	207	75
Belgium (Flemish) - pT	2	557	232	83	2	629	270	84
Bulgaria - pT	2	545	213	82	2	617	211	78
Canada - Br	5	544	80	59	5	619	321	90
Chile - Br	2	538	80	55	2	610	237	80
Chinese Taipei - Br	2	562	83	56	2	636	214	78
Croatia - Br	2	557	73	56	3	637	199	76
Cyprus - pT	3	563	203	78	2	637	206	74
Czech Republic - Br	2	558	101	65	2	636	260	84
Denmark - Br	2	555	71	56	2	628	185	73
England - Br	2	329	62	59	2	336	179	90
Finland - Br	3	560	99	63	3	634	250	83
France - Br	2	562	97	60	2	637	243	81
Georgia - Br	2	559	81	56	2	637	195	74
Germany - Br	2	563	75	59	2	637	197	76
Hong Kong SAR - Br	3	563	66	54	3	637	180	73
Hungary - Br	2	538	88	58	2	613	251	82
Iran, Islamic Rep. of - pT	2	563	299	90	2	637	191	73
Ireland - pT	3	563	229	83	3	637	217	78
Italy - Br	2	556	96	59	2	631	218	77
Japan - pT	2	552	209	82	2	635	219	79
Kazakhstan - pT	3	562	239	84	—	—	—	—
Korea, Rep. of - Br	2	549	77	57	2	636	233	81
Latvia - pT	3	561	224	82	—	—	—	—
Lithuania - Br	4	548	79	57	4	630	226	79
Netherlands - Br	2	323	64	61	2	619	225	82
New Zealand - pT	6	563	250	87	8	633	314	90
Northern Ireland - pT	2	503	174	79	3	589	155	71
Norway (5) - Br	3	483	94	64	3	636	216	80
Oman - pT	3	563	306	90	3	637	353	90
Poland - pT	2	558	244	85	2	616	237	81
Portugal - Br	2	561	80	54	2	636	234	79
Qatar - Br	3	562	74	56	3	632	259	83
Russian Federation - Br	2	537	106	62	2	613	246	81
Serbia - pT	2	562	219	83	2	628	201	76
Singapore - Br	2	539	94	62	2	637	322	90

Countries are shown as being either paperTIMSS (pT) or eTIMSS bridge (Br).

Exhibit 12.10: Conditioning Models for the paperTIMSS 2019 Grade 4 Data (Continued)

Country	2019				2015			
	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
Slovak Republic - Br	3	563	80	55	3	633	288	86
Spain - Br	6	555	83	57	5	628	319	90
Sweden - Br	2	537	84	60	2	611	207	78
United Arab Emirates - Br	5	563	112	68	5	637	346	90
United States - Br	10	327	82	65	10	330	184	90

Countries are shown as being either paperTIMSS (pT) or eTIMSS bridge (Br).

Exhibit 12.11: Conditioning Models for the paperTIMSS 2019 Grade 8 Data

Country	2019				2015			
	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
Australia - pT	2	539	241	90	2	478	245	90
Bahrain - pT	3	546	255	90	3	482	245	89
Chile - Br	2	545	76	59	2	481	242	89
Chinese Taipei - Br	2	532	78	63	2	481	231	90
Cyprus - pT	3	888	176	67	—	—	—	—
Egypt - pT	3	547	273	90	2	482	276	90
England - Br	2	545	79	61	2	482	240	89
Georgia - Br	2	886	65	49	2	850	201	72
Hong Kong SAR - Br	3	545	71	64	2	482	207	87
Hungary - Br	2	887	87	53	2	850	244	75
Iran, Islamic Rep. of - pT	2	547	257	90	2	482	261	90
Ireland - pT	3	547	205	84	3	482	235	88
Israel - Br	3	481	93	66	3	436	230	90
Italy - Br	2	546	101	65	2	482	224	87
Japan - pT	2	547	222	89	2	480	234	90
Jordan - pT	2	547	259	90	2	482	263	90
Kazakhstan - pT	3	887	222	77	—	—	—	—
Korea, Rep. of - Br	2	533	84	66	2	481	227	90
Kuwait - pT	3	541	228	86	3	474	225	85
Lebanon - pT	3	769	236	78	3	724	193	71
Lithuania - Br	4	881	84	54	4	845	217	73
Malaysia - Br	3	541	78	59	2	473	248	90
Morocco - pT	3	888	422	90	2	850	463	90
New Zealand - pT	7	547	246	90	8	478	245	90
Norway (9) - Br	3	503	100	67	3	482	234	89
Oman - pT	3	547	268	90	3	482	271	90
Qatar - Br	3	547	74	59	3	477	244	90
Romania - pT	2	888	224	74	—	—	—	—
Russian Federation - Br	2	888	104	57	2	849	239	76
Saudi Arabia - pT	3	541	266	90	3	482	187	79
Singapore - Br	2	523	93	65	2	482	246	90
South Africa (9) - pT	5	547	277	90	3	482	276	90
Sweden - Br	2	773	79	58	2	726	204	77
Turkey - Br	2	547	90	61	2	481	257	90
United Arab Emirates - Br	5	547	104	67	5	482	258	90
United States - Br	10	542	74	59	10	475	248	90
Benchmarking Participants								
Gauteng, RSA (9) - pT	3	547	265	90	—	—	—	—
Western Cape, RSA (9) - pT	3	547	262	90	—	—	—	—

Countries are shown as being either paperTIMSS (pT) or eTIMSS bridge (Br).

Generating Plausible values for the paperTIMSS and Bridge Data

Educational Testing Service's MGROUP program (Sheehan, 1985) was used to estimate the latent regression model and generate plausible values. This program takes as input the students' responses to the items they were given, the item parameters estimated at the calibration stage, and the conditioning variables, and generates as output the estimated regression effects and residual variance covariance, as well as plausible values that represent the posterior distribution of student proficiency given achievement and contextual data. A useful feature of MGROUP is its ability to estimate multi-dimensional latent regression models using the responses to all items across the proficiency scales and the correlations among the scales to improve the reliability of each individual scale. TIMSS capitalizes on this feature to estimate simultaneously overall mathematics proficiency and overall science proficiency at each grade using a two-dimensional MGROUP model. More details on the latent regression model are available in Chapter 11 of this volume.

The multi-dimensional scaling feature of MGROUP also was used to generate plausible values for the TIMSS 2019 content and cognitive domains. The estimation of plausible values for the mathematics and science content and cognitive domains relied on multidimensional IRT models using the item parameters estimated for the overall mathematics and overall science scales as well as the same set of conditioning variables. At the fourth grade, the content domain scaling used two four-dimensional models, one to estimate plausible values for the three content domains in mathematics with overall science and a second for the three science content domains with overall mathematics. At the eighth grade, the content domain scaling required two five-dimensional models because of the four content domains in each subject along with the other overall subject. The cognitive domain scaling relied on four four-dimensional models to estimate the three cognitive domains in mathematics and science, along with the other overall subject, at both fourth and eighth grades. All of these models were applied to each paperTIMSS country and benchmarking participant.

In addition to generating plausible values on the overall mathematics and science scales for the 2019 paperTIMSS and bridge data, the item parameters estimated at the calibration stage also were used to generate plausible values for the TIMSS 2015 assessment data for the countries included in the concurrent calibration at the fourth and eighth grades. These additional plausible values were used to establish the linear transformation necessary to place the paperTIMSS and bridge 2019 data on the appropriate TIMSS trend scales.

Transforming the Overall Scores to Measure Trends

To provide results for the TIMSS 2019 assessments on the existing TIMSS achievement scales, the 2019 plausible values for overall mathematics and overall science had to be transformed to the TIMSS reporting metric. This was accomplished through a set of linear transformations as part of the concurrent calibration approach. These linear transformations were given by:

$$PV_{ik}^{*} = A_{ik} + B_{ik} \times PV_{ik} \tag{12.1}$$

where

PV_{ik} is the TIMSS 2019 plausible value i of scale k prior to transformation;

PV_{ik}^{*} is the TIMSS 2019 plausible value i of scale k after transformation; and

A_{ik} and B_{ik} are the linear transformation constants.

The linear transformation constants were obtained by first computing the international means and standard deviations of the plausible values for the overall mathematics and science scales using the plausible values produced in 2015 based on the 2015 item calibrations for the trend countries. These were the plausible values published in 2015. Next, the same calculations were done using the plausible values from the re-scaled TIMSS 2015 assessment data based on the 2019 paperTIMSS and bridge concurrent item calibrations for the same set of trend countries. From these calculations, the linear transformation constants were defined as:

$$B_{ik} = \sigma_{ik} / \sigma_{ik}^{*} \tag{12.2}$$

$$A_{ik} = \mu_{ik} - B_{ik} \cdot \mu_{ik}^{*} \tag{12.3}$$

where

μ_{ik} is the international mean of scale k based on plausible value i published in 2015;

μ_{ik}^{*} is the international mean of scale k based on plausible value i from the 2015 assessment based on the 2019 concurrent calibration;

σ_{ik} is the international standard deviation of scale k based on plausible value i published in 2015;

σ_{ik}^{*} is the international standard deviation of scale k based on plausible value i from the 2015 assessment based on the 2019 concurrent calibration.

There are five sets of transformation constants for each scale, one for each plausible value. The trend countries contributed equally to the calculation of these transformation constants. Exhibits 12.12 and 12.13 show the TIMSS 2019 transformation constants for both subjects at the fourth grade and eighth grade, respectively. These transformation constants were applied to overall mathematics, overall science, and their respective content and cognitive domains. They also were applied across all TIMSS 2019 assessments: paperTIMSS, eTIMSS, bridge, and science for the countries participating in the less difficult assessment.

Exhibit 12.12: Transformation Constants for the TIMSS 2019 Grade 4 Data

Overall Mathematics	TIMSS 2015 Published Scores		TIMSS 2015 Re-Scaled Scores		A_{ik}	B_{ik}
	Mean	Standard Deviation	Mean	Standard Deviation		
PV1	511.25828	99.44908	-0.05863	1.03038	516.91736	96.51715
PV2	511.32879	99.96828	-0.05780	1.03165	516.92943	96.90131
PV3	511.72035	98.57866	-0.05873	1.02959	517.34333	95.74545
PV4	511.07161	99.70953	-0.05635	1.03198	516.51649	96.61951
PV5	510.86364	99.52263	-0.05629	1.02952	516.30491	96.66892

Overall Science	TIMSS 2015 Published Scores		TIMSS 2015 Re-Scaled Scores		A_{ik}	B_{ik}
	Mean	Standard Deviation	Mean	Standard Deviation		
PV1	506.83611	99.28332	-0.01640	0.98759	508.48461	100.53116
PV2	505.33314	99.77459	-0.01691	0.98878	507.03961	100.90681
PV3	505.66704	99.87928	-0.01688	0.98785	507.37344	101.10788
PV4	504.63307	100.51279	-0.01857	0.98947	506.51953	101.58220
PV5	506.56374	99.60458	-0.01817	0.98877	508.39371	100.73569

Exhibit 12.13: Transformation Constants for the TIMSS 2019 Grade 8 Data

Overall Mathematics	TIMSS 2015 Published Scores		TIMSS 2015 Re-Scaled Scores		A_{ik}	B_{ik}
	Mean	Standard Deviation	Mean	Standard Deviation		
PV1	481.22701	110.64947	-0.03701	0.99368	485.34848	111.35346
PV2	481.67630	111.33530	-0.03596	0.99442	485.70283	111.95996
PV3	481.41115	112.03210	-0.03787	0.99406	485.67858	112.70101
PV4	480.58759	112.61280	-0.03757	0.99614	484.83512	113.04950
PV5	481.40015	111.97331	-0.03629	0.99284	485.49337	112.78062

Overall Science	TIMSS 2015 Published Scores		TIMSS 2015 Re-Scaled Scores		A_{ik}	B_{ik}
	Mean	Standard Deviation	Mean	Standard Deviation		
PV1	484.92745	109.68547	-0.00066	0.94325	485.00413	116.28414
PV2	484.89171	109.07120	-0.00075	0.94167	484.97803	115.82786
PV3	486.02108	108.36916	-0.00139	0.94144	486.18070	115.10991
PV4	484.61364	110.12633	0.00005	0.94190	484.60824	116.91939
PV5	485.54639	109.34609	0.00021	0.94104	485.52214	116.19689

Evaluating Model Fit to the TIMSS Assessment Data

After scaling the TIMSS 2019 paperTIMSS and bridge data, extensive checks were performed to verify the fit of the IRT models applied to these data, in terms of item calibration, proficiency estimation, and link to the TIMSS trend reporting scales. One key method consisted of evaluating the fit of the estimated item characteristic curves to the empirical response data. A second critical method consisted of measuring the accuracy in re-estimating the TIMSS 2015 achievement results across the pool of trend countries, a crucial component in accurately reporting TIMSS 2019 results on the TIMSS trend scales. This involved quantifying the linking error between the 2015 and 2019 assessments.

Item Characteristic Curves

Model fit was assessed by visually comparing the item response function curves generated using the item parameters estimated from the data with the empirical item response function curves calculated from the latent abilities estimated for each student that responded to an item. The empirical functions are themselves based on an estimated latent ability distribution that uses the IRT model and are therefore also referred to item functions based on *pseudo counts*. When the empirical results for an item fall near the fitted curves, the IRT model for that item fits the data well and provides an accurate and reliable measurement of the underlying proficiency scale.

Plots of these response function curves are called item characteristic curves (ICC). The plots in Exhibits 12.14 and 12.15 show examples of the empirical and fitted item response functions for dichotomously scored (right/wrong) multiple-choice and constructed response items, respectively. In each plot, the horizontal axis represents the proficiency scale on the logit metric, and the vertical axis represents the probability of a correct response. The fitted curve based on the estimated item parameters is shown as a solid line, with the item slope parameter represented by the slope of the curve between the two inflexion points, the difficulty or location parameter represented by the point on the horizontal axis where the probability of a correct response is 50 percent, and, for multiple-choice items, a lower asymptote corresponding to the guessing parameter.

Empirical results based on *pseudo counts* are represented by circles. The empirical results are obtained by first dividing the logit proficiency scale into intervals of equal size and then counting the number of students responding to the item whose *estimated* latent abilities (EAP scores estimated by Parscale) fall in each interval. Then the proportion of students in each interval that responded correctly to the item is calculated. In the exhibits, the center of each circle represents this empirical proportion of correct responses. The size of each circle is proportional to the estimated number of students contributing to the empirical proportion correct in its corresponding interval.

Exhibit 12.14: Example Item Response Function for a Dichotomous Multiple-Choice Item from paperTIMSS 2019 Grade 8 Mathematics

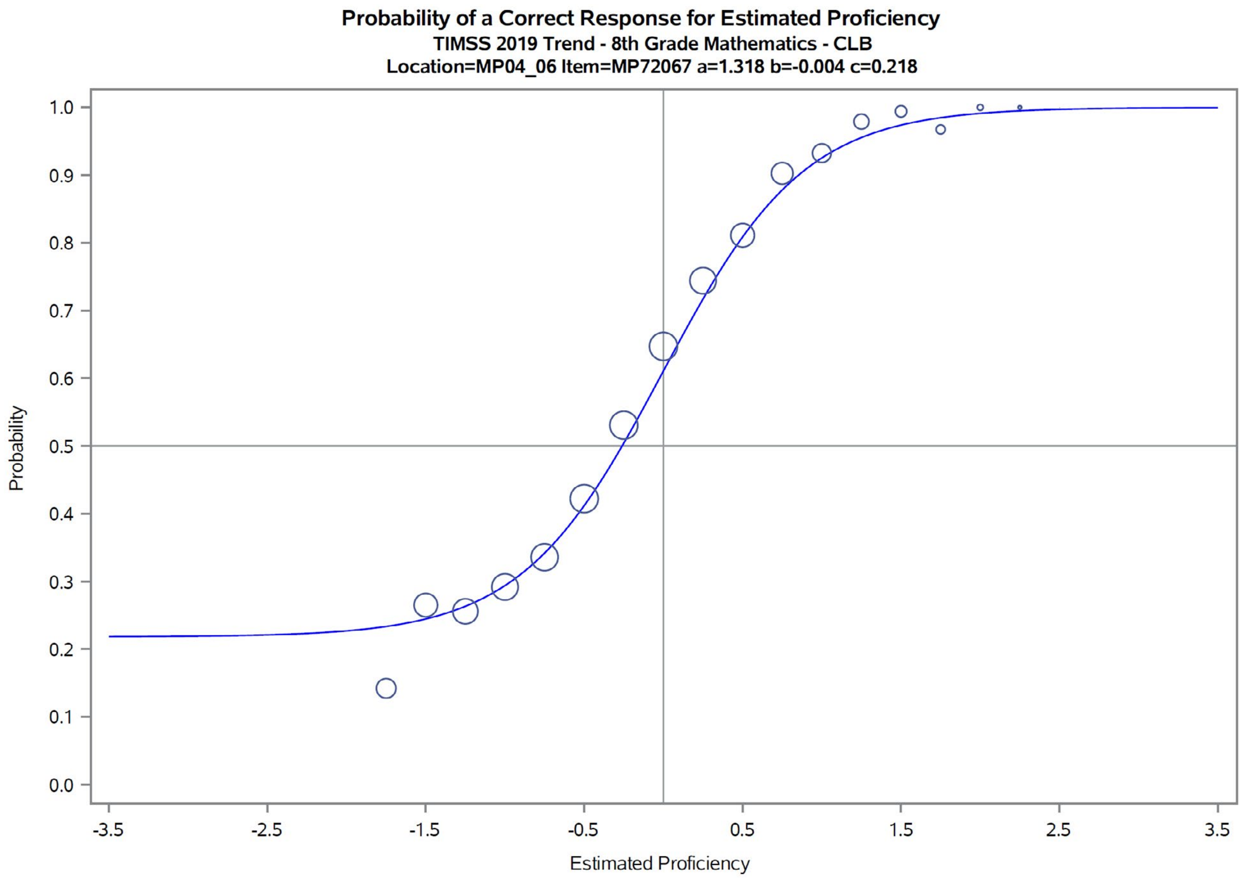
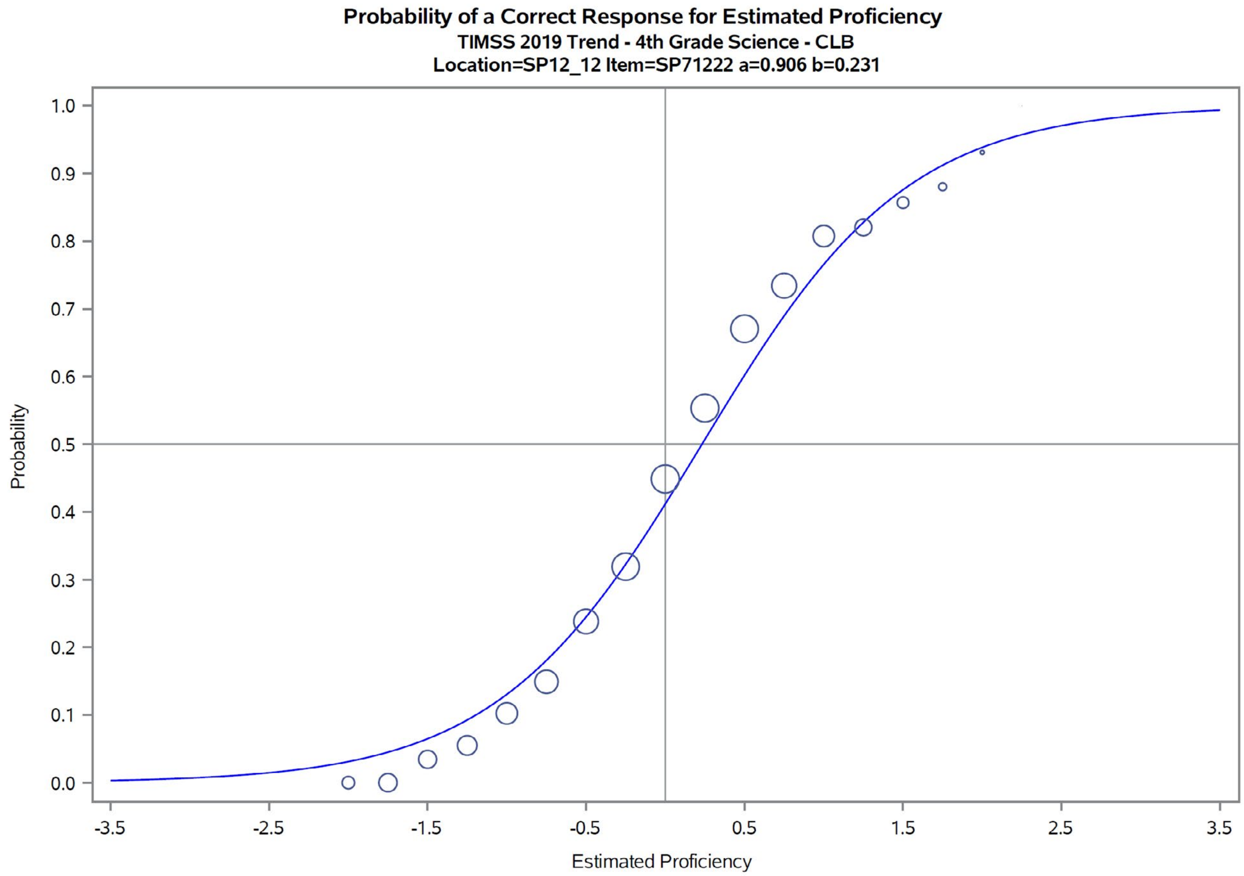


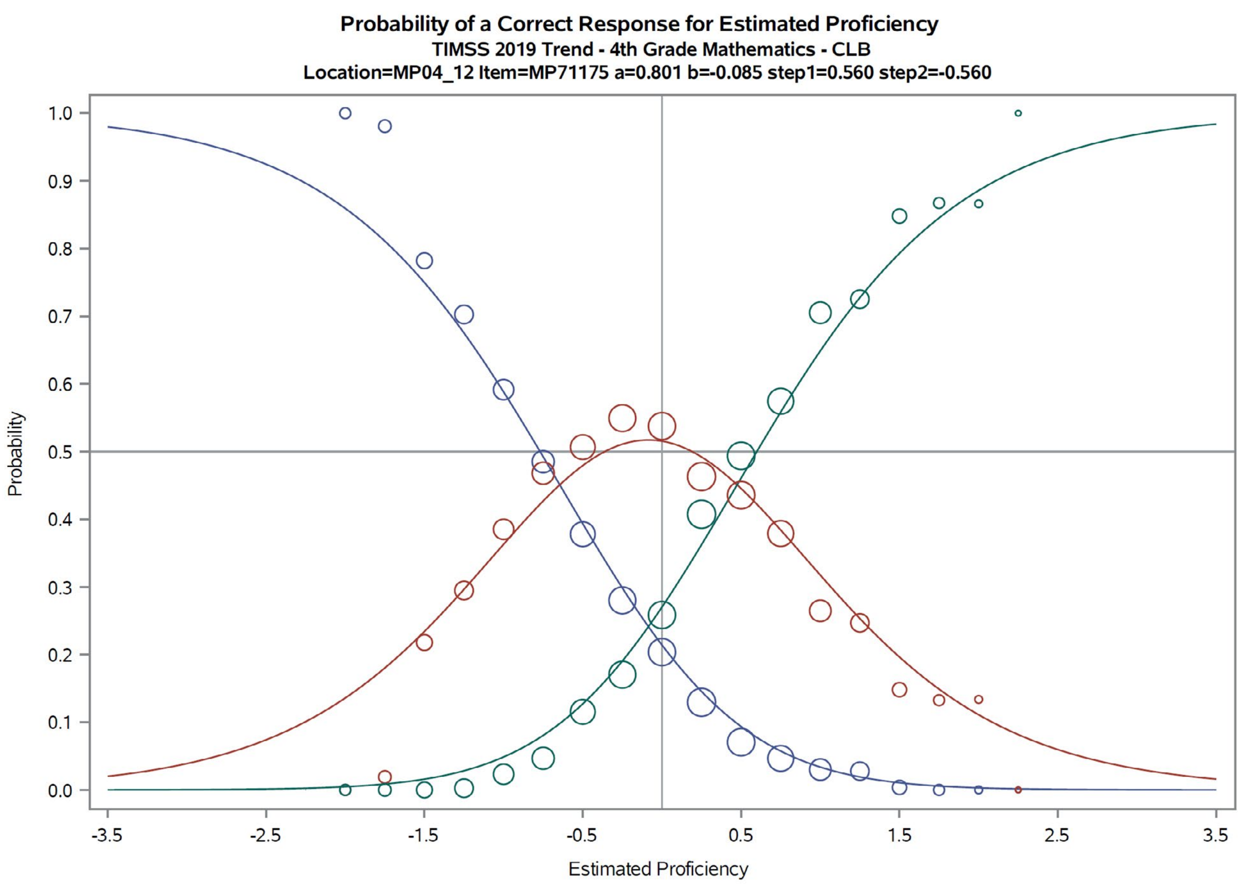
Exhibit 12.15: Example Item Response Function for a Dichotomous Constructed-Response Item from paperTIMSS 2019 Grade 4 Science



In addition to graphical model fit assessment, the fit of an item to the IRT model is quantified by the root mean square difference (RMSD) statistic. The RMSD is the square root of the average of squared differences (i.e., the area) between the empirical curve, shown as bubbles, and the fitted curve, shown as the straight line, weighted by the size of the empirical bubbles. The RMSD statistics for the items shown in exhibits 12.14 and 12.15 are 0.028 and 0.038, respectively. RMSD values less than 0.1 were considered to indicate good fit.

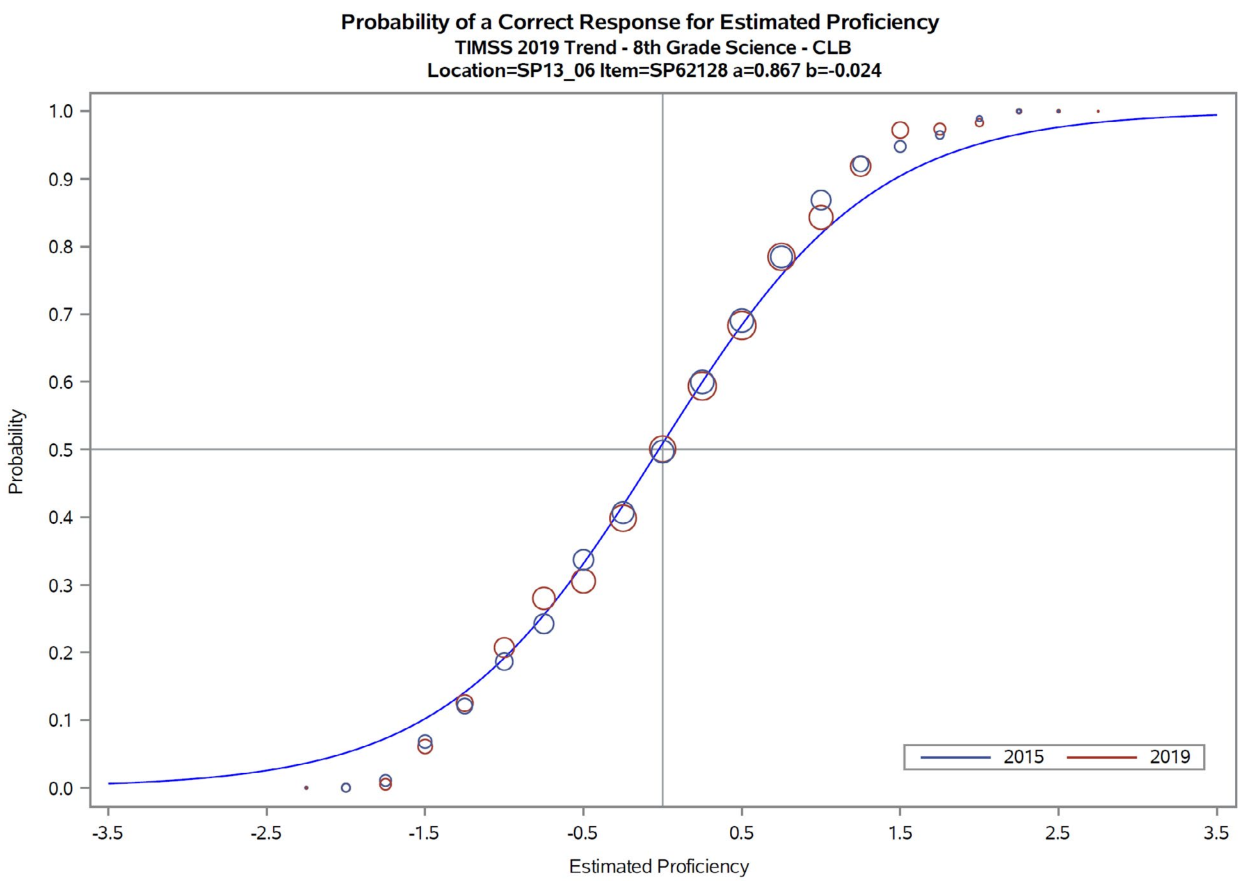
The ICC plot in Exhibit 12.16 shows the empirical and fitted item response functions for a polytomous item (scored 0, 1, or 2). As for the dichotomous item plots above, the horizontal axis represents the proficiency scale in logits, but in this example the vertical axis represents the probability of having a response in a given response category. The fitted curves based on the estimated item parameters are shown as solid lines and the empirical results are represented by circles. The interpretation of the circles is the same as in Exhibits 12.14 and 12.15. The curve starting at the top left of the chart plots the probability of a score of zero on the item. This probability should always decrease as proficiency increases. The bell-shaped curve shows the probability of a score of one point—partial credit, which should start low approaching zero for low-ability students, reaching a maximum for medium-ability students, and decreasing for high-ability students. The curve ending at the top right corner of the chart shows the probability of a score of two points—full credit, starting low for low-ability students and increasing as proficiency increases. For this particular item, the RMSD value is 0.035, calculated from all three response curves.

Exhibit 12.16: Example Item Response Function for a Polytomous Constructed-Response Item from paperTIMSS 2019 Grade 4 Mathematics



Although a single set of item parameters was estimated for any given item in the concurrent calibration for 2019, trend items have two empirical curves, one for each assessment cycle. Plotting both empirical curves from 2019 and 2015 allowed for a visual inspection of the invariance of the item parameters between cycles; a key aspect of the link to the trend scale. Exhibit 12.17 shows the ICC for a paperTIMSS 2019 trend item, with its single fitted curve and two empirical curves: the blue bubbles represent the empirical curve based on the TIMSS 2015 response data, the red curve the empirical curve based on the TIMSS 2019 response data. Thus, for trend items, there are two RMSD values. The RMSD values for this particular item are 0.027 based on the 2015 data and 0.028 based on the 2019 data.

Exhibit 12.17: Example Item Response Function for a Dichotomous Constructed-Response Trend Item from paperTIMSS 2019 Grade 8 Science



RMSD values were computed for all items included in the paperTIMSS 2019 concurrent calibrations. These values are shown in the item parameter exhibits of Appendices 12A through 12D. They are also presented graphically in Exhibits 12.18 through 12.21 for the fourth and eighth grades and for mathematics and science, respectively. In each exhibit, the items are sorted from smallest to largest RMSD values. For trend items with two RMSD values, the largest of the two determined the order. Across both grades and subjects, the vast majority of paperTIMSS items have RMSD values less than 0.04. All paperTIMSS items have RMSD values less than 0.10.

Exhibit 12.18: RMSD Statistics for Items in the paperTIMSS 2019 Concurrent Calibration—Grade 4 Mathematics

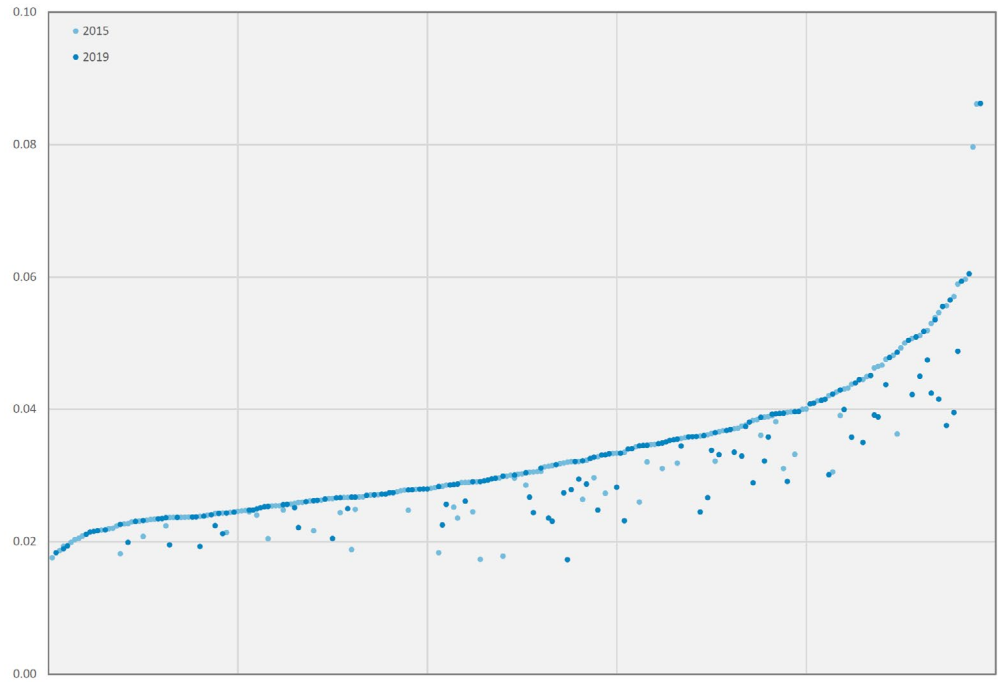


Exhibit 12.19: RMSD Statistics for Items in the paperTIMSS 2019 Concurrent Calibration—Grade 4 Science

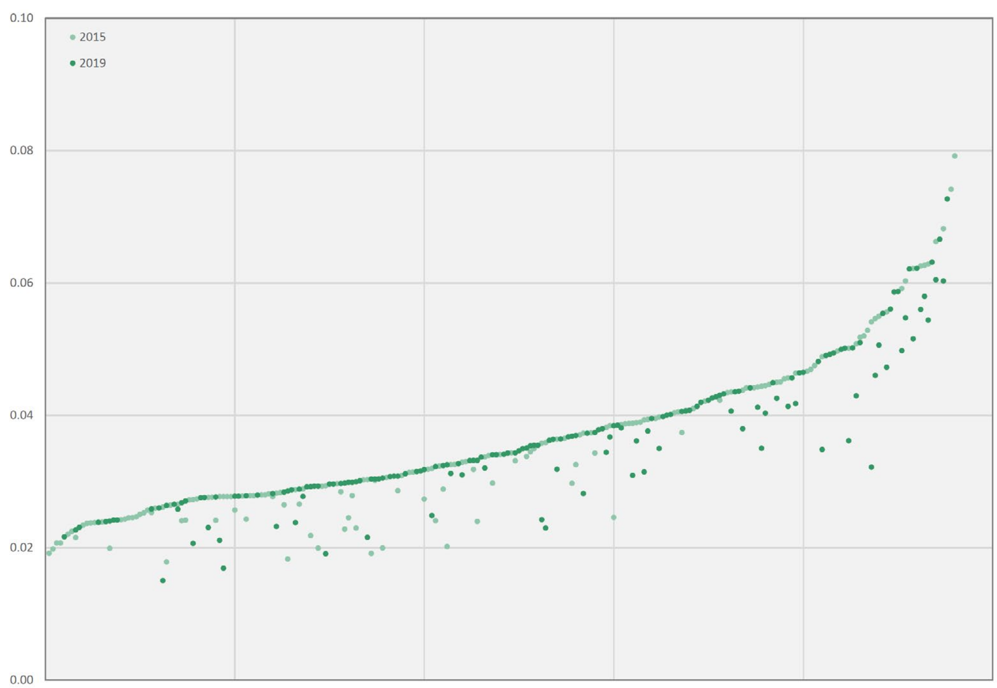


Exhibit 12.20: RMSD Statistics for Items in the paperTIMSS 2019 Concurrent Calibration—Grade 8 Mathematics

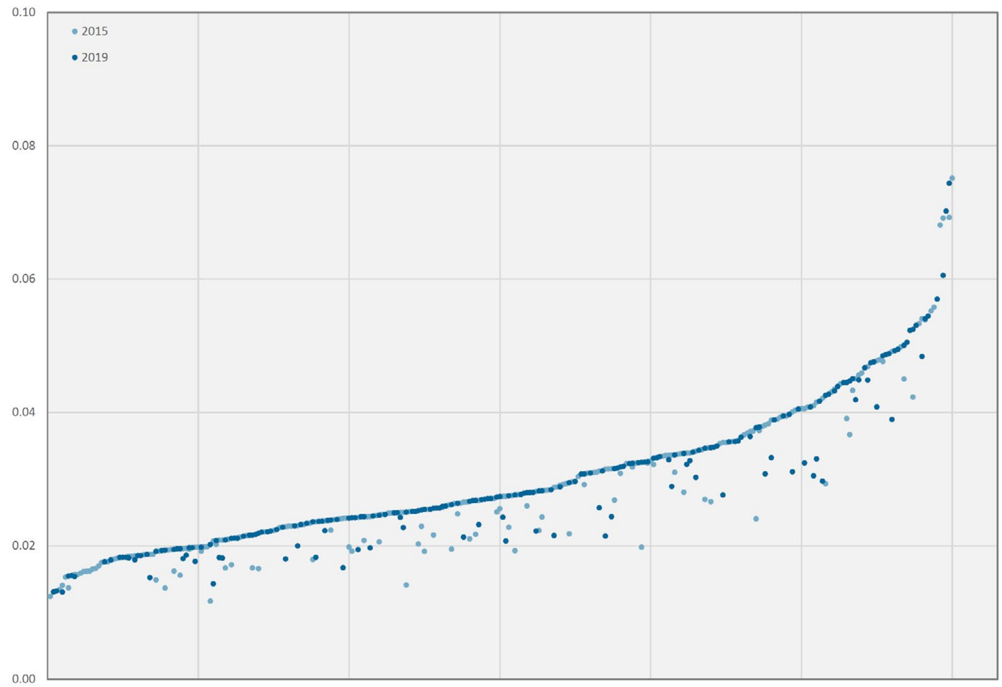
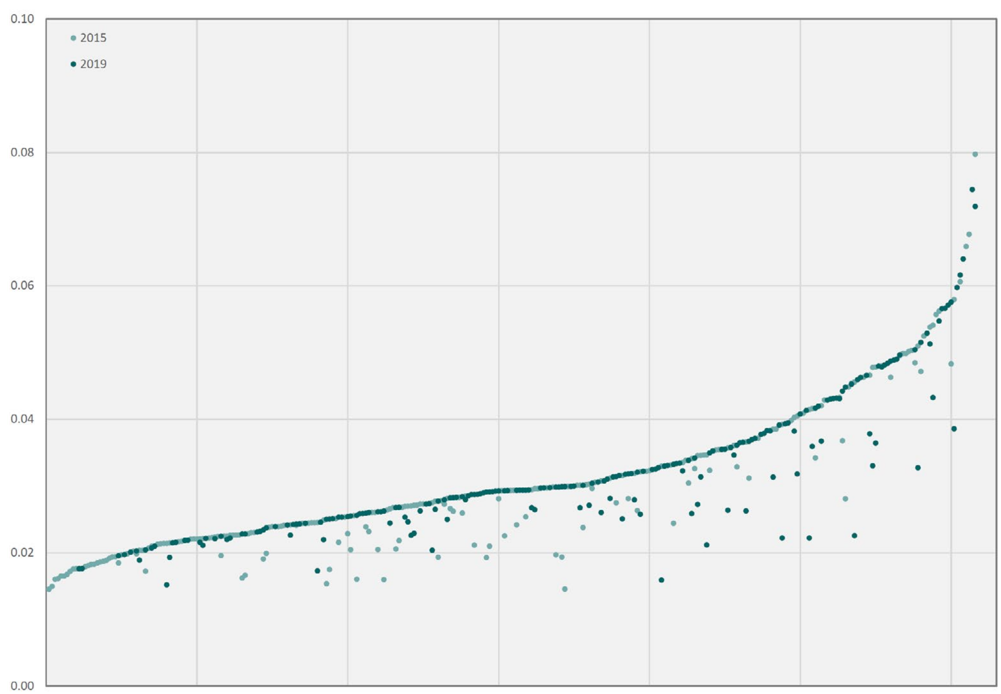


Exhibit 12.21: RMSD Statistics for Items in the paperTIMSS 2019 Concurrent Calibration—Grade 8 Science



Quantifying the Linking Error between 2015 and 2019

A key aspect of reporting the TIMSS 2019 results on the TIMSS trend scales is the ability to accurately re-estimate the TIMSS 2015 achievement results based on a concurrent calibration of the 2015 and 2019 data across a common set of trend countries. As described earlier, this re-estimation serves to establish the linear transformation that places the TIMSS 2019 results on the TIMSS trend scale. Although this transformation is set globally to match the overall mean and standard deviation across the trend countries, it also should achieve a good alignment of the 2015 results across calibrations for each individual trend country. The difference between a trend country's TIMSS 2015 achievement mean published back in 2015 and re-estimated in 2019 gives a good measure of quality of the link between the two assessments. The linking error is quantified by the standard error of difference, for each country and aggregated over the countries (see Martin, Mullis, Foy, Brossman, & Stanco, 2012).

Exhibits 12.22 through 12.25 provide results on the linking error associated with the paperTIMSS 2019 results for the fourth and eighth grades and for mathematics and science, respectively. Across both grades and subjects, there was good agreement between the countries' published and re-estimated 2015 results. Only one country shows a statistically significant difference for eighth grade science. In the vast majority of cases, the differences are within one point and the standard errors rarely exceed 2 points.

Exhibit 12.22: Trend Linking Error for the paperTIMSS 2019 Grade 4 Mathematics Scale

Country	TIMSS 2015 Published Results		TIMSS 2015 Re-Estimated Results		Difference (Linking Error)
	Mean	Standard Deviation	Mean	Standard Deviation	
Armenia	481 (3.4)	80 (1.5)	481 (3.5)	80 (1.6)	0 (0.9)
Australia	517 (3.1)	83 (1.8)	518 (3.0)	84 (1.6)	0 (0.9)
Bahrain	450 (2.1)	85 (1.8)	451 (2.1)	84 (1.5)	1 (1.6)
Belgium (Flemish)	546 (2.1)	61 (1.2)	545 (2.0)	61 (1.3)	0 (0.6)
Bulgaria	524 (5.3)	83 (2.6)	524 (5.2)	83 (2.7)	0 (1.4)
Canada	511 (2.3)	75 (1.9)	511 (2.5)	75 (2.1)	0 (1.0)
Chile	459 (2.4)	73 (1.5)	459 (2.6)	74 (1.4)	0 (0.8)
Chinese Taipei	597 (1.9)	71 (1.2)	597 (1.8)	71 (1.5)	0 (0.9)
Croatia	502 (1.8)	66 (1.0)	502 (2.0)	66 (1.4)	0 (1.4)
Cyprus	523 (2.7)	81 (1.2)	523 (2.5)	81 (1.2)	0 (0.8)
Czech Republic	528 (2.2)	70 (1.3)	528 (2.3)	70 (1.4)	0 (0.7)
Denmark	539 (2.7)	75 (1.6)	538 (2.7)	75 (1.4)	0 (1.3)
England	546 (2.8)	84 (2.2)	547 (2.9)	84 (2.1)	0 (0.7)
Finland	535 (2.0)	67 (1.2)	535 (1.9)	67 (1.2)	0 (0.8)
France	488 (2.9)	74 (1.3)	488 (2.8)	74 (1.5)	0 (1.2)
Georgia	463 (3.6)	87 (2.4)	463 (3.6)	86 (2.4)	0 (1.5)
Germany	522 (2.0)	65 (1.2)	522 (2.1)	65 (1.2)	0 (0.8)
Hong Kong SAR	615 (2.9)	66 (1.7)	614 (3.0)	66 (1.5)	-1 (0.8)
Hungary	529 (3.2)	88 (2.3)	529 (3.2)	88 (2.3)	0 (0.8)
Iran, Islamic Rep. of	424 (4.1)	101 (3.0)	424 (3.7)	100 (2.5)	1 (1.6)
Ireland	547 (2.1)	73 (1.2)	547 (2.1)	74 (1.4)	0 (0.9)
Italy	507 (2.6)	72 (1.7)	507 (2.5)	72 (1.8)	0 (0.9)
Japan	593 (2.0)	69 (1.0)	593 (1.9)	69 (1.1)	0 (0.8)
Korea, Rep. of	608 (2.2)	67 (1.4)	609 (2.1)	69 (1.7)	1 (0.9)
Kuwait	351 (4.8)	101 (2.4)	352 (4.8)	100 (2.3)	1 (2.0)
Lithuania	535 (2.5)	71 (1.5)	535 (2.5)	72 (1.4)	0 (0.9)
Morocco	379 (3.8)	91 (2.6)	380 (3.7)	91 (2.2)	0 (2.0)
Netherlands	530 (1.7)	56 (1.0)	529 (1.7)	56 (0.9)	0 (0.9)
New Zealand	491 (2.3)	90 (1.5)	491 (2.4)	90 (1.6)	0 (1.2)
Northern Ireland	570 (2.9)	86 (1.7)	570 (3.0)	86 (2.0)	-1 (1.0)
Norway (5)	549 (2.5)	71 (1.4)	549 (2.5)	71 (1.5)	0 (1.0)
Oman	425 (2.5)	101 (1.3)	426 (2.5)	101 (1.6)	0 (0.9)
Poland	535 (2.1)	71 (1.1)	535 (2.3)	72 (1.2)	0 (1.1)
Portugal	541 (2.2)	72 (1.2)	541 (2.3)	73 (1.2)	0 (1.0)
Qatar	439 (3.4)	97 (2.3)	439 (3.4)	96 (2.4)	0 (1.1)
Russian Federation	564 (3.4)	73 (2.4)	564 (3.4)	73 (2.4)	0 (0.9)
Saudi Arabia	383 (4.1)	92 (2.2)	383 (4.1)	92 (2.3)	0 (1.0)
Serbia	518 (3.5)	87 (2.8)	517 (3.5)	88 (2.4)	-1 (1.1)
Singapore	618 (3.8)	86 (2.6)	619 (3.9)	87 (2.6)	1 (0.5)
Slovak Republic	498 (2.5)	80 (1.7)	498 (2.3)	79 (1.8)	0 (0.8)

Linking error is the standard error associated with the difference. Standard errors are shown in parentheses.

Exhibit 12.22: Trend Linking Error for the paperTIMSS 2019 Grade 4 Mathematics Scale (continued)

Country	TIMSS 2015 Published Results		TIMSS 2015 Re-Estimated Results		Difference (Linking Error)
	Mean	Standard Deviation	Mean	Standard Deviation	
Spain	505 (2.5)	69 (1.3)	504 (2.4)	69 (1.3)	-1 (1.0)
Sweden	519 (2.8)	69 (1.7)	519 (2.8)	70 (1.7)	0 (0.8)
United Arab Emirates	452 (2.4)	105 (1.5)	452 (2.5)	105 (1.5)	1 (0.6)
United States	539 (2.3)	81 (1.3)	539 (2.2)	82 (1.3)	0 (0.7)
International Average	**511 (0.5)**	**79 (0.5)**	**511 (0.5)**	**79 (0.5)**	**0 (0.1)**

Exhibit 12.23: Trend Linking Error for the paperTIMSS 2019 Grade 4 Science Scale

Country	TIMSS 2015 Published Results		TIMSS 2015 Re-Estimated Results		Difference (Linking Error)
	Mean	Standard Deviation	Mean	Standard Deviation	
Armenia	444 (4.0)	87 (1.6)	442 (3.9)	87 (1.5)	-1 (1.0)
Australia	524 (2.9)	76 (1.9)	523 (2.8)	77 (1.8)	0 (1.3)
Bahrain	459 (2.6)	105 (1.7)	460 (2.7)	104 (1.5)	2 (1.6)
Belgium (Flemish)	512 (2.3)	62 (1.2)	510 (2.3)	63 (1.2)	-1 (0.8)
Bulgaria	536 (5.9)	95 (3.6)	537 (6.1)	96 (3.8)	1 (1.2)
Canada	525 (2.6)	73 (1.6)	525 (2.6)	74 (1.6)	0 (0.7)
Chile	478 (2.7)	74 (1.4)	477 (2.8)	74 (1.5)	-1 (1.1)
Chinese Taipei	555 (1.8)	68 (1.1)	555 (2.1)	68 (1.2)	0 (1.1)
Croatia	533 (2.1)	62 (1.1)	533 (2.4)	63 (1.2)	0 (1.2)
Cyprus	481 (2.6)	76 (1.4)	481 (2.7)	76 (1.3)	0 (1.3)
Czech Republic	534 (2.4)	70 (1.4)	535 (2.1)	70 (1.2)	1 (1.6)
Denmark	527 (2.1)	69 (1.3)	527 (2.4)	70 (1.6)	0 (1.2)
England	536 (2.4)	70 (1.7)	536 (2.4)	70 (1.6)	0 (1.0)
Finland	554 (2.3)	65 (1.7)	553 (2.2)	65 (1.6)	0 (0.8)
France	487 (2.7)	73 (1.4)	487 (2.7)	73 (1.3)	-1 (1.0)
Georgia	451 (3.7)	87 (2.5)	450 (3.9)	88 (2.5)	-1 (1.4)
Germany	528 (2.4)	70 (1.3)	528 (2.4)	70 (1.2)	-1 (0.9)
Hong Kong SAR	557 (2.9)	70 (1.4)	557 (3.1)	71 (1.6)	1 (1.0)
Hungary	542 (3.3)	83 (2.7)	542 (3.3)	83 (2.7)	0 (1.0)
Iran, Islamic Rep. of	421 (4.0)	103 (3.0)	421 (4.4)	102 (2.8)	0 (2.2)
Ireland	529 (2.4)	70 (2.0)	529 (2.6)	70 (1.5)	0 (1.5)
Italy	516 (2.6)	66 (1.3)	516 (2.6)	68 (1.4)	0 (1.5)
Japan	569 (1.8)	65 (1.0)	568 (2.0)	67 (1.2)	-1 (1.3)
Korea, Rep. of	589 (2.0)	62 (0.9)	589 (1.9)	63 (1.0)	0 (1.1)
Kuwait	337 (6.2)	126 (2.0)	340 (6.3)	124 (2.1)	3 (2.3)
Lithuania	528 (2.5)	69 (1.2)	528 (2.5)	70 (1.3)	0 (1.2)
Morocco	352 (4.7)	120 (2.7)	353 (4.6)	119 (2.5)	1 (1.6)
Netherlands	517 (2.7)	60 (1.3)	517 (2.4)	61 (1.2)	0 (1.6)
New Zealand	506 (2.7)	85 (1.6)	505 (2.4)	86 (1.6)	-1 (1.0)
Northern Ireland	520 (2.2)	70 (1.5)	519 (2.8)	71 (1.6)	-1 (1.6)
Norway (5)	538 (2.6)	63 (1.5)	537 (2.3)	63 (1.6)	0 (1.0)
Oman	431 (3.1)	119 (1.8)	432 (3.1)	119 (1.7)	1 (1.1)
Poland	547 (2.4)	69 (1.4)	548 (2.1)	70 (1.2)	1 (1.4)
Portugal	508 (2.2)	60 (1.0)	508 (2.1)	61 (1.0)	0 (0.7)
Qatar	436 (4.1)	111 (2.2)	437 (4.2)	111 (2.5)	1 (1.3)
Russian Federation	567 (3.2)	69 (1.9)	566 (3.0)	68 (2.0)	-1 (1.0)
Saudi Arabia	390 (4.9)	116 (2.9)	392 (4.7)	115 (2.6)	1 (1.7)
Serbia	525 (3.7)	81 (3.4)	524 (3.6)	82 (2.8)	-1 (1.2)
Singapore	590 (3.7)	85 (2.6)	591 (3.7)	85 (2.5)	1 (0.9)
Slovak Republic	520 (2.6)	85 (1.9)	520 (2.9)	86 (1.8)	-1 (1.2)

Linking error is the standard error associated with the difference. Standard errors are shown in parentheses.

Exhibit 12.23: Trend Linking Error for the paperTIMSS 2019 Grade 4 Science Scale (continued)

Country	TIMSS 2015 Published Results		TIMSS 2015 Re-Estimated Results		Difference (Linking Error)
	Mean	Standard Deviation	Mean	Standard Deviation	
Spain	518 (2.6)	69 (1.6)	518 (2.6)	70 (1.5)	0 (1.3)
Sweden	540 (3.6)	73 (2.5)	540 (3.3)	74 (2.5)	0 (1.5)
United Arab Emirates	451 (2.8)	121 (1.5)	452 (2.9)	120 (1.5)	1 (0.7)
United States	546 (2.2)	81 (1.2)	546 (2.3)	81 (1.4)	0 (0.5)
International Average	**506 (1.1)**	**80 (0.5)**	**506 (1.1)**	**81 (0.5)**	**0 (0.1)**

Exhibit 12.24: **Trend Linking Error for the paperTIMSS 2019 Grade 8 Mathematics Scale**

Country	TIMSS 2015 Published Results		TIMSS 2015 Re-Estimated Results		Difference (Linking Error)
	Mean	Standard Deviation	Mean	Standard Deviation	
Australia	505 (3.1)	82 (1.9)	505 (3.0)	83 (1.8)	0 (0.9)
Bahrain	454 (1.4)	80 (1.4)	453 (1.5)	80 (1.3)	0 (1.3)
Chile	427 (3.2)	80 (1.9)	427 (2.9)	80 (1.9)	0 (1.2)
Chinese Taipei	599 (2.4)	97 (1.7)	600 (2.5)	99 (1.7)	1 (0.8)
Egypt	392 (4.1)	99 (2.0)	392 (3.8)	97 (1.8)	0 (1.0)
England	518 (4.2)	80 (2.6)	518 (4.2)	81 (2.6)	0 (1.7)
Georgia	453 (3.4)	92 (1.7)	454 (3.8)	91 (2.3)	0 (1.2)
Hong Kong SAR	594 (4.6)	78 (2.8)	595 (4.7)	80 (2.7)	0 (0.8)
Hungary	514 (3.8)	93 (2.2)	514 (3.8)	94 (2.3)	0 (1.1)
Iran, Islamic Rep. of	436 (4.6)	94 (2.7)	436 (4.6)	94 (2.8)	-1 (1.8)
Ireland	523 (2.7)	74 (2.3)	523 (2.7)	74 (2.2)	-1 (1.0)
Israel	511 (4.1)	102 (2.3)	511 (4.1)	102 (2.2)	0 (0.6)
Italy	494 (2.5)	75 (1.8)	493 (2.5)	75 (1.5)	-1 (0.7)
Japan	586 (2.3)	89 (1.3)	587 (2.5)	89 (1.3)	0 (0.9)
Jordan	386 (3.2)	94 (1.7)	386 (3.2)	92 (1.5)	1 (0.9)
Korea, Rep. of	606 (2.6)	85 (1.1)	606 (2.8)	86 (1.4)	1 (1.3)
Kuwait	392 (4.6)	91 (3.3)	393 (4.5)	90 (3.2)	0 (1.7)
Lebanon	442 (3.6)	75 (1.7)	442 (3.8)	75 (1.9)	-1 (1.2)
Lithuania	511 (2.8)	77 (1.5)	511 (2.9)	79 (1.8)	-1 (1.1)
Malaysia	465 (3.6)	87 (2.1)	465 (3.5)	87 (1.9)	-1 (0.6)
Morocco	384 (2.3)	80 (1.3)	384 (2.1)	79 (1.3)	0 (0.6)
New Zealand	493 (3.4)	88 (2.0)	493 (3.3)	88 (1.9)	0 (0.8)
Norway (9)	512 (2.3)	70 (1.2)	512 (2.2)	70 (1.1)	0 (0.7)
Oman	403 (2.4)	96 (1.3)	403 (2.6)	94 (1.6)	0 (1.3)
Qatar	437 (3.0)	102 (2.2)	437 (2.8)	102 (1.8)	0 (1.2)
Russian Federation	538 (4.7)	82 (1.8)	537 (4.8)	83 (1.8)	-1 (1.2)
Saudi Arabia	368 (4.6)	86 (2.9)	368 (4.2)	85 (2.7)	0 (2.4)
Singapore	621 (3.2)	82 (2.2)	622 (3.3)	83 (2.3)	1 (0.7)
South Africa (9)	372 (4.5)	87 (3.0)	373 (4.5)	85 (3.1)	1 (0.8)
Sweden	501 (2.8)	72 (1.9)	501 (2.8)	72 (1.5)	0 (1.1)
Turkey	458 (4.7)	105 (2.8)	458 (4.5)	105 (2.2)	0 (1.4)
United Arab Emirates	465 (2.0)	98 (1.5)	464 (2.0)	97 (1.5)	0 (0.6)
United States	518 (3.1)	83 (1.6)	518 (3.1)	84 (1.6)	0 (0.6)
International Average	**481 (0.7)**	**87 (0.7)**	**481 (0.7)**	**87 (0.7)**	**0 (0.1)**

Linking error is the standard error associated with the difference. Standard errors are shown in parentheses.

Exhibit 12.25: Trend Linking Error for the paperTIMSS 2019 Grade 8 Science Scale

Country	TIMSS 2015 Published Results		TIMSS 2015 Re-Estimated Results		Difference (Linking Error)	
	Mean	Standard Deviation	Mean	Standard Deviation		
Australia	512 (2.7)	82 (1.5)	512 (2.6)	82 (1.6)	0 (0.8)	
Bahrain	466 (2.2)	106 (1.8)	466 (2.1)	105 (1.7)	1 (1.2)	
Chile	454 (3.1)	81 (1.5)	454 (3.1)	81 (1.5)	0 (1.1)	
Chinese Taipei	569 (2.1)	83 (1.2)	570 (2.1)	84 (1.3)	0 (1.1)	
Egypt	371 (4.3)	115 (1.9)	371 (4.4)	113 (2.0)	0 (1.2)	
England	537 (3.8)	81 (2.3)	537 (3.8)	82 (2.2)	0 (0.9)	
Georgia	443 (3.1)	87 (1.7)	443 (3.1)	87 (1.6)	0 (2.4)	
Hong Kong SAR	546 (3.9)	72 (2.2)	546 (3.9)	72 (2.3)	0 (0.8)	
Hungary	527 (3.4)	85 (2.3)	526 (3.5)	86 (2.2)	-1 (1.6)	
Iran, Islamic Rep. of	456 (4.0)	89 (2.3)	456 (4.0)	90 (2.4)	-1 (1.2)	
Ireland	530 (2.8)	80 (2.5)	530 (2.9)	81 (2.5)	0 (0.8)	
Israel	507 (3.9)	104 (2.5)	506 (3.9)	105 (2.3)	-1 (0.8)	
Italy	499 (2.4)	76 (1.7)	499 (2.4)	76 (1.6)	0 (1.4)	
Japan	571 (1.8)	75 (1.3)	571 (1.8)	76 (1.1)	0 (0.8)	
Jordan	426 (3.4)	101 (2.1)	426 (3.2)	101 (2.2)	0 (1.2)	
Korea, Rep. of	556 (2.2)	78 (1.1)	556 (2.1)	78 (1.0)	0 (1.1)	
Kuwait	411 (5.2)	110 (3.7)	411 (5.3)	110 (3.6)	0 (1.8)	
Lebanon	398 (5.3)	102 (2.6)	398 (5.6)	102 (3.0)	-1 (1.5)	
Lithuania	519 (2.8)	78 (1.8)	518 (2.7)	77 (1.6)	-1 (1.0)	
Malaysia	471 (4.1)	94 (2.7)	471 (4.1)	93 (2.7)	0 (1.0)	
Morocco	393 (2.5)	84 (1.4)	393 (2.3)	83 (1.2)	0 (0.8)	
New Zealand	513 (3.1)	90 (1.9)	512 (3.3)	91 (1.8)	0 (1.1)	
Norway (9)	509 (2.8)	78 (1.6)	508 (2.8)	79 (1.4)	0 (1.1)	
Oman	455 (2.7)	98 (1.6)	455 (2.6)	98 (1.9)	0 (0.8)	
Qatar	457 (3.0)	112 (2.0)	457 (3.0)	112 (2.3)	0 (1.1)	
Russian Federation	544 (4.2)	77 (1.9)	544 (4.2)	78 (2.1)	0 (0.9)	
Saudi Arabia	396 (4.5)	98 (2.7)	396 (4.6)	97 (2.7)	0 (1.5)	
Singapore	597 (3.2)	86 (2.3)	597 (3.3)	88 (2.3)	1 (0.8)	
South Africa (9)	358 (5.6)	108 (3.6)	361 (5.7)	106 (3.6)	3 (1.2)	▲
Sweden	522 (3.4)	86 (2.4)	522 (3.4)	86 (2.2)	0 (1.2)	
Turkey	493 (4.0)	96 (2.0)	493 (4.0)	97 (2.0)	-1 (1.1)	
United Arab Emirates	477 (2.3)	105 (1.6)	477 (2.3)	106 (1.7)	0 (0.9)	
United States	530 (2.8)	82 (1.4)	530 (2.8)	82 (1.4)	0 (0.7)	
International Average	**485 (0.9)**	**90 (0.7)**	**485 (0.9)**	**90 (0.6)**	**0 (0.1)**	

▲ statistically significant difference

Linking error is the standard error associated with the difference. Standard errors are shown in parentheses.

Scaling the Fourth Grade Less Difficult Mathematics Data

All 11 countries that participated in the TIMSS 2019 less difficult mathematics assessment required additional item calibration models to estimate appropriate item parameters and plausible values on the TIMSS trend scales. By its very nature, the less difficult mathematics assessment and the pool of countries that participated required special consideration. Although there were four mathematics item blocks shared with the regular fourth grade mathematics assessment, they proved to be more challenging to students than expected, and solely relying on these four blocks did not produce precise achievement results. Instead, linking the TIMSS 2019 less difficult mathematics assessment by including the data from the TIMSS Numeracy 2015 assessment provided a stronger design with eight shared mathematics item blocks. Item parameters were estimated using the concurrent calibration approach as described for paperTIMSS above, combining TIMSS Numeracy 2015 data from 7 countries and TIMSS 2019 less difficult mathematics data from 11 countries. Exhibit 12.26 shows the number of mathematics items present for the TIMSS 2019 less difficult mathematics concurrent calibration by item type and mathematics content and cognitive domain.

Exhibit 12.26: Mathematics Items for the TIMSS 2019 Grade 4 Less Difficult Concurrent Calibration

Item Type	Points	Items Released in 2015		Items Common in 2015 and 2019		Items Introduced in 2019		Total	
		Items	Points	Items	Points	Items	Points	Items	Points
Multiple Choice	1	11	11	45	45	35	35	91	91
Constructed Response	1	12	12	50	50	34	34	96	96
	2	1	2	5	10	8	16	14	28
Total		**24**	**25**	**100**	**105**	**77**	**85**	**201**	**215**

Items by Content and Cognitive Domains

Mathematics Content Domains	Items Released in 2015		Items Common in 2015 and 2019		Items Introduced in 2019		Total	
	Items	Points	Items	Points	Items	Points	Items	Points
Number	15	16	60	61	36	39	111	116
Measurement and Geometry	8	8	29	32	21	22	58	62
Data	1	1	11	12	20	24	32	37
Mathematics Cognitive Domains	**Items Released in 2015**		**Items Common in 2015 and 2019**		**Items Introduced in 2019**		**Total**	
	Items	**Points**	**Items**	**Points**	**Items**	**Points**	**Items**	**Points**
Knowing	10	11	52	52	27	28	89	91
Applying	11	11	34	35	32	37	77	83
Reasoning	3	3	14	18	18	20	35	41
Total	**24**	**25**	**100**	**105**	**77**	**85**	**201**	**215**

The fourth grade science assessment administered to these countries did not have a counterpart in the TIMSS Numeracy 2015 administration. Consequently, the scaling approach adopted for the science data relied entirely on the TIMSS 2019 data with fixed item parameters for trend items, as estimated from the regular fourth grade science item calibration, and allowing item parameters for the new science items to be estimated appropriately based on the 11 countries that participated in the less difficult assessment. Exhibit 12.27 shows the number of science items present for the TIMSS 2019 less difficult calibration by item type and science content and cognitive domain.

Exhibit 12.27: Science Items for the TIMSS 2019 Grade 4 Less Difficult Calibration

Item Type	Points	Trend Items		New Items		Total	
		Items	Points	Items	Points	Items	Points
Multiple Choice	1	47	47	40	40	87	87
Constructed Response	1	45	45	31	31	76	76
	2	3	6	2	4	5	10
Total		**95**	**98**	**73**	**75**	**168**	**173**

Items by Content and Cognitive Domains

Science Content Domains	Trend Items		New Items		Total	
	Items	Points	Items	Points	Items	Points
Life Science	41	43	32	34	73	77
Physical Science	36	37	25	25	61	62
Earth Science	18	18	16	16	34	34
Science Cognitive Domains	**Trend Items**		**New Items**		**Total**	
	Items	**Points**	**Items**	**Points**	**Items**	**Points**
Knowing	40	43	29	30	69	73
Applying	34	34	30	30	64	64
Reasoning	21	21	14	15	35	36
Total	**95**	**98**	**73**	**75**	**168**	**173**

Exhibit 12.28 shows the sample sizes for scaling the TIMSS 2019 less difficult data, both for item calibration and for proficiency estimation. The mathematics concurrent calibration made use of all TIMSS 2019 data from 11 participating countries and 61,884 students, as well as all the TIMSS Numeracy 2015 data from 7 participating countries and 40,684 students. Three countries participated in both assessments—Kuwait, Morocco, and South Africa. The science calibration made use of the TIMSS 2019 data only.

Exhibit 12.28: Sample Sizes for Scaling the TIMSS 2019 Less Difficult Data

Country	Item Calibration & Proficiency Estimation	
	2019	TIMSS Numeracy 2015
Albania	4,426	—
Bahrain	—	4,429
Bosnia and Herzegovina	5,617	—
Indonesia	—	4,294
Iran, Islamic Rep. of	—	4,105
Jordan	—	7,861
Kosovo	4,496	—
Kuwait	4,437	3,703
Montenegro	5,076	—
Morocco	7,723	5,360
North Macedonia	3,270	—
Pakistan	3,980	—
Philippines	5,515	—
Saudi Arabia	5,453	—
South Africa (5)	11,891	10,932
TOTAL	**61,884**	**40,684**

The item parameters estimated from the TIMSS 2019 less difficult calibrations for mathematics and science are presented in Appendices 12E and 12F, respectively. Appendices 12E and 12F also include the RMSD values computed to measure item-model fit. These item parameters we used to estimate student proficiency in both mathematics and science for all countries participating in the TIMSS 2019 less difficult assessment. They also were used to re-estimate student mathematics proficiency in TIMSS Numeracy 2015 for the countries that participated in that assessment, which was necessary to set the linear transformation that placed the TIMSS 2019 less difficult mathematics proficiency results on the TIMSS fourth grade mathematics trend scale.

Scaling the TIMSS 2019 less difficult assessment also required conditioning to enhance the reliability of student plausible values using student and parent context variables, as described for paperTIMSS earlier. Exhibit 12.29 provides details on the conditioning models used for proficiency estimation of the TIMSS 2019 less difficult data. The conditioning models for the 2015 data were relevant only for estimating mathematics proficiency for the TIMSS Numeracy 2015 data.

Exhibit 12.29: **Conditioning Models for the TIMSS 2019 Less Difficult Data**

Country	2019				TIMSS Numeracy 2015			
	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
Albania	2	557	221	81	—	—	—	—
Bahrain	—	—	—	—	3	637	221	77
Bosnia and Herzegovina	4	551	280	89	—	—	—	—
Indonesia	—	—	—	—	2	617	214	76
Iran, Islamic Rep. of	—	—	—	—	2	637	205	75
Jordan	—	—	—	—	2	637	334	90
Kosovo	2	557	224	81	—	—	—	—
Kuwait	3	539	221	80	3	629	185	72
Montenegro	2	563	253	85	—	—	—	—
Morocco	2	563	310	90	2	637	268	82
North Macedonia	3	563	163	72	—	—	—	—
Pakistan	4	563	199	80	—	—	—	—
Philippines	2	559	275	85	—	—	—	—
Saudi Arabia	3	557	272	85	—	—	—	—
South Africa (5)	3	563	320	90	3	533	301	90

Estimating fourth grade mathematics proficiency and science proficiency for the TIMSS 2019 less difficult data followed the same approach as the paperTIMSS 2019 fourth grade data described earlier, incorporating the TIMSS 2019 less difficult response data, item parameters, and conditioning models. A two-dimensional MGROUP model was used to estimate simultaneously overall mathematics proficiency and overall science proficiency. The same fourth grade multi-dimensional MGROUP models described for paperTIMSS were used to estimate proficiency in the fourth grade mathematics and science content and cognitive domains.

With respect to the TIMSS 2019 less difficult mathematics assessment, the item parameters estimated at the concurrent calibration stage also were used to generate mathematics plausible values for the TIMSS Numeracy 2015 assessment data. These TIMSS 2015 plausible values were used to establish the linear transformation necessary to place the TIMSS 2019 less difficult mathematics data on the appropriate TIMSS fourth grade mathematics trend scale. Setting this linear transformation was done in the same manner described earlier in equations (12.1) through (12.3). It required aligning the re-estimated TIMSS Numeracy 2015 student ability distribution with the TIMSS Numeracy 2015 ability distribution that was estimated and published back in 2015. This linear transformation was then applied to the TIMSS 2019 less difficult mathematics proficiency plausible values to place them on the TIMSS fourth grade mathematics trend scale. Exhibit 12.30 shows the transformation constants for the TIMSS 2019 less difficult mathematics assessment.

Exhibit 12.30: Transformation Constants for the TIMSS 2019 Less Difficult Mathematics Data

Less Difficult Mathematics	TIMSS Numeracy 2015 Published Scores		TIMSS Numeracy 2015 Re-scaled Scores		A_{ik}	B_{ik}
	Mean	Standard Deviation	Mean	Standard Deviation		
PV1	398.74801	105.35161	-0.05006	0.92525	404.44767	113.86313
PV2	398.36696	105.67865	-0.05089	0.92905	404.15595	113.74883
PV3	400.00943	104.28663	-0.04945	0.92667	405.57442	112.53916
PV4	398.59399	105.37049	-0.04897	0.92428	404.17690	114.00325
PV5	398.38464	105.37384	-0.04913	0.92296	403.99395	114.16995

No additional linear transformation was required for the fourth grade science data from the TIMSS 2019 less difficult assessment. Because of the fixed item parameter scaling approach applied to the science data, the regular TIMSS 2019 fourth grade science linear transformation constants shown in the science panel of Exhibit 12.12 were used to place the science plausible values of the TIMSS 2019 less difficult assessment on the TIMSS 2019 fourth grade science trend scale.

Scaling the eTIMSS Data

The main objective in this third phase of the scaling effort was to derive TIMSS 2019 student plausible values from the eTIMSS assessment data, suitable for reporting and publication, adjusting for any mode effect between the two assessment modes (see Chapter 11). The eTIMSS data and bridge data from the eTIMSS 2019 trend countries were submitted to a series of calibration models to estimate TIMSS 2019 student proficiency results from the eTIMSS assessment data relying on group equivalence between the eTIMSS and bridge samples and the presence of comparable items, that is, items that functioned equivalently under both modes of administration. Chapter 11 provides the conceptual framework and describes the models implemented and described in this section to address the presence of a mode effect.

Exhibits 12.31 through 12.34 show the numbers of items present in the eTIMSS 2019 calibrations by item type and content and cognitive domains for both grades and subjects. The bridge data consist of the paperTIMSS trend items. Consequently, the numbers for bridge items in Exhibits 12.31 through 12.34 match the numbers of trend items shown in Exhibits 12.3 through 12.6, respectively. They also match the numbers of eTIMSS trend items in their respective exhibits, with the exception of eighth grade science since one eTIMSS trend item (SE52134) was removed from scaling because it did not have suitable psychometric properties in its digital form (see Appendix 10E).

Exhibit 12.31: Mathematics Items for the eTIMSS 2019 Calibration—Grade 4

Item Type	Points	Bridge Items		eTIMSS Items					
				Trend		New		Total	
		Items	Points	Items	Points	Items	Points	Items	Points
Multiple Choice	1	42	42	42	42	26	26	68	68
Constructed Response	1	45	45	45	45	46	46	91	91
	2	5	10	5	10	7	14	12	24
Total		**92**	**97**	**92**	**97**	**79**	**86**	**171**	**183**

Items by Content and Cognitive Domains

Mathematics Content Domains	Bridge Items		eTIMSS Items					
			Trend		New		Total	
	Items	Points	Items	Points	Items	Points	Items	Points
Number	55	59	55	59	28	29	83	88
Measurement and Geometry	26	27	26	27	26	29	52	56
Data	11	11	11	11	25	28	36	39
Mathematics Cognitive Domains	**Bridge Items**		**eTIMSS Items**					
			Trend		**New**		**Total**	
	Items	**Points**	**Items**	**Points**	**Items**	**Points**	**Items**	**Points**
Knowing	34	34	34	34	25	25	59	59
Applying	40	42	40	42	34	38	74	80
Reasoning	18	21	18	21	20	23	38	44
Total	**92**	**97**	**92**	**97**	**79**	**86**	**171**	**183**

Exhibit 12.32: Science Items for the eTIMSS 2019 Calibration—Grade 4

Item Type	Points	Bridge Items		eTIMSS Items					
				Trend		New		Total	
		Items	Points	Items	Points	Items	Points	Items	Points
Multiple Choice	1	47	47	47	47	41	41	88	88
Constructed Response	1	45	45	45	45	31	31	76	76
	2	3	6	3	6	2	4	5	10
Total		**95**	**98**	**95**	**98**	**74**	**76**	**169**	**174**

Items by Content and Cognitive Domains

Science Content Domains	Bridge Items		eTIMSS Items					
			Trend		New		Total	
	Items	Points	Items	Points	Items	Points	Items	Points
Life Science	41	43	41	43	32	34	73	77
Physical Science	36	37	36	37	25	25	61	62
Earth Science	18	18	18	18	17	17	35	35

Science Cognitive Domains	Bridge Items		eTIMSS Items					
			Trend		New		Total	
	Items	Points	Items	Points	Items	Points	Items	Points
Knowing	40	43	40	43	29	30	69	73
Applying	34	34	34	34	30	30	64	64
Reasoning	21	21	21	21	15	16	36	37
Total	**95**	**98**	**95**	**98**	**74**	**76**	**169**	**174**

Exhibit 12.33: Mathematics Items for the eTIMSS 2019 Calibration—Grade 8

Item Type	Points	Bridge Items		eTIMSS Items					
				Trend		New		Total	
		Items	Points	Items	Points	Items	Points	Items	Points
Multiple Choice	1	62	62	62	62	28	28	90	90
Constructed Response	1	46	46	46	46	59	59	105	105
	2	6	12	6	12	5	10	11	22
Total		**114**	**120**	**114**	**120**	**92**	**97**	**206**	**217**

Items by Content and Cognitive Domains

Mathematics Content Domains	Bridge Items		eTIMSS Items					
			Trend		New		Total	
	Items	Points	Items	Points	Items	Points	Items	Points
Number	36	37	36	37	27	29	63	66
Algebra	30	31	30	31	31	31	61	62
Geometry	25	28	25	28	18	21	43	49
Data and Probability	23	24	23	24	16	16	39	40

Mathematics Cognitive Domains	Bridge Items		eTIMSS Items					
			Trend		New		Total	
	Items	Points	Items	Points	Items	Points	Items	Points
Knowing	35	35	35	35	29	31	64	66
Applying	57	60	57	60	39	39	96	99
Reasoning	22	25	22	25	24	27	46	52
Total	**114**	**120**	**114**	**120**	**92**	**97**	**206**	**217**

Exhibit 12.34: Science Items for the eTIMSS 2019 Calibration—Grade 8

Item Type	Points	Bridge Items		eTIMSS Items					
				Trend		New		Total	
		Items	Points	Items	Points	Items	Points	Items	Points
Multiple Choice	1	59	59	58	58	49	49	107	107
Constructed Response	1	48	48	48	48	34	34	82	82
	2	11	22	11	22	11	22	22	44
Total		**118**	**129**	**117**	**128**	**94**	**105**	**211**	**233**

Items by Content and Cognitive Domains

Science Content Domains	Bridge Items		eTIMSS Items					
			Trend		New		Total	
	Items	Points	Items	Points	Items	Points	Items	Points
Biology	39	48	39	48	36	41	75	89
Chemistry	22	23	21	22	21	24	42	46
Physics	30	30	30	30	22	24	52	54
Earth Science	27	28	27	28	15	16	42	44
Science Cognitive Domains	**Bridge Items**		**eTIMSS Items**					
			Trend		**New**		**Total**	
	Items	**Points**	**Items**	**Points**	**Items**	**Points**	**Items**	**Points**
Knowing	43	48	42	47	33	34	75	81
Applying	44	48	44	48	36	44	80	92
Reasoning	31	33	31	33	25	27	56	60
Total	**118**	**129**	**117**	**128**	**94**	**105**	**211**	**233**

Exhibit 12.35 shows the sample sizes for scaling the fourth grade eTIMSS 2019 and bridge data, both for item calibration and for proficiency estimation. Twenty-seven countries contributed bridge and eTIMSS data to the item calibration and 30 countries and 6 benchmarking participants were included in proficiency estimation. Although Austria was not a trend country and did not contribute to the fourth grade paperTIMSS concurrent calibration, they administered the bridge booklets and thus were included in the fourth grade eTIMSS 2019 item calibration.

Exhibit 12.35: Sample Sizes for Scaling the eTIMSS 2019 Grade 4 Data

Country	Item Calibration		Proficiency Estimation	
	eTIMSS	Bridge	eTIMSS	Bridge
Austria	4,464	1,964	4,464	1,964
Canada	13,653	1,604	13,653	1,604
Chile	4,174	1,612	4,174	1,612
Chinese Taipei	3,765	1,663	3,765	1,663
Croatia	3,785	1,472	3,785	1,472
Czech Republic	4,692	2,030	4,692	2,030
Denmark	3,227	1,432	3,227	1,432
England	3,396	1,242	3,396	1,242
Finland	4,730	1,983	4,730	1,983
France	4,186	1,948	4,186	1,948
Georgia	3,787	1,632	3,787	1,632
Germany	3,437	1,505	3,437	1,505
Hong Kong SAR	—	—	2,968	1,329
Hungary	4,571	1,778	4,571	1,778
Italy	3,741	1,921	3,741	1,921
Korea, Rep. of	3,893	1,541	3,893	1,541
Lithuania	3,741	1,587	3,741	1,587
Malta	—	—	3,630	—
Netherlands	3,355	1,295	3,355	1,295
Norway (5)	3,951	1,899	3,951	1,899
Portugal	4,300	1,612	4,300	1,612
Qatar	4,933	1,486	4,933	1,486
Russian Federation	4,022	2,128	4,022	2,128
Singapore	5,986	1,881	5,986	1,881
Slovak Republic	4,247	1,610	4,247	1,610
Spain	9,555	1,670	9,555	1,670
Sweden	3,965	1,697	3,965	1,697
Turkey (5)	—	—	4,028	—
United Arab Emirates	25,834	2,243	25,834	2,243
United States	8,776	1,652	8,776	1,652
Benchmarking Participants				
Ontario, Canada	—	—	3,830	—
Quebec, Canada	—	—	3,837	—
Moscow City, Russian Fed.	—	—	3,843	—
Madrid, Spain	—	—	3,390	—
Abu Dhabi, UAE	—	—	9,037	—
Dubai, UAE	—	—	7,265	—
TOTAL	**152,166**	**46,087**	**193,994**	**47,416**

Exhibit 12.36 shows the sample sizes for scaling the eighth grade eTIMSS 2019 and bridge data, both for item calibration and for proficiency estimation. Eighteen countries contributed bridge and eTIMSS data to the item calibration and 22 countries and 5 benchmarking participants were included in proficiency estimation.

Hong Kong SAR, despite being a trend country and having administered the bridge booklets, was excluded from the eTIMSS item calibration at both grades due to inconsistent mode differences at both grades and subjects. The response differences in Hong Kong SAR were more pronounced for science at both grades, particularly at the eighth grade, which ran counter to the general pattern of other eTIMSS countries where mathematics showed larger mode differences (see Exhibit 13.6 in Chapter 13).

Exhibit 12.36: Sample Sizes for Scaling the eTIMSS 2019 Grade 8 Data

Country	Item Calibration		Proficiency Estimation	
	eTIMSS	Bridge	eTIMSS	Bridge
Chile	4,115	1,526	4,115	1,526
Chinese Taipei	4,915	1,578	4,915	1,578
England	3,365	1,592	3,365	1,592
Finland	—	—	4,874	—
France	—	—	3,874	—
Georgia	3,315	1,314	3,315	1,314
Hong Kong SAR	—	—	3,265	1,423
Hungary	4,569	1,751	4,569	1,751
Israel	3,731	1,863	3,731	1,863
Italy	3,619	2,032	3,619	2,032
Korea, Rep. of	3,861	1,693	3,861	1,693
Lithuania	3,826	1,687	3,826	1,687
Malaysia	7,065	1,560	7,065	1,560
Norway (9)	4,575	2,018	4,575	2,018
Portugal	—	—	3,377	—
Qatar	3,884	1,490	3,884	1,490
Russian Federation	3,901	2,083	3,901	2,083
Singapore	4,853	1,871	4,853	1,871
Sweden	3,996	1,582	3,996	1,582
Turkey	4,077	1,819	4,077	1,819
United Arab Emirates	22,334	2,089	22,334	2,089
United States	8,698	1,484	8,698	1,484
Benchmarking Participants				
Ontario, Canada	—	—	3,776	—
Quebec, Canada	—	—	3,178	—
Moscow City, Russian Fed.	—	—	3,783	—
Abu Dhabi, UAE	—	—	8,204	—
Dubai, UAE	—	—	5,728	—
TOTAL	**98,699**	**31,032**	**138,758**	**32,455**

Identifying Invariant Items

As described earlier, an item equivalence study was carried out before the TIMSS 2019 assessment (Fishbein et al., 2018). This study led to the expectation that around 80 percent of the trend items could be considered comparable in terms of presentation and item content. That is, a large proportion of trend items, after being adapted to the digital interface for computer delivery, were from a visual and response requirement perspective deemed comparable to their paper counterpart. To confirm this comparability assessment, the starting point for scaling the eTIMSS 2019 data was the application of an interim item calibration model that made no assumption about the presence of a difference in mode of administration, thus relying exclusively on the group equivalence between the eTIMSS and bridge samples (see Chapter 9 for information about the samples). Combining eTIMSS and bridge data from all eTIMSS trend countries, item parameters were estimated for all eTIMSS 2019 items and then compared to their paperTIMSS counterparts. This "full non-invariance model" served as a baseline to provide statistical evidence of item equivalence, or invariance, between both modes of administration.

From the outset, many TIMSS 2019 items were expected to have similar behavior in both modes of administration based on the item equivalence study, in particular trend items that had been designed for paper-based administration in past TIMSS assessments. Some of the new eTIMSS 2019 items designed to capitalize on the digital environment of computer-based assessments were not expected to behave the same (Fishbein et al., 2018). Extensive analyses of item percent correct statistics and IRT parameters between eTIMSS and paperTIMSS, as well as RMSD statistics for the difference between paperTIMSS and eTIMSS ICC curves determined that three response input types showed more similarity in psychometric properties between modes and could be further analyzed for item equivalence. Consequently, the identification of equivalent or invariant items focused on the three major item types whose student responses were expected to be similar in both modes of administration based on detailed examination of items: traditional multiple-choice items, keyboard items, and number pad items.

Finalizing the groups of equivalent items was achieved, first, by using a modified version of the Root Mean Square Difference (RMSD) statistic, as described earlier. In the context of the full non-invariance model, the RMSD statistic measured the difference between an item's two empirical item characteristic curves, one based on the paperTIMSS item response data (including bridge for trend items) and the other based on the eTIMSS item response data. Appendices 12G through 12J show the item parameters estimated by the full non-invariance model for all eTIMSS items at both grades and both subjects, including the RMSD statistic for quantifying item invariance.[4] Items from the three major item types with RMSD values less than 0.1 were deemed suitable to serve as anchor items between modes.

All other items, including items with other input types (e.g., not multiple-choice, keyboard, or number pad), were left as non-equivalent items to have item parameters freely estimated in the final model.

4 The bridge item parameters are not presented in these Appendices since they were identical to the item parameters shown in Appendices 12A through 12D, respectively, for paperTIMSS trend items.

The group of equivalent items was further refined after estimating the adjusted model described below, where equivalent items had item parameters fixed to equal the paperTIMSS item parameters adjusted by a constant. After running the adjusted model, RMSD statistics for the fit of the empirical eTIMSS ICC curve to the theoretical eTIMSS ICC curve were examined. Any equivalent items with an RMSD greater than 0.1 were made non-equivalent for the subsequent model, so that, consistent with paperTIMSS, all eTIMSS items had good fit.

At the fourth grade, the full non-invariance model and resulting RMSD statistics identified 124 of 171 mathematics items as invariant. The results of the first adjusted model identified one item as having poor fit, resulting in 123 invariant items. In science, the full non-invariance model identified 148 of 169 items as invariant. The results of the first adjusted model identified one item as having poor fit, resulting in 147 invariant items. At the eighth grade, the full non-invariance model identified 170 of 206 mathematics items and 185 of 211 science items as invariant. In eighth grade mathematics, the first adjusted model identified three items as having poor model fit, resulting in 167 invariant items in the final model. In science, the adjusted model was estimated twice to finalize the invariant items—the first identified five poorly fitting items and the second identified two additional, resulting in 178 invariant items in the final model.

Exhibit 12.37 shows the numbers of equivalent and non-equivalent items in the final calibration models. The percentage of equivalent eTIMSS items ranged from 72 to 87 percent across fourth and eighth grades for mathematics and science. As could be anticipated, somewhat higher percentages of eTIMSS trend items were equivalent—ranging from 80 to 91 percent. Having a substantial percentage of equivalent items between paperTIMSS and eTIMSS strengthened the validity and interpretability of achievement results based on linking the two modes utilizing equivalent items as anchor, and estimating the mode adjustment based on the equivalence of the samples prior to mode assignment.

Exhibit 12.37: eTIMSS 2019 Achievement Items by Equivalence Classification

eTIMSS 2019 Fourth Grade Item Equivalence

Item Type		Mathematics			Science		
		Trend	New	Total	Trend	New	Total
Equivalent Items	Multiple Choice Items	41	24	65	47	39	86
	Keyboard Items	3	3	6	39	22	61
	Number Pad Items	30	22	52	—	—	—
All Equivalent Items		**74**	**49**	**123**	**86**	**61**	**147**
All Non–Equivalent Items		**18**	**30**	**48**	**9**	**13**	**22**
All Items		**92**	**79**	**171**	**95**	**74**	**169**
Percentage of Equivalent Items		**80%**	**62%**	**72%**	**91%**	**82%**	**87%**

eTIMSS 2019 Eighth Grade Item Equivalence

Item Type		Mathematics			Science		
		Trend	New	Total	Trend	New	Total
Equivalent Items	Multiple Choice Items	60	26	86	58	44	102
	Keyboard Items	9	10	19	47	26	73
	Number Pad Items	33	29	62	2	1	3
All Equivalent Items		**102**	**65**	**167**	**107**	**71**	**178**
All Non–Equivalent Items		**12**	**27**	**39**	**10**	**23**	**33**
All Items		**114**	**92**	**206**	**117**	**94**	**211**
Percentage of Equivalent Items		**89%**	**71%**	**81%**	**91%**	**76%**	**84%**

Estimating International Mode Effect Parameters for Equivalent Items

Dealing effectively with the two modes of administration in TIMSS 2019 required applying an overall mode adjustment constant to the difficulty parameter of invariant eTIMSS items. An extensive examination of percent correct statistics of paper bridge and eTIMSS trend items revealed there was a small but significant average international difference favoring paper bridge in each subject at both grades, with a smaller difference in science than mathematics (see Exhibit 13.6 in Chapter 13). These observed international mode differences required accounting for in the eTIMSS achievement results using an international adjustment of the invariant item parameters. While non-invariant eTIMSS items had distinct item parameters estimated for them, invariant items inherited the item parameters of their paperTIMSS counterparts with their location, or difficulty, parameter shifted by an international mode effect parameter to account for the average international difference (the international mode effect) between the paper and eTIMSS versions. Chapter 11 provides a description of how this adjustment is based on a simple extension of IRT models that utilizes the features of customary IRT linking methods.

Estimating an international mode effect parameter for each grade and subject was done using a weighted mean mode transformation. Simply stated, this adjustment parameter is the difference of the mean location parameters between the paper-based and computer-based versions of the eTIMSS items that were deemed invariant. The weighting factor assigned to each invariant eTIMSS item was the amount of information present in the two location parameters, quantified by the inverse of the two location parameter estimation error variances.

For each grade and subject combination, the international mode effect parameter δ_m was estimated over all invariant items (i = 1, …, N) as follows:

$$\delta_m = \frac{\sum_{i=1}^{N} w_i \left(b_{eT,i} - b_{pT,i} \right)}{\sum_{i=1}^{N} w_i} \qquad (12.4)$$

where

$b_{eT,i}$ is the estimated computer-based location parameter for invariant item i;

$b_{pT,i}$ is the estimated paper-based location parameter for invariant item i;

and the weight factor w_i is as follows:

$$w_i = \frac{1}{SD\left(b_{eT,i}\right)^2 + SD\left(b_{pT,i}\right)^2} \qquad (12.5)$$

where

SD($b_{eT,i}$) is the estimated standard deviation of the computer-based location parameter for invariant item i; and

SD($b_{pT,i}$) is the estimated standard deviation of the paper-based location parameter for invariant item i.

Thus, the shifted location parameter b_i^* for invariant item i was calculated as follows:

$$b_i^* = b_i + \delta_m \qquad (12.6)$$

Exhibit 12.38 shows the four estimated international mode effect parameters. The paper-based location parameters and their standard deviations were estimated from the paperTIMSS concurrent calibrations and are shown in Appendices 12A through 12D. The computer-based location parameters and their standard deviations were estimated from the eTIMSS full non-invariance calibration models and are shown in Appendices 12G through 12J.

Exhibit 12.38: eTIMSS 2019 Estimated International Mode Effect Parameters

		Mode Effect Parameter
Grade 4	Mathematics	0.09342
	Science	0.05894
Grade 8	Mathematics	0.10983
	Science	0.06766

eTIMSS 2019 Final Item Calibration

In the final eTIMSS 2019 item calibration models, which combined bridge and eTIMSS data, item parameters for the bridge items were fixed at their values from the paperTIMSS concurrent calibrations. Item parameters for eTIMSS items found to be invariant also were fixed at the values of their paperTIMSS counterparts with an additional international mode effect parameter estimated for each grade and subject added to the location parameters as shown in equation (12.6). Finally, item parameters for the remaining eTIMSS items—found to be non-invariant, i.e., affected by mode differences—were estimated freely. Appendices 12K through 12N show the item parameters for all eTIMSS 2019 items based on the final calibration models. These appendices include RMSD statistics to quantify model fit. All RMSD values are less than 0.10, the vast majority are less than 0.05, indicating good model fit.

Generating Plausible Values for the eTIMSS Data

Estimating student plausible values for the eTIMSS 2019 data followed the same general approach as for the paperTIMSS 2019 data. Conditioning was used to enhance the psychometric properties of student plausible values using student and parent context variables, as described for paperTIMSS earlier. Exhibits 12.39 and 12.40 provide details on the conditioning models used for proficiency estimation at the fourth grade and eighth grade, respectively.

Exhibit 12.39: Conditioning Models for the eTIMSS 2019 Grade 4 Data

Country	eTIMSS 2019			
	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
Austria	2	618	223	79
Canada	5	605	314	90
Chile	2	598	208	76
Chinese Taipei	2	622	188	75
Croatia	2	619	189	76
Czech Republic	2	618	234	82
Denmark	2	614	161	71
England	2	389	169	83
Finland	3	622	236	82
France	2	623	209	77
Georgia	2	620	189	73
Germany	2	623	171	73
Hong Kong SAR	3	623	148	69
Hungary	2	599	228	80
Italy	2	617	187	72
Korea, Rep. of	2	613	194	77
Lithuania	4	608	187	74
Malta	2	603	181	71
Netherlands	2	383	167	82
Norway (5)	4	543	197	80
Portugal	2	623	215	77
Qatar	3	622	246	81
Russian Federation	2	597	201	76
Singapore	2	599	299	90
Slovak Republic	3	623	212	77
Spain	6	616	317	90
Sweden	2	597	198	77
Turkey (5)	2	599	201	76
United Arab Emirates	5	623	326	90
United States	10	387	220	90
Benchmarking Participants				
Ontario, Canada	3	604	191	76
Quebec, Canada	3	605	191	76
Moscow City, Russian Fed.	2	591	192	75
Madrid, Spain	2	616	169	70
Abu Dhabi, UAE	3	623	318	90
Dubai, UAE	3	623	306	90

Exhibit 12.40: Conditioning Models for the eTIMSS 2019 Grade 8 Data

Country	eTIMSS 2019			
	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
Chile	2	638	205	79
Chinese Taipei	2	638	245	87
England	2	637	168	74
Finland	3	979	243	76
France	2	952	193	74
Georgia	2	978	165	63
Hong Kong SAR	3	639	163	78
Hungary	2	980	228	72
Israel	3	573	186	78
Italy	2	639	180	74
Korea, Rep. of	2	626	193	82
Lithuania	4	973	191	67
Malaysia	3	633	304	90
Norway (9)	4	595	228	85
Portugal	2	957	168	71
Qatar	3	639	194	77
Russian Federation	2	980	195	68
Singapore	2	615	242	86
Sweden	2	865	199	74
Turkey	2	639	203	77
United Arab Emirates	5	639	313	90
United States	10	634	300	90
Benchmarking Participants				
Ontario, Canada	3	639	188	77
Quebec, Canada	3	639	158	72
Moscow City, Russian Fed.	2	972	189	67
Abu Dhabi, UAE	3	639	311	90
Dubai, UAE	3	639	286	89

Mathematics proficiency and science proficiency for the eTIMSS 2019 data at both grades were estimated using the same psychometric models as for the paperTIMSS 2019 data, as described earlier in this chapter, incorporating the eTIMSS 2019 response data, item parameters, and conditioning models. A two-dimensional MGROUP model was used to estimate simultaneously overall mathematics proficiency and overall science proficiency. The same paperTIMSS multi-dimensional MGROUP models were used to estimate proficiency in the mathematics and science content and cognitive domains at both grades.

Because the eTIMSS 2019 item calibrations were anchored to the paperTIMSS concurrent calibrations via the bridge items, the scale transformations calculated and applied to the paperTIMSS data, as shown in Exhibits 12.12 and 12.13, were appropriate for placing the estimated eTIMSS 2019 student plausible values in mathematics and science on the TIMSS trend scales. These scale transformations also were applied to the eTIMSS 2019 mathematics and science plausible values in the content and cognitive domains at both grades.

Conclusion

Scaling the TIMSS 2019 achievement data was successful in estimating plausible values from its paperTIMSS and eTIMSS assessments, including the less difficult mathematics assessment at the fourth grade. The psychometric methods implemented and described in this chapter relied on past experience for scaling the paperTIMSS data. Scaling the eTIMSS data required careful consideration of any potential mode effect, which was dealt with effectively with the use of a paper bridge assessment administered in eTIMSS trend countries. The conceptual framework and mode effect models for linking the paperTIMSS and eTIMSS achievement data are described in Chapter 11.

The major outcome was the successful linking of all TIMSS 2019 assessments to the TIMSS trend scales such that results from the paper-based and the computer-based 2019 assessments can be compared directly without further need for adjustments. They also can be compared reliably with past TIMSS assessments. The high levels of comparability of the item parameters between modes of administration was established, and the mode-adjusted item parameters can be used in the population model to generate plausible values for estimating group level results and to examine the relation between the constructs of interest and additional variables. Chapter 13 provides valuable insight into the comparability of achievement results between both modes of administration.

References

Beaton, A. E. (1969). Criterion scaling of questionnaire items. *Socio-Economic Planning Sciences*, 2, 355–362.

Fishbein, B., Martin, M. O., Mullis, I. V. S., & Foy, P. (2018). The TIMSS 2019 item equivalence study: Examining mode effects for computer-based assessment and implications for measuring trends. *Large-scale Assessments in Education*, 6(1), 11. https://doi.org/10.1186/s40536-018-0064-z

Martin, M. O., Mullis, I. V. S., & Foy, P. (2017). TIMSS 2019 assessment design. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2019 assessment frameworks* (pp. 79-91). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/

Martin, M. O., Mullis, I. V. S., Foy, P., Brossman, B., & Stanco, G. M. (2012). Estimating linking error in PIRLS. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, 5, 35–47. Retrieved from https://www.ierinstitute.org/dissemination-area.html

Mazzeo, J., & von Davier, M. (2014). Linking scales in international large-scale assessments. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues and methods of data analysis* (pp. 229-258). Boca Raton: Chapman & Hall/CRC.

Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 assessment frameworks*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/

Muraki, E., & Bock, R. D. (1997). PARSCALE: IRT item analysis and test scoring for ratingscale data [computer software]. Chicago: Scientific Software.

Sheehan, K. M. (1985). M-Group: Estimation of group effects in multivariate models [computer software, Version 3.2] Princeton, NJ: Educational Testing Service.

Appendix 12A: Mathematics Item Parameters from the paperTIMSS 2019 Concurrent Calibration—Grade 4

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
Items Released in 2015:								
M01_01	M041004	0.050	—	0.969 (0.064)	-1.538 (0.116)	0.207 (0.055)		
M01_02	M041023	0.028	—	1.559 (0.091)	-0.857 (0.049)	0.184 (0.029)		
M01_03	M041034	0.024	—	0.928 (0.061)	-0.084 (0.064)	0.159 (0.027)		
M01_04	M041087	0.030	—	0.766 (0.032)	-0.203 (0.033)			
M01_05	M041124	0.027	—	0.938 (0.036)	-0.346 (0.029)			
M01_06A	M041302A	0.026	—	1.038 (0.062)	-0.705 (0.069)	0.161 (0.033)		
M01_06B	M041302B	0.040	—	0.602 (0.028)	-0.419 (0.042)			
M01_06C	M041302C	0.023	—	1.034 (0.039)	-0.415 (0.028)			
M01_07	M041254	0.025	—	0.671 (0.063)	0.211 (0.109)	0.222 (0.036)		
M01_08	M041153	0.020	—	1.020 (0.065)	0.072 (0.051)	0.139 (0.022)		
M01_09	M041132	0.033	—	0.476 (0.056)	0.907 (0.125)	0.131 (0.036)		
M01_10	M041165	0.040	—	0.352 (0.013)	0.396 (0.039)		-0.984 (0.083)	0.984 (0.088)
M01_11	M041174	0.030	—	1.077 (0.042)	-0.785 (0.032)			
M01_12	M041191	0.028	—	0.997 (0.075)	-1.157 (0.122)	0.336 (0.051)		
M02_01	M061272	0.023	—	0.817 (0.034)	0.080 (0.030)			
M02_02	M061243	0.030	—	0.468 (0.014)	-0.315 (0.030)		-0.992 (0.072)	0.992 (0.068)
M02_03	M061029	0.028	—	1.087 (0.064)	-0.367 (0.055)	0.145 (0.026)		
M02_04	M061031	0.028	—	1.411 (0.078)	0.493 (0.027)	0.069 (0.011)		
M02_05	M061050	0.022	—	1.275 (0.089)	0.500 (0.039)	0.181 (0.017)		
M02_06	M061167	0.045	—	0.692 (0.031)	-0.973 (0.048)			
M02_07	M061206	0.022	—	0.712 (0.063)	0.736 (0.067)	0.121 (0.024)		
M02_08A	M061265A	0.026	—	0.953 (0.039)	0.371 (0.028)			
M02_08B	M061265B	0.025	—	0.912 (0.090)	1.082 (0.059)	0.179 (0.018)		
M02_09	M061185	0.031	—	0.963 (0.059)	-0.575 (0.070)	0.145 (0.032)		
M02_10	M061239	0.041	—	1.356 (0.053)	-0.734 (0.027)			
M03_01	M051205	0.037	—	0.709 (0.031)	-0.367 (0.036)			
M03_02	M051039	0.029	—	1.082 (0.041)	-0.204 (0.025)			
M03_03	M051055	0.023	—	1.076 (0.046)	0.853 (0.031)			
M03_04	M051006	0.035	—	0.522 (0.019)	1.049 (0.039)		-0.539 (0.058)	0.539 (0.072)
M03_05	M051070	0.026	—	1.344 (0.108)	0.913 (0.038)	0.178 (0.014)		
M03_06	M051018	0.019	—	0.864 (0.077)	0.530 (0.068)	0.227 (0.025)		
M03_07	M051407	0.024	—	0.852 (0.065)	0.016 (0.076)	0.197 (0.030)		
M03_08	M051410	0.022	—	0.883 (0.069)	0.445 (0.060)	0.166 (0.023)		
M03_09	M051059	0.060	—	0.685 (0.032)	-1.412 (0.060)			
M03_10	M051093	0.021	—	0.768 (0.069)	0.658 (0.069)	0.164 (0.025)		
M03_11	M051134	0.034	—	1.187 (0.046)	0.332 (0.023)			
M03_12	M051077	0.027	—	1.117 (0.064)	0.104 (0.040)	0.085 (0.017)		

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
M05_01	M041291	0.043	—	0.689 (0.031)	-0.796 (0.045)			
M05_02	M041289	0.023	—	1.059 (0.084)	0.160 (0.065)	0.299 (0.026)		
M05_03	M041068	0.031	—	1.105 (0.067)	0.503 (0.036)	0.082 (0.015)		
M05_04A	M041065A	0.024	—	1.435 (0.099)	0.596 (0.034)	0.182 (0.015)		
M05_04B	M041065B	0.024	—	0.929 (0.042)	1.027 (0.039)			
M05_05	M041096	0.024	—	0.989 (0.066)	0.514 (0.044)	0.106 (0.018)		
M05_06	M041125	0.020	—	1.114 (0.089)	0.765 (0.045)	0.186 (0.017)		
M05_07	M041135	0.032	—	0.731 (0.066)	-0.732 (0.166)	0.358 (0.053)		
M05_08	M041257	0.027	—	0.728 (0.032)	0.246 (0.034)			
M05_09	M041268	0.024	—	1.731 (0.147)	0.979 (0.034)	0.227 (0.012)		
M05_10	M041151	0.037	—	0.483 (0.047)	-0.528 (0.214)	0.191 (0.060)		
M05_11	M041264	0.027	—	0.508 (0.062)	0.471 (0.164)	0.225 (0.046)		
M05_12	M041182	0.080	—	0.769 (0.037)	-1.784 (0.069)			
M05_13	M041200	0.049	—	0.447 (0.017)	-0.623 (0.039)		-0.221 (0.073)	0.221 (0.062)
M06_01	M051140	0.029	—	0.664 (0.057)	0.143 (0.101)	0.176 (0.035)		
M06_02	M051017	0.018	—	0.924 (0.089)	0.628 (0.069)	0.298 (0.024)		
M06_03	M051111	0.025	—	0.706 (0.034)	0.911 (0.045)			
M06_04	M051089	0.030	—	1.104 (0.045)	0.657 (0.027)			
M06_05	M051094	0.027	—	1.059 (0.078)	0.399 (0.051)	0.201 (0.021)		
M06_06	M051227	0.028	—	1.009 (0.046)	1.115 (0.039)			
M06_07	M051060	0.029	—	0.593 (0.059)	0.535 (0.105)	0.163 (0.034)		
M06_08Z	M051061Z	0.028	—	0.700 (0.033)	0.659 (0.040)			
M06_09	M051129	0.035	—	0.645 (0.055)	-0.311 (0.133)	0.203 (0.045)		
M06_10	M051236	0.040	—	0.846 (0.035)	0.035 (0.030)			
M06_11A	M051125A	0.086	—	0.796 (0.038)	-1.791 (0.067)			
M06_11B	M051125B	0.030	—	0.642 (0.064)	0.001 (0.138)	0.253 (0.043)		
M07_01	M041298	0.043	—	0.930 (0.065)	-0.782 (0.099)	0.253 (0.041)		
M07_02	M041007	0.027	—	0.807 (0.066)	0.321 (0.071)	0.182 (0.027)		
M07_03	M041280	0.023	—	0.731 (0.077)	0.780 (0.082)	0.233 (0.027)		
M07_04	M041059	0.036	—	0.689 (0.030)	-0.315 (0.036)			
M07_05	M041046	0.025	—	1.255 (0.074)	0.176 (0.037)	0.117 (0.017)		
M07_06	M041048	0.021	—	1.309 (0.105)	0.557 (0.044)	0.277 (0.018)		
M07_07	M041169	0.025	—	0.942 (0.069)	0.051 (0.066)	0.205 (0.027)		
M07_08	M041333	0.023	—	0.963 (0.072)	0.565 (0.049)	0.147 (0.019)		
M07_09	M041262	0.022	—	0.799 (0.082)	0.984 (0.068)	0.197 (0.022)		
M07_10	M041267	0.026	—	0.558 (0.029)	0.771 (0.052)			
M07_11	M041177	0.047	—	0.809 (0.055)	-0.472 (0.088)	0.158 (0.035)		
M07_12	M041271	0.048	—	0.860 (0.051)	-0.668 (0.076)	0.115 (0.032)		
M07_13A	M041276A	0.038	—	0.948 (0.038)	0.042 (0.027)			
M07_13B	M041276B	0.025	—	0.885 (0.038)	0.573 (0.032)			

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
Items Common in 2015 and 2019:								
MP01_01	MP51043	0.055	0.042	0.489 (0.017)	-0.065 (0.030)			
MP01_02	MP51040	0.027	0.025	1.162 (0.066)	-0.042 (0.052)	0.422 (0.020)		
MP01_03	MP51008	0.021	0.023	1.270 (0.034)	0.917 (0.018)			
MP01_04A	MP51031A	0.032	0.023	1.449 (0.034)	0.085 (0.013)			
MP01_04B	MP51031B	0.032	0.027	1.619 (0.038)	0.159 (0.012)			
MP01_05	MP51508	0.021	0.024	1.256 (0.030)	0.097 (0.014)			
MP01_06A	MP51216A	0.024	0.019	1.272 (0.062)	0.498 (0.029)	0.237 (0.013)		
MP01_06B	MP51216B	0.051	0.045	0.576 (0.039)	-0.876 (0.174)	0.270 (0.055)		
MP01_07	MP51221	0.053	0.042	0.571 (0.033)	-1.000 (0.144)	0.168 (0.051)		
MP01_08	MP51115	0.036	0.039	0.591 (0.052)	1.613 (0.066)	0.113 (0.017)		
MP01_09A	MP51507A	0.044	0.036	0.704 (0.021)	-0.657 (0.028)			
MP01_09B	MP51507B	0.019	0.019	1.101 (0.030)	0.768 (0.018)			
MP03_01	MP61026	0.054	0.054	0.904 (0.034)	-0.833 (0.053)	0.098 (0.025)		
MP03_02	MP61273	0.031	0.031	0.779 (0.039)	0.241 (0.049)	0.138 (0.020)		
MP03_03	MP61034	0.017	0.029	1.187 (0.030)	0.601 (0.016)			
MP03_04	MP61040	0.032	0.017	1.504 (0.065)	0.590 (0.021)	0.174 (0.010)		
MP03_05	MP61228	0.025	0.029	0.734 (0.015)	0.872 (0.017)		-0.255 (0.027)	0.255 (0.033)
MP03_06	MP61166	0.031	0.042	1.106 (0.027)	-0.356 (0.017)			
MP03_07	MP61171	0.033	0.028	1.310 (0.054)	-0.343 (0.036)	0.231 (0.019)		
MP03_08	MP61080	0.029	0.026	0.765 (0.022)	0.541 (0.022)			
MP03_09	MP61222	0.042	0.030	0.853 (0.056)	0.483 (0.057)	0.323 (0.020)		
MP03_10	MP61076	0.051	0.042	0.553 (0.018)	-0.697 (0.034)			
MP03_11	MP61084	0.024	0.027	1.010 (0.028)	0.777 (0.020)			
MP05_01	MP51206	0.059	0.049	0.591 (0.019)	-0.887 (0.035)			
MP05_02	MP51052	0.036	0.034	0.824 (0.048)	-0.010 (0.070)	0.297 (0.026)		
MP05_03	MP51049	0.032	0.029	1.341 (0.051)	0.037 (0.026)	0.143 (0.013)		
MP05_04	MP51045	0.039	0.039	1.066 (0.026)	-0.109 (0.016)			
MP05_05	MP51098	0.027	0.033	0.990 (0.047)	0.660 (0.030)	0.121 (0.012)		
MP05_06	MP51030	0.038	0.037	0.945 (0.028)	1.093 (0.025)			
MP05_07	MP51502	0.023	0.024	0.961 (0.057)	1.098 (0.035)	0.153 (0.012)		
MP05_08	MP51224	0.036	0.025	0.938 (0.051)	-0.013 (0.058)	0.301 (0.023)		
MP05_09	MP51207	0.019	0.027	0.799 (0.062)	0.794 (0.061)	0.341 (0.019)		
MP05_10	MP51427	0.018	0.023	1.053 (0.050)	0.659 (0.029)	0.136 (0.012)		
MP05_11	MP51533	0.022	0.026	1.056 (0.027)	0.075 (0.016)			
MP05_12	MP51080	0.037	0.034	0.957 (0.025)	-0.162 (0.018)			
MP06_01	MP61018	0.025	0.028	0.860 (0.023)	0.026 (0.019)			
MP06_02	MP61274	0.052	0.047	0.665 (0.037)	-0.686 (0.108)	0.197 (0.040)		
MP06_03	MP61248	0.046	0.039	0.828 (0.019)	0.346 (0.014)		0.401 (0.021)	-0.401 (0.023)
MP06_04	MP61039	0.026	0.022	1.068 (0.027)	0.233 (0.016)			
MP06_05	MP61079	0.036	0.027	1.238 (0.031)	0.637 (0.016)			

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
MP06_06	MP61179	0.025	0.026	1.141 (0.047)	-0.023 (0.033)	0.157 (0.016)		
MP06_07	MP61052	0.031	0.024	0.945 (0.038)	0.022 (0.035)	0.091 (0.016)		
MP06_08	MP61207	0.033	0.023	1.429 (0.053)	0.282 (0.021)	0.113 (0.010)		
MP06_09	MP61236	0.043	0.040	0.795 (0.022)	0.182 (0.020)			
MP06_10	MP61266	0.031	0.035	0.466 (0.010)	0.671 (0.021)		-0.844 (0.043)	0.844 (0.047)
MP06_11	MP61106	0.029	0.030	0.974 (0.046)	-0.126 (0.050)	0.219 (0.022)		
MP07_01	MP51401	0.031	0.039	0.784 (0.022)	0.447 (0.021)			
MP07_02	MP51075	0.025	0.025	1.297 (0.088)	1.044 (0.033)	0.326 (0.011)		
MP07_03	MP51402	0.026	0.032	0.917 (0.024)	0.377 (0.018)			
MP07_04	MP51226	0.023	0.020	1.302 (0.067)	0.588 (0.029)	0.270 (0.012)		
MP07_05	MP51131	0.038	0.029	0.731 (0.021)	-0.032 (0.021)			
MP07_06	MP51103	0.020	0.025	1.258 (0.060)	0.174 (0.034)	0.280 (0.016)		
MP07_07	MP51217	0.024	0.020	1.153 (0.029)	0.576 (0.016)			
MP07_08	MP51079	0.024	0.025	0.851 (0.023)	0.257 (0.019)			
MP07_09	MP51211	0.039	0.036	0.783 (0.045)	-0.198 (0.078)	0.274 (0.029)		
MP07_10	MP51102	0.028	0.023	0.948 (0.050)	0.699 (0.034)	0.159 (0.014)		
MP07_11	MP51009	0.048	0.044	0.777 (0.021)	-0.032 (0.020)			
MP07_12	MP51100	0.032	0.028	0.642 (0.041)	0.123 (0.085)	0.195 (0.029)		
MP09_01	MP61275	0.039	0.032	0.709 (0.039)	-0.570 (0.096)	0.212 (0.036)		
MP09_02	MP61027	0.057	0.040	0.893 (0.024)	-0.577 (0.022)			
MP09_03	MP61255	0.026	0.025	0.812 (0.016)	0.483 (0.013)		-0.182 (0.024)	0.182 (0.026)
MP09_04	MP61021	0.024	0.029	0.825 (0.023)	0.621 (0.021)			
MP09_05	MP61043	0.031	0.027	1.232 (0.030)	0.300 (0.014)			
MP09_06	MP61151	0.025	0.029	1.203 (0.046)	-0.159 (0.031)	0.132 (0.016)		
MP09_07	MP61172	0.018	0.028	1.520 (0.065)	0.756 (0.019)	0.123 (0.008)		
MP09_08	MP61223	0.046	0.039	0.725 (0.033)	-0.726 (0.078)	0.119 (0.032)		
MP09_09	MP61269	0.037	0.033	0.851 (0.037)	-0.464 (0.058)	0.130 (0.026)		
MP09_10A	MP61081A	0.030	0.030	1.002 (0.027)	0.721 (0.019)			
MP09_10B	MP61081B	0.039	0.043	0.719 (0.024)	1.055 (0.031)			
MP11_01	MP61178	0.030	0.033	0.829 (0.023)	0.048 (0.019)			
MP11_02	MP61246	0.025	0.027	0.953 (0.038)	0.052 (0.034)	0.090 (0.015)		
MP11_03	MP61271	0.056	0.038	0.618 (0.019)	-0.720 (0.031)			
MP11_04	MP61256	0.038	0.039	0.835 (0.023)	0.125 (0.019)			
MP11_05	MP61182	0.026	0.034	1.210 (0.035)	1.079 (0.021)			
MP11_06	MP61049	0.040	0.029	0.910 (0.048)	-0.482 (0.073)	0.310 (0.029)		
MP11_07	MP61232	0.032	0.035	0.970 (0.063)	0.660 (0.045)	0.321 (0.016)		
MP11_08	MP61095	0.029	0.026	0.915 (0.024)	-0.128 (0.018)			
MP11_09	MP61264	0.036	0.034	0.577 (0.013)	0.389 (0.017)		-0.100 (0.031)	0.100 (0.034)
MP11_10	MP61108	0.033	0.025	0.520 (0.042)	0.554 (0.103)	0.182 (0.032)		
MP11_11A	MP61211A	0.022	0.024	1.222 (0.030)	0.148 (0.014)			
MP11_11B	MP61211B	0.024	0.021	1.512 (0.078)	0.626 (0.026)	0.276 (0.011)		

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
MP13_01	MP61240	0.024	0.022	0.751 (0.022)	0.518 (0.022)			
MP13_02	MP61254	0.045	0.035	0.901 (0.023)	-0.007 (0.018)			
MP13_03	MP61244	0.018	0.030	0.931 (0.044)	-0.161 (0.053)	0.220 (0.023)		
MP13_04	MP61041	0.032	0.029	1.209 (0.072)	0.997 (0.030)	0.242 (0.011)		
MP13_05	MP61173	0.033	0.040	0.706 (0.020)	-0.303 (0.023)			
MP13_06	MP61252	0.027	0.020	1.157 (0.049)	0.590 (0.024)	0.113 (0.011)		
MP13_07	MP61261	0.032	0.036	1.261 (0.030)	0.115 (0.014)			
MP13_08	MP61224	0.031	0.024	0.825 (0.023)	0.541 (0.020)			
MP13_09	MP61077	0.037	0.033	0.830 (0.035)	-0.161 (0.046)	0.093 (0.020)		
MP13_10A	MP61069A	0.036	0.049	0.725 (0.021)	-0.791 (0.028)			
MP13_10B	MP61069B	0.032	0.035	0.732 (0.021)	-0.114 (0.022)			
Items Introduced in 2019:								
MP02_01	MP71219	—	0.059	0.709 (0.084)	-1.165 (0.256)	0.032 (0.121)		
MP02_02	MP71021	—	0.033	1.146 (0.110)	0.098 (0.065)	0.089 (0.031)		
MP02_03	MP71167	—	0.027	1.192 (0.081)	0.849 (0.049)			
MP02_04	MP71041	—	0.037	1.375 (0.131)	-0.313 (0.071)	0.143 (0.039)		
MP02_05	MP71162	—	0.033	0.479 (0.029)	1.451 (0.090)		-0.840 (0.112)	0.840 (0.149)
MP02_06	MP71078	—	0.041	0.715 (0.051)	-0.194 (0.054)			
MP02_07	MP71090	—	0.026	1.102 (0.124)	0.183 (0.080)	0.164 (0.037)		
MP02_08	MP71151	—	0.023	0.593 (0.028)	0.897 (0.050)		-1.236 (0.109)	1.236 (0.122)
MP02_09	MP71119	—	0.056	0.589 (0.049)	-1.308 (0.104)			
MP02_10A	MP71217A	—	0.052	0.909 (0.059)	-0.627 (0.052)			
MP02_11	MP71142	—	0.044	1.190 (0.073)	-0.435 (0.040)			
MP02_12	MP71204	—	0.024	1.334 (0.084)	0.475 (0.037)			
MP04_01	MP71013	—	0.033	1.155 (0.143)	-0.260 (0.116)	0.234 (0.056)		
MP04_02	MP71026	—	0.035	1.118 (0.076)	0.161 (0.041)			
MP04_03	MP71036	—	0.051	0.945 (0.067)	-0.538 (0.054)			
MP04_04	MP71040	—	0.021	1.391 (0.146)	0.338 (0.056)	0.103 (0.027)		
MP04_05	MP71068	—	0.034	0.492 (0.118)	0.419 (0.336)	0.113 (0.109)		
MP04_06A	MP71075A	—	0.023	1.256 (0.084)	0.266 (0.038)			
MP04_06B	MP71075B	—	0.024	1.471 (0.103)	0.647 (0.039)			
MP04_07	MP71080	—	0.027	1.595 (0.236)	0.637 (0.069)	0.303 (0.029)		
MP04_08	MP71211	—	0.035	0.632 (0.054)	0.080 (0.066)			
MP04_09	MP71178	—	0.027	0.762 (0.061)	0.508 (0.062)			
MP04_10B	MP71135B	—	0.036	0.681 (0.056)	-0.549 (0.072)			
MP04_11	MP71201	—	0.027	0.787 (0.069)	0.987 (0.080)			
MP04_12	MP71175	—	0.035	0.801 (0.052)	-0.085 (0.040)		0.560 (0.068)	-0.560 (0.062)
MP08_01	MP71018	—	0.036	1.371 (0.140)	0.177 (0.060)	0.160 (0.029)		
MP08_02	MP71009	—	0.045	1.248 (0.075)	0.209 (0.035)			
MP08_03	MP71037	—	0.035	0.908 (0.058)	0.158 (0.045)			
MP08_04	MP71051	—	0.025	1.170 (0.081)	0.913 (0.052)			

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
MP08_05	MP71064	—	0.044	0.724 (0.122)	0.756 (0.124)	0.155 (0.045)		
MP08_06	MP71169	—	0.029	1.317 (0.082)	0.506 (0.037)			
MP08_07	MP71083	—	0.040	1.202 (0.150)	0.507 (0.072)	0.209 (0.030)		
MP08_09	MP71184	—	0.027	1.635 (0.258)	1.059 (0.064)	0.244 (0.020)		
MP08_10	MP71141	—	0.029	0.957 (0.066)	0.733 (0.054)			
MP08_11	MP71194	—	0.086	0.743 (0.056)	-1.035 (0.074)			
MP08_12	MP71193	—	0.033	0.585 (0.028)	0.449 (0.043)		-0.802 (0.092)	0.802 (0.100)
MP08_13	MP71192	—	0.018	0.499 (0.024)	0.947 (0.057)		-2.150 (0.161)	2.150 (0.173)
MP10_02	MP71016	—	0.024	0.949 (0.066)	-0.049 (0.047)			
MP10_03	MP71163	—	0.027	1.762 (0.208)	0.966 (0.048)	0.076 (0.015)		
MP10_04	MP71045	—	0.024	1.087 (0.135)	0.257 (0.087)	0.163 (0.040)		
MP10_05	MP71213	—	0.024	0.941 (0.069)	0.435 (0.051)			
MP10_06	MP71070	—	0.038	0.354 (0.108)	-0.609 (1.060)	0.021 (0.287)		
MP10_07	MP71181	—	0.026	0.733 (0.060)	0.629 (0.068)			
MP10_08	MP71179	—	0.021	0.852 (0.072)	1.061 (0.078)			
MP10_09	MP71067	—	0.032	0.543 (0.028)	0.961 (0.058)		-1.542 (0.138)	1.542 (0.152)
MP10_10A	MP71147A	—	0.041	1.302 (0.087)	-0.429 (0.042)			
MP10_10B	MP71147B	—	0.026	0.886 (0.066)	0.298 (0.052)			
MP10_11	MP71189	—	0.056	0.903 (0.072)	-1.359 (0.088)			
MP10_12A	MP71187A	—	0.048	0.813 (0.063)	-0.932 (0.076)			
MP10_12B	MP71187B	—	0.060	0.676 (0.056)	-0.354 (0.068)			
MP12_01	MP71001	—	0.050	0.857 (0.103)	-1.079 (0.211)	0.087 (0.107)		
MP12_02	MP71010	—	0.039	0.694 (0.055)	-0.186 (0.062)			
MP12_03	MP71062	—	0.027	1.337 (0.208)	1.169 (0.073)	0.129 (0.021)		
MP12_04A	MP71216A	—	0.032	1.253 (0.082)	-0.382 (0.042)			
MP12_04B	MP71216B	—	0.037	0.831 (0.065)	0.295 (0.057)			
MP12_05	MP71117	—	0.035	0.646 (0.053)	-0.414 (0.070)			
MP12_06	MP71071	—	0.022	1.248 (0.198)	0.517 (0.094)	0.332 (0.037)		
MP12_07	MP71098	—	0.028	0.729 (0.047)	0.762 (0.048)		0.060 (0.068)	-0.060 (0.086)
MP12_08A	MP71134A	—	0.030	1.769 (0.165)	-0.046 (0.047)	0.092 (0.026)		
MP12_08B	MP71134B	—	0.036	1.454 (0.097)	0.254 (0.035)			
MP12_09	MP71202	—	0.036	0.681 (0.057)	-0.492 (0.071)			
MP12_10	MP71190	—	0.026	1.052 (0.073)	-0.112 (0.045)			
MP12_11	MP71218	—	0.025	1.098 (0.094)	1.196 (0.072)			
MP14_01	MP71024	—	0.025	0.921 (0.066)	0.160 (0.048)			
MP14_02	MP71008	—	0.028	1.118 (0.123)	-0.198 (0.095)	0.128 (0.047)		
MP14_03	MP71165	—	0.022	1.277 (0.154)	0.200 (0.076)	0.190 (0.037)		
MP14_04	MP71049	—	0.041	0.805 (0.060)	-0.370 (0.057)			
MP14_05	MP71063	—	0.028	1.050 (0.073)	0.220 (0.044)			
MP14_06	MP71079	—	0.019	1.179 (0.170)	0.696 (0.078)	0.192 (0.032)		
MP14_07	MP71081	—	0.034	1.007 (0.069)	-0.105 (0.046)			

Item		RMSD		Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
		2015	2019					
MP14_08	MP71094	—	0.024	1.007 (0.175)	0.648 (0.111)	0.280 (0.041)		
MP14_09	MP71177	—	0.029	0.606 (0.054)	0.389 (0.073)			
MP14_10	MP71206	—	0.042	0.681 (0.105)	-0.620 (0.282)	0.125 (0.114)		
MP14_11A	MP71138A	—	0.029	0.798 (0.060)	0.032 (0.054)			
MP14_11B	MP71138B	—	0.022	0.984 (0.076)	0.747 (0.058)			
MP14_12	MP71203	—	0.028	0.653 (0.139)	1.178 (0.143)	0.106 (0.047)		
MP14_13	MP71205	—	0.024	1.108 (0.079)	0.366 (0.044)			

Appendix 12B: Science Item Parameters from the paperTIMSS 2019 Concurrent Calibration— Grade 4

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
Items Released in 2015:								
S01_01	S041010	0.044	—	0.967 (0.067)	-0.786 (0.093)	0.259 (0.038)		
S01_02	S041034	0.034	—	0.647 (0.064)	-0.096 (0.134)	0.274 (0.042)		
S01_03	S041017	0.021	—	0.963 (0.108)	1.018 (0.061)	0.249 (0.021)		
S01_04	S041124	0.024	—	1.020 (0.102)	0.812 (0.057)	0.263 (0.022)		
S01_05	S041186	0.031	—	0.638 (0.036)	1.080 (0.057)			
S01_06	S041037	0.040	—	0.543 (0.020)	-0.186 (0.028)		-0.129 (0.055)	0.129 (0.050)
S01_07	S041119	0.028	—	1.115 (0.095)	-0.024 (0.076)	0.400 (0.029)		
S01_08	S041105	0.039	—	0.933 (0.060)	-0.130 (0.060)	0.141 (0.026)		
S01_10Z	S041149Z	0.024	—	0.606 (0.019)	1.033 (0.032)		-1.084 (0.062)	1.084 (0.072)
S01_11	S041032	0.079	—	0.827 (0.038)	-1.450 (0.061)			
S01_12	S041068	0.030	—	0.715 (0.035)	0.250 (0.033)			
S01_13	S041303	0.026	—	0.676 (0.085)	0.799 (0.102)	0.272 (0.033)		
S02_01	S061105	0.028	—	0.701 (0.078)	0.107 (0.135)	0.383 (0.039)		
S02_02	S061010	0.041	—	0.419 (0.026)	0.038 (0.053)			
S02_03	S061028	0.020	—	0.843 (0.119)	1.177 (0.083)	0.321 (0.024)		
S02_04	S061065	0.039	—	1.003 (0.067)	-0.227 (0.065)	0.198 (0.029)		
S02_05	S061130	0.031	—	0.797 (0.037)	0.420 (0.031)			
S02_06	S061081	0.028	—	0.926 (0.044)	0.828 (0.034)			
S02_07	S061060	0.041	—	0.829 (0.036)	-0.010 (0.030)			
S02_08	S061075	0.043	—	0.604 (0.050)	-0.260 (0.117)	0.145 (0.039)		
S02_09	S061031	0.034	—	0.992 (0.046)	0.875 (0.033)			
S02_10A	S061049A	0.047	—	0.773 (0.050)	-0.348 (0.074)	0.105 (0.028)		
S02_10B	S061049B	0.031	—	0.618 (0.057)	0.244 (0.101)	0.159 (0.034)		
S02_11	S061098	0.019	—	0.757 (0.105)	1.217 (0.088)	0.264 (0.026)		
S02_12	S061172	0.023	—	0.566 (0.034)	1.057 (0.061)			
S03_01	S051041	0.022	—	0.862 (0.098)	0.675 (0.081)	0.348 (0.027)		
S03_02	S051037	0.042	—	0.787 (0.035)	0.038 (0.031)			
S03_03	S051008	0.024	—	0.870 (0.045)	1.141 (0.046)			
S03_04	S051004	0.034	—	1.361 (0.088)	-0.099 (0.048)	0.248 (0.024)		
S03_05Z	S051026Z	0.028	—	0.532 (0.031)	0.752 (0.052)			
S03_06	S051130	0.022	—	0.530 (0.035)	1.478 (0.088)			
S03_07	S051114	0.024	—	1.155 (0.098)	0.607 (0.049)	0.251 (0.021)		
S03_08Z	S051121Z	0.045	—	0.414 (0.026)	0.097 (0.054)			
S03_09	S051147	0.027	—	0.841 (0.043)	0.978 (0.041)			
S03_10	S051105	0.031	—	1.005 (0.082)	-0.115 (0.081)	0.339 (0.032)		
S03_11	S051110	0.033	—	0.871 (0.066)	0.052 (0.072)	0.195 (0.029)		
S03_12	S051111	0.021	—	1.114 (0.101)	0.299 (0.068)	0.374 (0.027)		

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
S05_01	S041009	0.044	—	0.774 (0.059)	-0.931 (0.127)	0.246 (0.045)		
S05_02	S041223	0.028	—	1.000 (0.091)	0.432 (0.064)	0.297 (0.025)		
S05_03	S041026	0.037	—	0.536 (0.051)	0.222 (0.111)	0.126 (0.035)		
S05_04	S041177	0.026	—	0.424 (0.022)	1.054 (0.051)		0.377 (0.058)	-0.377 (0.078)
S05_05	S041183	0.052	—	0.646 (0.021)	0.207 (0.028)		1.136 (0.044)	-1.136 (0.045)
S05_06	S041008	0.026	—	1.171 (0.101)	0.666 (0.046)	0.237 (0.020)		
S05_08	S041195	0.018	—	0.618 (0.042)	1.664 (0.093)			
S05_09A	S041134A	0.028	—	0.804 (0.041)	0.953 (0.043)			
S05_09B	S041134B	0.045	—	0.768 (0.035)	0.162 (0.031)			
S05_09C	S041134C	0.025	—	0.756 (0.069)	0.471 (0.074)	0.184 (0.028)		
S05_10	S041191	0.024	—	0.841 (0.100)	0.802 (0.078)	0.309 (0.026)		
S05_11	S041107	0.047	—	0.394 (0.014)	-0.825 (0.047)		-0.797 (0.089)	0.797 (0.075)
S05_12	S041113	0.028	—	0.755 (0.037)	0.398 (0.033)			
S06_01	S051185	0.028	—	1.044 (0.075)	0.365 (0.048)	0.170 (0.021)		
S06_02	S051048	0.039	—	0.670 (0.025)	0.058 (0.023)		0.255 (0.043)	-0.255 (0.041)
S06_03	S051164	0.040	—	0.839 (0.051)	1.575 (0.072)			
S06_04	S051186	0.045	—	0.635 (0.030)	-1.065 (0.058)			
S06_05	S051137	0.050	—	0.661 (0.048)	-1.053 (0.139)	0.163 (0.047)		
S06_06	S051007	0.033	—	0.835 (0.036)	-0.131 (0.031)			
S06_07	S051087	0.032	—	1.020 (0.071)	-0.533 (0.080)	0.258 (0.034)		
S06_08Z	S051188Z	0.029	—	0.597 (0.031)	0.255 (0.039)			
S06_10	S051201	0.036	—	0.663 (0.033)	0.381 (0.036)			
S06_11	S051102	0.024	—	0.815 (0.068)	0.035 (0.085)	0.235 (0.032)		
S06_12	S051095	0.053	—	0.540 (0.028)	-0.429 (0.049)			
S07_01	S041027	0.074	—	0.715 (0.035)	-1.989 (0.083)			
S07_02	S041043	0.048	—	0.608 (0.030)	-0.664 (0.049)			
S07_03	S041050	0.025	—	0.459 (0.060)	0.656 (0.157)	0.181 (0.044)		
S07_04	S041070	0.025	—	0.797 (0.072)	0.411 (0.075)	0.212 (0.029)		
S07_05	S041006	0.037	—	0.453 (0.021)	0.601 (0.036)		0.354 (0.056)	-0.354 (0.065)
S07_06	S041052	0.027	—	0.918 (0.075)	-0.432 (0.103)	0.349 (0.038)		
S07_07	S041301	0.025	—	0.569 (0.033)	0.822 (0.052)			
S07_09	S041033	0.025	—	0.854 (0.043)	1.000 (0.042)			
S07_11	S041077	0.032	—	0.745 (0.035)	0.328 (0.033)			
S07_12	S041209	0.028	—	0.689 (0.070)	0.690 (0.080)	0.167 (0.029)		
S07_13	S041081	0.030	—	0.540 (0.019)	0.495 (0.028)		-0.440 (0.055)	0.440 (0.059)
S07_14	S041102	0.028	—	0.941 (0.070)	-0.248 (0.081)	0.244 (0.034)		

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
Items Common in 2015 and 2019:								
SP01_01	SP51054	0.049	0.035	0.934 (0.044)	-0.419 (0.058)	0.261 (0.024)		
SP01_02	SP51024	0.038	0.034	0.612 (0.021)	0.674 (0.028)			
SP01_03A	SP51132A	0.026	0.015	0.881 (0.031)	1.254 (0.032)			
SP01_03B	SP51132B	0.039	0.031	0.810 (0.027)	1.065 (0.029)			
SP01_04	SP51040	0.030	0.037	0.453 (0.018)	0.606 (0.036)			
SP01_05	SP51193	0.033	0.037	0.940 (0.048)	-0.126 (0.053)	0.274 (0.022)		
SP01_06	SP51063	0.018	0.029	1.148 (0.066)	0.754 (0.030)	0.222 (0.013)		
SP01_07	SP51012	0.027	0.029	0.989 (0.052)	0.268 (0.042)	0.253 (0.018)		
SP01_08	SP51115	0.054	0.032	1.090 (0.028)	0.146 (0.015)			
SP01_09	SP51180	0.034	0.037	0.880 (0.054)	0.057 (0.064)	0.360 (0.022)		
SP01_10	SP51106	0.018	0.026	1.024 (0.061)	0.721 (0.034)	0.215 (0.014)		
SP01_11	SP51148	0.025	0.038	1.049 (0.050)	0.043 (0.041)	0.241 (0.018)		
SP03_01	SP61141	0.028	0.021	1.235 (0.068)	0.519 (0.032)	0.300 (0.014)		
SP03_02	SP61023	0.034	0.035	0.770 (0.022)	0.015 (0.020)			
SP03_03	SP61054	0.046	0.042	0.479 (0.010)	0.643 (0.024)		1.489 (0.034)	-1.489 (0.043)
SP03_04	SP61007	0.040	0.035	0.647 (0.036)	-0.209 (0.079)	0.163 (0.028)		
SP03_05	SP61006	0.056	0.047	0.785 (0.022)	-0.650 (0.026)			
SP03_06	SP61108	0.025	0.026	1.050 (0.061)	0.233 (0.047)	0.352 (0.018)		
SP03_07	SP61109	0.029	0.032	0.583 (0.050)	0.710 (0.081)	0.235 (0.026)		
SP03_08	SP61080	0.024	0.024	0.968 (0.053)	0.297 (0.044)	0.264 (0.018)		
SP03_09	SP61088	0.028	0.017	0.672 (0.026)	1.417 (0.046)			
SP03_10	SP61151	0.033	0.031	0.952 (0.026)	0.440 (0.017)			
SP03_11	SP61150	0.045	0.043	0.624 (0.021)	0.408 (0.025)			
SP03_12	SP61169	0.024	0.032	1.077 (0.053)	0.079 (0.041)	0.268 (0.018)		
SP05_01	SP51044	0.034	0.035	0.503 (0.018)	0.201 (0.028)			
SP05_03	SP51003	0.044	0.038	0.711 (0.034)	-0.122 (0.054)	0.104 (0.021)		
SP05_04	SP51168	0.066	0.060	0.704 (0.021)	-0.475 (0.026)			
SP05_05	SP51010	0.039	0.038	0.766 (0.022)	0.076 (0.020)			
SP05_06	SP51035	0.024	0.030	1.249 (0.101)	1.298 (0.037)	0.236 (0.010)		
SP05_07	SP51059	0.035	0.035	0.584 (0.020)	0.104 (0.025)			
SP05_08	SP51142	0.036	0.023	0.802 (0.050)	0.598 (0.046)	0.199 (0.018)		
SP05_09A	SP51131A	0.030	0.034	1.014 (0.045)	-0.089 (0.041)	0.193 (0.019)		
SP05_09B	SP51131B	0.023	0.030	0.988 (0.055)	0.576 (0.035)	0.197 (0.015)		
SP05_10	SP51151	0.063	0.058	0.918 (0.026)	-1.120 (0.030)			
SP05_11	SP51157	0.030	0.022	0.739 (0.057)	0.999 (0.049)	0.190 (0.017)		
SP06_01	SP61071	0.051	0.043	0.335 (0.028)	-1.372 (0.337)	0.197 (0.071)		
SP06_02	SP61138	0.055	0.046	0.616 (0.020)	0.002 (0.024)			
SP06_03A	SP61016A	0.032	0.025	0.926 (0.050)	0.365 (0.041)	0.216 (0.017)		
SP06_03B	SP61016B	0.038	0.037	0.990 (0.027)	0.509 (0.017)			
SP06_04	SP61011	0.059	0.050	0.733 (0.021)	-0.536 (0.026)			

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SP06_06	SP61083	0.055	0.051	0.726 (0.021)	-1.025 (0.034)			
SP06_07	SP61034	0.026	0.028	0.788 (0.027)	1.088 (0.030)			
SP06_08	SP61044	0.030	0.030	0.740 (0.023)	0.551 (0.022)			
SP06_09A	SP61142A	0.034	0.034	0.623 (0.021)	0.351 (0.024)			
SP06_09B	SP61142B	0.027	0.026	0.788 (0.027)	1.034 (0.029)			
SP06_10A	SP61115A	0.033	0.031	1.468 (0.068)	0.346 (0.026)	0.264 (0.013)		
SP06_10B	SP61115B	0.037	0.028	1.345 (0.081)	0.662 (0.030)	0.328 (0.013)		
SP07_01	SP51161	0.036	0.032	0.488 (0.051)	1.007 (0.099)	0.217 (0.029)		
SP07_02	SP51051	0.027	0.021	1.391 (0.122)	1.370 (0.037)	0.281 (0.009)		
SP07_03Z	SP51138Z	0.033	0.034	0.583 (0.020)	0.313 (0.025)			
SP07_04	SP51194	0.024	0.027	0.970 (0.030)	1.014 (0.024)			
SP07_05	SP51029	0.022	0.023	0.518 (0.055)	1.220 (0.083)	0.202 (0.026)		
SP07_06	SP51077	0.046	0.041	0.747 (0.022)	-0.167 (0.022)			
SP07_07	SP51200	0.023	0.030	0.679 (0.025)	1.196 (0.037)			
SP07_08	SP51075	0.062	0.052	0.670 (0.020)	-0.586 (0.029)			
SP07_09	SP51065	0.037	0.041	0.870 (0.049)	-0.215 (0.070)	0.333 (0.026)		
SP07_10	SP51191	0.024	0.033	1.342 (0.065)	0.578 (0.025)	0.205 (0.012)		
SP07_11	SP51099	0.024	0.027	0.868 (0.049)	0.332 (0.047)	0.216 (0.019)		
SP07_12	SP51175	0.020	0.031	0.978 (0.030)	0.968 (0.023)			
SP09_01	SP61135	0.050	0.036	0.758 (0.041)	-0.598 (0.085)	0.268 (0.030)		
SP09_02	SP61069	0.044	0.041	0.400 (0.016)	-0.481 (0.041)			
SP09_03	SP61134	0.039	0.038	0.651 (0.036)	0.181 (0.060)	0.126 (0.022)		
SP09_04	SP61140	0.029	0.024	1.039 (0.064)	0.601 (0.039)	0.296 (0.016)		
SP09_05	SP61019	0.024	0.028	0.887 (0.028)	0.943 (0.024)			
SP09_06	SP61022	0.028	0.030	0.656 (0.044)	0.183 (0.079)	0.241 (0.026)		
SP09_07	SP61036	0.029	0.028	0.951 (0.029)	0.903 (0.022)			
SP09_08	SP61160	0.052	0.051	0.761 (0.022)	-0.954 (0.032)			
SP09_09	SP61159	0.063	0.054	0.826 (0.023)	-0.788 (0.027)			
SP09_10	SP61091	0.029	0.031	0.452 (0.014)	1.170 (0.032)		-0.176 (0.038)	0.176 (0.050)
SP09_11	SP61118	0.020	0.029	1.001 (0.056)	0.542 (0.036)	0.217 (0.016)		
SP09_12	SP61097	0.024	0.028	0.798 (0.055)	0.517 (0.056)	0.275 (0.021)		
SP11_01	SP61132	0.028	0.023	0.710 (0.048)	0.539 (0.058)	0.213 (0.021)		
SP11_02	SP61120	0.028	0.028	0.884 (0.047)	0.333 (0.043)	0.197 (0.018)		
SP11_03	SP61025	0.041	0.041	0.531 (0.018)	-0.366 (0.031)			
SP11_04A	SP61133A	0.028	0.023	1.370 (0.067)	0.245 (0.032)	0.326 (0.015)		
SP11_04B	SP61133B	0.028	0.030	1.701 (0.073)	0.792 (0.016)	0.114 (0.008)		
SP11_05	SP61074	0.044	0.035	0.772 (0.023)	0.219 (0.020)			
SP11_06	SP61093	0.063	0.056	0.761 (0.016)	-0.057 (0.016)		0.937 (0.026)	-0.937 (0.022)
SP11_07	SP61161	0.034	0.032	0.614 (0.021)	0.664 (0.028)			
SP11_08A	SP61042A	0.020	0.024	1.366 (0.077)	0.806 (0.025)	0.239 (0.011)		
SP11_08B	SP61042B	0.022	0.029	0.791 (0.047)	0.640 (0.042)	0.150 (0.017)		

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SP11_09A	SP61041A	0.032	0.033	0.871 (0.024)	0.116 (0.018)			
SP11_09B	SP61041B	0.044	0.041	0.719 (0.022)	0.167 (0.021)			
SP11_10	SP61155	0.044	0.040	0.735 (0.043)	-0.488 (0.093)	0.286 (0.032)		
SP13_02	SP61014	0.039	0.036	0.495 (0.018)	0.425 (0.030)			
SP13_03	SP61056	0.068	0.060	0.853 (0.023)	-0.738 (0.026)			
SP13_04	SP61015	0.060	0.055	0.692 (0.020)	-0.395 (0.025)			
SP13_05	SP61113	0.036	0.024	0.760 (0.025)	0.954 (0.028)			
SP13_06	SP61107	0.020	0.032	1.001 (0.054)	0.641 (0.032)	0.180 (0.014)		
SP13_07	SP61046	0.019	0.030	1.164 (0.068)	0.804 (0.029)	0.227 (0.012)		
SP13_08	SP61047	0.042	0.043	0.751 (0.043)	-0.518 (0.089)	0.313 (0.030)		
SP13_09	SP61048	0.027	0.032	1.300 (0.062)	0.509 (0.026)	0.221 (0.012)		
SP13_10	SP61096	0.029	0.019	1.100 (0.066)	0.730 (0.033)	0.257 (0.014)		
SP13_11	SP61124	0.026	0.028	0.590 (0.023)	1.242 (0.043)			
SP13_12	SP61116	0.039	0.031	0.681 (0.021)	0.159 (0.022)			
Items Introduced in 2019:								
SP02_01	SP71002	—	0.046	0.572 (0.047)	0.043 (0.065)			
SP02_02	SP71402	—	0.048	1.119 (0.135)	-0.253 (0.108)	0.299 (0.045)		
SP02_03	SP71017	—	0.035	0.710 (0.054)	0.271 (0.056)			
SP02_04	SP71077	—	0.036	1.100 (0.071)	0.226 (0.038)			
SP02_05	SP71072	—	0.022	1.212 (0.186)	0.786 (0.072)	0.232 (0.027)		
SP02_06	SP71054	—	0.042	0.941 (0.064)	0.213 (0.043)			
SP02_07	SP71115	—	0.028	0.848 (0.159)	0.797 (0.110)	0.249 (0.039)		
SP02_08	SP71140	—	0.043	0.703 (0.110)	-0.071 (0.182)	0.240 (0.062)		
SP02_09	SP71128	—	0.040	0.852 (0.133)	0.016 (0.152)	0.330 (0.052)		
SP02_10	SP71147	—	0.044	0.883 (0.113)	-0.224 (0.134)	0.241 (0.052)		
SP02_11A	SP71920A	—	0.038	0.802 (0.059)	0.344 (0.052)			
SP02_11B	SP71920B	—	0.031	0.956 (0.070)	0.612 (0.051)			
SP02_12	SP71268	—	0.023	0.941 (0.204)	1.253 (0.119)	0.203 (0.029)		
SP04_01	SP71013	—	0.049	0.852 (0.106)	-0.766 (0.181)	0.278 (0.067)		
SP04_02	SP71902	—	0.031	0.272 (0.040)	1.509 (0.259)			
SP04_03	SP71076	—	0.050	0.860 (0.091)	-0.563 (0.126)	0.134 (0.052)		
SP04_04	SP71041	—	0.036	0.778 (0.049)	0.977 (0.050)		0.021 (0.060)	-0.021 (0.084)
SP04_05	SP71046	—	0.033	0.803 (0.059)	0.442 (0.053)			
SP04_06	SP71095	—	0.040	0.654 (0.051)	0.225 (0.059)			
SP04_07	SP71129	—	0.042	0.855 (0.118)	-0.617 (0.192)	0.346 (0.066)		
SP04_08	SP71102	—	0.032	0.751 (0.059)	0.669 (0.064)			
SP04_09	SP71124	—	0.031	1.132 (0.159)	0.510 (0.079)	0.252 (0.032)		
SP04_10	SP71112	—	0.062	0.743 (0.094)	-1.183 (0.243)	0.216 (0.090)		
SP04_11	SP71265	—	0.030	0.708 (0.157)	0.628 (0.170)	0.341 (0.052)		
SP04_12	SP71223	—	0.059	0.548 (0.100)	-1.573 (0.565)	0.298 (0.161)		
SP08_02	SP71033	—	0.038	0.544 (0.123)	0.276 (0.275)	0.289 (0.076)		

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SP08_03	SP71065	—	0.049	0.670 (0.048)	-0.637 (0.066)			
SP08_04	SP71025	—	0.043	0.270 (0.095)	-0.329 (1.177)	0.000 (0.251)		
SP08_05	SP71081	—	0.027	0.949 (0.162)	1.051 (0.091)	0.157 (0.027)		
SP08_06	SP71056	—	0.034	0.635 (0.055)	0.853 (0.083)			
SP08_07	SP71145	—	0.046	0.516 (0.094)	-0.290 (0.313)	0.181 (0.091)		
SP08_08	SP71104	—	0.067	0.795 (0.053)	-0.850 (0.064)			
SP08_09	SP71144	—	0.044	0.515 (0.083)	-0.087 (0.229)	0.081 (0.073)		
SP08_10	SP71150	—	0.046	1.055 (0.065)	-0.402 (0.044)			
SP08_11	SP71201	—	0.031	1.048 (0.133)	-0.026 (0.106)	0.285 (0.043)		
SP08_12	SP71237	—	0.044	1.086 (0.070)	0.213 (0.039)			
SP08_13	SP71260	—	0.024	0.735 (0.145)	1.105 (0.119)	0.151 (0.036)		
SP10_01	SP71009	—	0.073	0.591 (0.033)	-0.470 (0.049)		1.127 (0.086)	-1.127 (0.069)
SP10_02	SP71093	—	0.049	0.727 (0.050)	-0.409 (0.057)			
SP10_03	SP71069	—	0.028	0.946 (0.213)	1.140 (0.118)	0.295 (0.032)		
SP10_04	SP71051	—	0.029	0.748 (0.058)	0.622 (0.062)			
SP10_05	SP71039	—	0.034	0.766 (0.101)	0.150 (0.117)	0.147 (0.045)		
SP10_06	SP71080	—	0.026	0.929 (0.170)	0.928 (0.099)	0.235 (0.033)		
SP10_07	SP71137	—	0.063	0.705 (0.050)	-0.283 (0.057)			
SP10_08	SP71103	—	0.035	0.815 (0.127)	0.275 (0.130)	0.259 (0.046)		
SP10_09	SP71106	—	0.040	0.629 (0.051)	0.442 (0.067)			
SP10_10	SP71100	—	0.029	0.910 (0.155)	0.275 (0.136)	0.374 (0.045)		
SP10_12	SP71220	—	0.030	0.998 (0.160)	0.732 (0.088)	0.232 (0.033)		
SP10_13	SP71254	—	0.030	0.704 (0.057)	0.652 (0.068)			
SP12_01	SP71031	—	0.043	0.630 (0.048)	0.021 (0.060)			
SP12_02	SP71090	—	0.041	0.767 (0.053)	0.011 (0.051)			
SP12_03	SP71048	—	0.024	1.433 (0.269)	1.191 (0.078)	0.220 (0.021)		
SP12_04	SP71071	—	0.028	0.990 (0.075)	0.875 (0.058)			
SP12_05	SP71011	—	0.045	1.209 (0.119)	-0.421 (0.085)	0.193 (0.040)		
SP12_06	SP71142	—	0.037	0.826 (0.149)	0.493 (0.133)	0.323 (0.044)		
SP12_07	SP71138	—	0.055	0.771 (0.052)	-0.619 (0.059)			
SP12_08	SP71127	—	0.040	0.920 (0.127)	0.034 (0.123)	0.288 (0.045)		
SP12_10	SP71500	—	0.035	0.792 (0.106)	0.333 (0.103)	0.140 (0.040)		
SP12_11	SP71257	—	0.033	1.395 (0.431)	1.384 (0.132)	0.431 (0.023)		
SP12_12	SP71222	—	0.038	0.906 (0.062)	0.231 (0.045)			
SP12_13	SP71252	—	0.030	0.988 (0.146)	0.352 (0.104)	0.290 (0.039)		
SP14_01	SP71063	—	0.050	0.407 (0.040)	-0.311 (0.090)			
SP14_02	SP71900	—	0.036	1.029 (0.149)	-0.022 (0.125)	0.373 (0.046)		
SP14_04	SP71043	—	0.024	0.644 (0.065)	1.381 (0.127)			
SP14_05	SP71005	—	0.062	1.021 (0.065)	-0.584 (0.049)			
SP14_06	SP71118	—	0.028	1.130 (0.170)	0.827 (0.073)	0.188 (0.027)		
SP14_07	SP71139	—	0.041	0.952 (0.143)	0.007 (0.135)	0.359 (0.048)		

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SP14_08	SP71114	—	0.059	0.784 (0.054)	-0.391 (0.054)			
SP14_09	SP71131	—	0.050	0.577 (0.047)	-0.028 (0.065)			
SP14_10	SP71152	—	0.029	1.235 (0.178)	0.479 (0.078)	0.300 (0.033)		
SP14_11	SP71218	—	0.056	0.795 (0.112)	-0.626 (0.205)	0.309 (0.070)		
SP14_12	SP71214	—	0.037	1.098 (0.123)	0.119 (0.076)	0.167 (0.035)		
SP14_13	SP71213	—	0.034	1.005 (0.081)	0.950 (0.063)			

Appendix 12C: Mathematics Item Parameters from the paperTIMSS 2019 Concurrent Calibration—Grade 8

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
Items Released in 2015:								
M01_01	M042182	0.018	—	1.565 (0.137)	0.177 (0.053)	0.360 (0.022)		
M01_02	M042081	0.032	—	0.838 (0.040)	0.581 (0.038)			
M01_03	M042049	0.038	—	1.031 (0.088)	-0.045 (0.083)	0.261 (0.034)		
M01_04	M042052	0.030	—	1.712 (0.106)	-0.132 (0.035)	0.126 (0.019)		
M01_05	M042076	0.031	—	1.049 (0.087)	0.408 (0.059)	0.179 (0.024)		
M01_06A	M042302A	0.026	—	0.987 (0.032)	0.317 (0.021)		-0.175 (0.039)	0.175 (0.041)
M01_06B	M042302B	0.027	—	0.984 (0.029)	0.411 (0.020)		-0.617 (0.048)	0.617 (0.050)
M01_06C	M042302C	0.035	—	0.510 (0.022)	1.723 (0.066)		-1.007 (0.087)	1.007 (0.115)
M01_07	M042100	0.031	—	1.181 (0.097)	0.042 (0.066)	0.258 (0.028)		
M01_08	M042202	0.017	—	1.471 (0.121)	0.392 (0.047)	0.260 (0.020)		
M01_09	M042240	0.025	—	1.319 (0.090)	0.131 (0.045)	0.141 (0.021)		
M01_10	M042093	0.022	—	1.710 (0.086)	1.091 (0.029)			
M01_11	M042271	0.028	—	1.111 (0.079)	0.156 (0.054)	0.132 (0.023)		
M01_12	M042268	0.017	—	1.519 (0.140)	1.053 (0.041)	0.168 (0.013)		
M01_13	M042159	0.075	—	0.453 (0.029)	-0.917 (0.075)			
M01_14	M042164	0.023	—	1.451 (0.062)	0.424 (0.025)			
M01_15	M042167	0.012	—	1.380 (0.064)	0.757 (0.029)			
M02_01	M062208	0.044	—	0.983 (0.042)	-0.180 (0.031)			
M02_02	M062153	0.024	—	0.897 (0.086)	0.495 (0.077)	0.210 (0.029)		
M02_03A	M062111A	0.033	—	1.326 (0.054)	0.095 (0.025)			
M02_03B	M062111B	0.018	—	1.673 (0.073)	0.591 (0.023)			
M02_04	M062237	0.018	—	1.636 (0.080)	1.024 (0.029)			
M02_05	M062314	0.023	—	1.072 (0.054)	1.182 (0.043)			
M02_06	M062074	0.021	—	0.908 (0.119)	1.172 (0.080)	0.276 (0.023)		
M02_07	M062183	0.031	—	0.949 (0.042)	0.245 (0.032)			
M02_08	M062202	0.039	—	1.136 (0.085)	-0.106 (0.066)	0.196 (0.030)		
M02_09	M062246	0.017	—	2.108 (0.194)	1.073 (0.033)	0.172 (0.011)		
M02_10	M062286	0.018	—	1.095 (0.044)	1.329 (0.031)		-0.179 (0.043)	0.179 (0.057)
M02_11	M062325	0.019	—	0.896 (0.126)	1.034 (0.093)	0.366 (0.025)		
M02_12	M062106	0.041	—	0.425 (0.064)	0.789 (0.222)	0.177 (0.057)		
M02_13	M062124	0.020	—	1.455 (0.103)	0.516 (0.037)	0.123 (0.015)		
M03_01	M052209	0.041	—	1.397 (0.091)	-0.170 (0.046)	0.148 (0.024)		
M03_02	M052142	0.028	—	1.004 (0.088)	0.761 (0.055)	0.145 (0.020)		
M03_03	M052006	0.016	—	1.266 (0.131)	0.939 (0.054)	0.273 (0.018)		
M03_04	M052035	0.029	—	1.480 (0.061)	0.249 (0.023)			
M03_05	M052016	0.036	—	1.467 (0.061)	0.367 (0.024)			
M03_06	M052064	0.025	—	1.296 (0.109)	0.512 (0.050)	0.233 (0.020)		

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
M03_07	M052126	0.017	—	1.790 (0.089)	1.067 (0.028)			
M03_08	M052103	0.040	—	0.964 (0.070)	0.063 (0.064)	0.122 (0.027)		
M03_09	M052066	0.026	—	1.304 (0.103)	0.367 (0.049)	0.213 (0.021)		
M03_10	M052041	0.034	—	1.226 (0.063)	1.273 (0.042)			
M03_11	M052057	0.053	—	0.661 (0.060)	-0.008 (0.124)	0.150 (0.044)		
M03_12	M052417	0.037	—	0.947 (0.041)	0.225 (0.032)			
M03_13	M052501	0.026	—	0.866 (0.042)	0.892 (0.043)			
M03_14	M052410	0.033	—	0.904 (0.096)	0.574 (0.085)	0.281 (0.029)		
M03_15	M052170	0.016	—	1.115 (0.130)	1.116 (0.064)	0.271 (0.019)		
M05_01	M042183	0.038	—	0.703 (0.062)	-0.152 (0.123)	0.163 (0.045)		
M05_02	M042060	0.025	—	1.318 (0.093)	-0.006 (0.050)	0.179 (0.024)		
M05_03	M042019	0.046	—	0.796 (0.037)	0.379 (0.038)			
M05_04	M042023	0.024	—	1.260 (0.053)	0.379 (0.027)			
M05_05	M042197	0.030	—	0.993 (0.047)	0.869 (0.039)			
M05_06	M042234	0.020	—	1.323 (0.093)	0.196 (0.045)	0.157 (0.020)		
M05_07	M042066	0.039	—	0.693 (0.034)	0.133 (0.040)			
M05_08	M042243	0.021	—	1.804 (0.114)	0.277 (0.029)	0.103 (0.013)		
M05_09	M042248	0.023	—	1.434 (0.062)	0.607 (0.026)			
M05_10Z	M042229Z	0.016	—	1.295 (0.044)	0.592 (0.019)		-0.119 (0.033)	0.119 (0.037)
M05_11A	M042080A	0.068	—	0.772 (0.037)	0.437 (0.040)			
M05_11B	M042080B	0.034	—	1.286 (0.068)	1.274 (0.041)			
M05_12	M042120	0.035	—	1.012 (0.087)	-0.090 (0.087)	0.266 (0.035)		
M05_13	M042203	0.025	—	1.404 (0.093)	0.005 (0.043)	0.140 (0.021)		
M05_14	M042264	0.025	—	0.795 (0.043)	1.192 (0.056)			
M05_15	M042255	0.055	—	0.661 (0.053)	-0.443 (0.127)	0.128 (0.047)		
M05_16	M042224	0.056	—	0.921 (0.040)	-0.185 (0.033)			
M06_01	M052017	0.028	—	1.167 (0.086)	0.006 (0.059)	0.185 (0.027)		
M06_02	M052217	0.019	—	1.371 (0.060)	0.667 (0.027)			
M06_03	M052021	0.019	—	1.035 (0.033)	0.566 (0.021)		-0.305 (0.041)	0.305 (0.045)
M06_04	M052095	0.016	—	1.606 (0.067)	0.390 (0.023)			
M06_05	M052094	0.019	—	1.188 (0.058)	1.067 (0.037)			
M06_06	M052131	0.013	—	1.130 (0.107)	0.730 (0.057)	0.233 (0.020)		
M06_07	M052090	0.019	—	1.161 (0.110)	0.776 (0.055)	0.213 (0.020)		
M06_08A	M052121A	0.029	—	0.994 (0.070)	0.197 (0.055)	0.100 (0.023)		
M06_08B	M052121B	0.022	—	1.810 (0.107)	1.439 (0.036)			
M06_09	M052042	0.020	—	0.873 (0.040)	0.460 (0.036)			
M06_10	M052047	0.024	—	1.126 (0.048)	0.248 (0.028)			
M06_11	M052044	0.024	—	1.581 (0.201)	1.115 (0.056)	0.391 (0.016)		
M06_12A	M052422A	0.034	—	0.754 (0.073)	-0.292 (0.146)	0.258 (0.052)		
M06_12B	M052422B	0.034	—	0.691 (0.060)	0.117 (0.104)	0.127 (0.038)		
M06_13	M052505	0.050	—	1.232 (0.096)	-0.860 (0.091)	0.262 (0.049)		

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
M07_01	M042015	0.048	—	0.863 (0.065)	-0.598 (0.107)	0.167 (0.048)		
M07_02	M042196	0.029	—	1.090 (0.069)	-0.042 (0.051)	0.088 (0.023)		
M07_03	M042194	0.039	—	1.195 (0.050)	-0.519 (0.029)			
M07_04A	M042114A	0.027	—	1.522 (0.062)	-0.108 (0.023)			
M07_04B	M042114B	0.035	—	1.553 (0.064)	0.169 (0.022)			
M07_05	M042112	0.043	—	0.871 (0.119)	1.140 (0.088)	0.313 (0.024)		
M07_06	M042109	0.016	—	1.527 (0.142)	0.968 (0.043)	0.214 (0.015)		
M07_07	M042050	0.015	—	1.074 (0.048)	0.628 (0.032)			
M07_08A	M042074A	0.037	—	1.019 (0.045)	0.487 (0.032)			
M07_08B	M042074B	0.037	—	0.954 (0.044)	0.662 (0.036)			
M07_08C	M042074C	0.023	—	1.690 (0.080)	0.922 (0.026)			
M07_09	M042151	0.032	—	0.818 (0.037)	-0.040 (0.035)			
M07_10	M042132	0.021	—	1.867 (0.185)	1.136 (0.038)	0.204 (0.012)		
M07_11	M042257	0.025	—	0.731 (0.071)	0.789 (0.077)	0.114 (0.026)		
M07_12	M042158	0.028	—	0.723 (0.081)	0.117 (0.144)	0.295 (0.046)		
M07_13	M042252	0.023	—	1.126 (0.099)	0.730 (0.053)	0.182 (0.020)		
M07_14	M042261	0.031	—	0.728 (0.060)	-0.140 (0.109)	0.139 (0.042)		
Items Common in 2015 and 2019:								
MP01_01	MP52024	0.027	0.024	1.646 (0.082)	0.441 (0.026)	0.232 (0.012)		
MP01_02A	MP52058A	0.043	0.045	1.281 (0.035)	-0.364 (0.017)			
MP01_02B	MP52058B	0.014	0.015	1.504 (0.043)	0.882 (0.018)			
MP01_03	MP52125	0.022	0.024	1.196 (0.054)	0.575 (0.027)	0.098 (0.011)		
MP01_04	MP52229	0.039	0.033	0.887 (0.025)	0.000 (0.021)			
MP01_05	MP52063	0.035	0.028	1.320 (0.068)	0.562 (0.030)	0.196 (0.013)		
MP01_06	MP52072	0.041	0.030	1.009 (0.049)	-0.003 (0.046)	0.146 (0.021)		
MP01_07A	MP52146A	0.042	0.030	0.859 (0.025)	0.182 (0.022)			
MP01_07B	MP52146B	0.021	0.021	1.533 (0.048)	1.153 (0.020)			
MP01_08	MP52092	0.022	0.026	1.244 (0.090)	1.514 (0.037)	0.151 (0.008)		
MP01_09	MP52046	0.023	0.025	1.125 (0.086)	1.477 (0.041)	0.188 (0.010)		
MP01_10	MP52083	0.018	0.018	1.501 (0.080)	0.882 (0.025)	0.169 (0.010)		
MP01_11	MP52082	0.034	0.030	1.202 (0.057)	0.161 (0.036)	0.174 (0.017)		
MP01_12	MP52161	0.042	0.033	1.187 (0.056)	-0.210 (0.044)	0.189 (0.022)		
MP01_13A	MP52418A	0.034	0.032	1.908 (0.089)	0.649 (0.020)	0.147 (0.009)		
MP01_13B	MP52418B	0.021	0.014	1.916 (0.100)	0.554 (0.023)	0.250 (0.011)		
MP03_01	MP62005	0.025	0.023	0.871 (0.061)	0.478 (0.064)	0.304 (0.022)		
MP03_02	MP62139	0.020	0.019	0.986 (0.028)	0.583 (0.021)			
MP03_03	MP62164	0.028	0.022	1.357 (0.061)	0.075 (0.031)	0.172 (0.015)		
MP03_04	MP62142	0.034	0.033	0.916 (0.026)	-0.261 (0.021)			
MP03_05	MP62084	0.017	0.022	1.393 (0.102)	1.553 (0.035)	0.144 (0.007)		
MP03_06	MP62351	0.022	0.027	0.804 (0.071)	1.405 (0.056)	0.207 (0.015)		
MP03_07	MP62223	0.031	0.034	1.420 (0.064)	-0.163 (0.033)	0.188 (0.018)		

Item		RMSD		Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
		2015	2019					
MP03_08	MP62027	0.027	0.021	0.772 (0.024)	0.556 (0.026)			
MP03_09	MP62174	0.020	0.018	1.403 (0.092)	0.862 (0.034)	0.319 (0.012)		
MP03_10	MP62244	0.019	0.028	0.971 (0.028)	0.462 (0.021)			
MP03_11	MP62261	0.020	0.026	1.889 (0.128)	1.460 (0.025)	0.132 (0.006)		
MP03_12	MP62300	0.029	0.029	0.752 (0.015)	0.412 (0.016)		-0.488 (0.033)	0.488 (0.035)
MP03_13	MP62254	0.024	0.028	0.744 (0.028)	1.490 (0.044)			
MP03_14A	MP62132A	0.037	0.045	1.185 (0.033)	-0.296 (0.018)			
MP03_14B	MP62132B	0.027	0.021	1.049 (0.070)	0.777 (0.044)	0.263 (0.016)		
MP05_01	MP52413	0.034	0.033	1.063 (0.060)	0.027 (0.054)	0.286 (0.023)		
MP05_02	MP52134	0.037	0.036	1.261 (0.053)	-0.270 (0.036)	0.130 (0.019)		
MP05_03	MP52078	0.026	0.026	0.990 (0.061)	0.884 (0.040)	0.183 (0.014)		
MP05_04	MP52034	0.020	0.033	1.216 (0.071)	0.549 (0.038)	0.279 (0.015)		
MP05_05A	MP52174A	0.032	0.032	1.088 (0.030)	0.213 (0.018)			
MP05_05B	MP52174B	0.024	0.019	1.118 (0.034)	1.021 (0.023)			
MP05_06	MP52130	0.019	0.015	1.232 (0.071)	0.970 (0.031)	0.173 (0.011)		
MP05_07	MP52073	0.021	0.018	1.385 (0.066)	0.473 (0.028)	0.174 (0.012)		
MP05_08	MP52110	0.019	0.020	1.464 (0.040)	0.653 (0.016)			
MP05_09	MP52105	0.025	0.026	1.172 (0.040)	1.428 (0.029)			
MP05_10	MP52407	0.012	0.020	1.344 (0.082)	0.359 (0.042)	0.378 (0.016)		
MP05_11	MP52036	0.034	0.029	0.730 (0.023)	0.439 (0.026)			
MP05_12	MP52502	0.045	0.042	1.165 (0.032)	-0.249 (0.018)			
MP05_13	MP52117	0.027	0.035	0.625 (0.028)	2.096 (0.075)			
MP05_14	MP52426	0.069	0.061	0.785 (0.040)	-0.797 (0.092)	0.142 (0.042)		
MP06_01	MP62150	0.039	0.044	1.111 (0.030)	-0.303 (0.019)			
MP06_02	MP62335	0.041	0.032	1.377 (0.061)	-0.106 (0.033)	0.175 (0.017)		
MP06_03	MP62219	0.019	0.020	2.050 (0.112)	0.851 (0.021)	0.218 (0.009)		
MP06_04	MP62002	0.027	0.032	0.703 (0.023)	0.620 (0.028)			
MP06_05	MP62149	0.031	0.032	1.089 (0.052)	0.507 (0.032)	0.111 (0.013)		
MP06_06	MP62241	0.024	0.017	1.708 (0.047)	0.633 (0.014)			
MP06_08	MP62105	0.026	0.027	0.757 (0.015)	0.850 (0.017)		-1.718 (0.062)	1.718 (0.064)
MP06_09	MP62040	0.027	0.023	0.769 (0.061)	0.947 (0.060)	0.224 (0.020)		
MP06_10	MP62288	0.024	0.022	0.776 (0.017)	1.140 (0.020)		-0.880 (0.041)	0.880 (0.047)
MP06_11	MP62173	0.025	0.027	1.119 (0.033)	0.812 (0.021)			
MP06_12	MP62133	0.014	0.019	1.315 (0.071)	0.616 (0.031)	0.214 (0.013)		
MP06_13A	MP62123A	0.021	0.027	1.562 (0.085)	0.354 (0.032)	0.306 (0.014)		
MP06_13B	MP62123B	0.020	0.025	1.444 (0.070)	0.704 (0.025)	0.138 (0.010)		
MP07_01	MP52079	0.026	0.028	0.966 (0.060)	0.424 (0.052)	0.271 (0.020)		
MP07_02	MP52204	0.031	0.026	0.871 (0.051)	0.396 (0.052)	0.180 (0.020)		
MP07_03	MP52364	0.045	0.050	1.177 (0.031)	-0.093 (0.017)			
MP07_04	MP52215	0.043	0.043	0.878 (0.025)	-0.248 (0.022)			
MP07_05	MP52147	0.016	0.020	1.572 (0.091)	0.762 (0.028)	0.275 (0.011)		

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
MP07_06	MP52067	0.032	0.033	1.063 (0.059)	0.067 (0.051)	0.263 (0.021)		
MP07_07	MP52068	0.016	0.015	1.417 (0.085)	1.264 (0.028)	0.132 (0.008)		
MP07_08	MP52087	0.022	0.028	1.622 (0.051)	1.139 (0.019)			
MP07_09	MP52048	0.021	0.024	1.019 (0.032)	1.148 (0.027)			
MP07_10	MP52039	0.018	0.018	1.235 (0.033)	0.272 (0.017)			
MP07_11	MP52208	0.018	0.018	2.264 (0.113)	1.111 (0.017)	0.081 (0.005)		
MP07_12A	MP52419A	0.048	0.048	0.888 (0.034)	-0.373 (0.042)	0.050 (0.018)		
MP07_12B	MP52419B	0.054	0.048	1.372 (0.055)	-0.672 (0.036)	0.104 (0.022)		
MP07_13	MP52115	0.031	0.021	1.738 (0.068)	0.348 (0.018)	0.080 (0.008)		
MP07_14	MP52421	0.038	0.031	0.824 (0.025)	0.641 (0.025)			
MP09_01	MP62329	0.069	0.074	0.793 (0.043)	-0.836 (0.103)	0.184 (0.046)		
MP09_02	MP62151	0.019	0.025	1.247 (0.035)	0.717 (0.019)			
MP09_03	MP62346	0.024	0.038	1.185 (0.033)	0.646 (0.019)			
MP09_04	MP62212	0.015	0.019	1.397 (0.077)	1.090 (0.026)	0.124 (0.008)		
MP09_05	MP62056	0.021	0.018	1.244 (0.039)	1.127 (0.023)			
MP09_06	MP62317	0.017	0.021	1.328 (0.038)	0.823 (0.018)			
MP09_07	MP62350	0.016	0.019	1.389 (0.099)	1.538 (0.034)	0.129 (0.007)		
MP09_08	MP62078	0.029	0.031	1.441 (0.040)	0.612 (0.016)			
MP09_09	MP62284	0.042	0.052	0.676 (0.056)	0.412 (0.100)	0.290 (0.031)		
MP09_10	MP62245	0.019	0.024	1.273 (0.069)	0.642 (0.031)	0.204 (0.013)		
MP09_11	MP62287	0.022	0.029	1.283 (0.044)	1.390 (0.027)			
MP09_12A	MP62345A	0.047	0.045	0.589 (0.016)	0.447 (0.021)		0.267 (0.034)	-0.267 (0.038)
MP09_13	MP62115	0.024	0.018	1.507 (0.108)	1.358 (0.031)	0.202 (0.009)		
MP11_01	MP62271	0.040	0.031	1.536 (0.081)	0.526 (0.029)	0.252 (0.012)		
MP11_02	MP62152	0.014	0.025	1.197 (0.032)	0.348 (0.017)			
MP11_03	MP62215	0.023	0.027	0.889 (0.019)	0.655 (0.015)		-0.188 (0.027)	0.188 (0.030)
MP11_04	MP62143	0.023	0.020	1.655 (0.047)	0.804 (0.016)			
MP11_05	MP62230	0.020	0.024	1.555 (0.112)	1.358 (0.031)	0.224 (0.008)		
MP11_06	MP62095	0.014	0.013	1.586 (0.080)	0.550 (0.026)	0.219 (0.011)		
MP11_07	MP62076	0.017	0.022	1.745 (0.089)	0.231 (0.028)	0.291 (0.014)		
MP11_08	MP62030	0.054	0.054	0.536 (0.020)	0.058 (0.032)			
MP11_09	MP62171	0.048	0.041	0.832 (0.042)	-0.145 (0.062)	0.128 (0.027)		
MP11_10	MP62301	0.018	0.024	1.080 (0.032)	0.998 (0.024)			
MP11_11	MP62194	0.049	0.039	1.025 (0.058)	-0.273 (0.066)	0.290 (0.028)		
MP11_12	MP62344	0.032	0.033	0.874 (0.028)	1.092 (0.030)			
MP11_13	MP62320	0.020	0.018	1.899 (0.077)	0.470 (0.018)	0.092 (0.008)		
MP11_14	MP62296	0.029	0.043	1.222 (0.033)	0.049 (0.017)			
MP13_01	MP62001	0.020	0.021	1.007 (0.077)	0.847 (0.050)	0.339 (0.016)		
MP13_02	MP62214	0.024	0.020	1.151 (0.031)	0.389 (0.018)			
MP13_03	MP62146	0.023	0.018	1.444 (0.068)	0.705 (0.023)	0.124 (0.010)		
MP13_04	MP62154	0.028	0.034	1.359 (0.036)	-0.086 (0.016)			

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
MP13_05	MP62067	0.037	0.038	1.159 (0.068)	0.096 (0.051)	0.335 (0.020)		
MP13_06	MP62341	0.027	0.035	0.932 (0.088)	1.643 (0.057)	0.218 (0.012)		
MP13_07	MP62242	0.032	0.024	1.269 (0.059)	0.175 (0.033)	0.171 (0.016)		
MP13_08A	MP62250A	0.025	0.024	1.207 (0.032)	0.138 (0.017)			
MP13_08B	MP62250B	0.021	0.025	1.403 (0.040)	0.817 (0.018)			
MP13_09	MP62170	0.087	0.083	0.535 (0.016)	0.921 (0.027)		0.551 (0.035)	-0.551 (0.046)
MP13_10	MP62192	0.017	0.021	1.044 (0.033)	1.120 (0.026)			
MP13_11	MP62072	0.046	0.045	1.024 (0.028)	0.110 (0.019)			
MP13_13	MP62120	0.029	0.022	1.250 (0.062)	0.465 (0.031)	0.166 (0.013)		
Items Introduced in 2019:								
MP02_01	MP72007	—	0.032	0.528 (0.034)	1.023 (0.082)		-0.407 (0.102)	0.407 (0.137)
MP02_02	MP72025	—	0.023	1.492 (0.214)	0.629 (0.067)	0.195 (0.024)		
MP02_03	MP72017	—	0.024	1.319 (0.106)	1.017 (0.065)			
MP02_04	MP72190	—	0.048	0.740 (0.057)	-0.038 (0.059)			
MP02_05	MP72068	—	0.044	1.285 (0.156)	-0.020 (0.080)	0.185 (0.038)		
MP02_06	MP72076	—	0.036	0.859 (0.127)	0.550 (0.100)	0.092 (0.038)		
MP02_07	MP72056	—	0.028	1.159 (0.082)	0.551 (0.053)			
MP02_08	MP72098	—	0.022	1.597 (0.122)	0.813 (0.049)			
MP02_09	MP72103	—	0.019	1.249 (0.176)	0.645 (0.073)	0.150 (0.026)		
MP02_10	MP72121	—	0.049	1.309 (0.084)	-0.264 (0.037)			
MP02_11	MP72180	—	0.027	0.671 (0.057)	0.634 (0.086)			
MP02_12	MP72198	—	0.024	1.233 (0.089)	0.610 (0.052)			
MP02_13	MP72227	—	0.028	1.507 (0.107)	0.578 (0.045)			
MP02_14	MP72170	—	0.033	0.875 (0.064)	0.071 (0.054)			
MP02_15	MP72209	—	0.018	1.057 (0.097)	1.360 (0.099)			
MP04_01	MP72178	—	0.025	0.933 (0.076)	1.032 (0.082)			
MP04_02	MP72234	—	0.028	0.959 (0.195)	0.942 (0.118)	0.258 (0.034)		
MP04_03	MP72020	—	0.040	0.639 (0.035)	-0.020 (0.042)		-0.266 (0.082)	0.266 (0.087)
MP04_04	MP72027	—	0.025	1.225 (0.150)	0.211 (0.074)	0.154 (0.033)		
MP04_05	MP72052	—	0.036	0.814 (0.080)	1.554 (0.133)			
MP04_06	MP72067	—	0.028	1.318 (0.164)	-0.004 (0.081)	0.218 (0.038)		
MP04_07A	MP72083A	—	0.049	1.406 (0.090)	-0.091 (0.036)			
MP04_07B	MP72083B	—	0.033	0.776 (0.116)	0.469 (0.113)	0.076 (0.044)		
MP04_08A	MP72108A	—	0.049	0.728 (0.056)	-0.011 (0.060)			
MP04_08B	MP72108B	—	0.031	1.025 (0.074)	0.513 (0.056)			
MP04_09	MP72181	—	0.024	1.211 (0.087)	0.634 (0.053)			
MP04_10	MP72126	—	0.041	0.679 (0.037)	0.900 (0.058)		-0.811 (0.102)	0.811 (0.123)
MP04_11	MP72164	—	0.020	0.858 (0.071)	0.981 (0.086)			
MP04_12A	MP72185A	—	0.025	1.612 (0.112)	0.447 (0.039)			
MP04_12B	MP72185B	—	0.027	1.506 (0.105)	0.429 (0.041)			

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
MP08_01	MP72002	—	0.026	1.517 (0.106)	0.542 (0.042)			
MP08_02	MP72188	—	0.022	1.280 (0.183)	0.770 (0.071)	0.138 (0.023)		
MP08_03	MP72035	—	0.023	1.132 (0.081)	0.551 (0.053)			
MP08_04	MP72055	—	0.023	1.391 (0.102)	0.707 (0.050)			
MP08_05	MP72222	—	0.047	0.603 (0.125)	0.651 (0.177)	0.098 (0.065)		
MP08_06	MP72090	—	0.025	1.211 (0.203)	0.877 (0.085)	0.198 (0.026)		
MP08_07	MP72233	—	0.022	1.075 (0.220)	0.692 (0.119)	0.367 (0.035)		
MP08_08A	MP72106A	—	0.047	1.068 (0.071)	-0.298 (0.043)			
MP08_08B	MP72106B	—	0.024	1.376 (0.097)	0.569 (0.046)			
MP08_08C	MP72106C	—	0.032	1.344 (0.104)	0.887 (0.058)			
MP08_09A	MP72128A	—	0.027	0.999 (0.073)	0.544 (0.058)			
MP08_09B	MP72128B	—	0.042	0.892 (0.058)	1.035 (0.058)		0.042 (0.065)	-0.042 (0.098)
MP08_10	MP72119	—	0.043	0.826 (0.063)	0.425 (0.064)			
MP08_11A	MP72153A	—	0.036	1.021 (0.072)	0.378 (0.053)			
MP08_11B	MP72153B	—	0.018	1.548 (0.140)	1.231 (0.068)			
MP08_12	MP72172	—	0.033	1.048 (0.116)	0.094 (0.075)	0.060 (0.033)		
MP10_01	MP72187	—	0.070	0.770 (0.057)	-0.336 (0.055)			
MP10_02	MP72022	—	0.020	1.631 (0.322)	1.070 (0.083)	0.279 (0.021)		
MP10_04	MP72045	—	0.025	1.307 (0.089)	0.461 (0.046)			
MP10_05	MP72049	—	0.039	0.986 (0.068)	0.059 (0.048)			
MP10_06	MP72069	—	0.052	1.335 (0.085)	-0.062 (0.038)			
MP10_07	MP72074	—	0.027	1.162 (0.090)	0.926 (0.066)			
MP10_08	MP72013	—	0.031	1.126 (0.152)	0.594 (0.075)	0.120 (0.027)		
MP10_09	MP72095	—	0.034	1.416 (0.098)	0.514 (0.045)			
MP10_10	MP72109	—	0.021	1.467 (0.122)	1.084 (0.062)			
MP10_11	MP72125	—	0.026	2.017 (0.268)	0.820 (0.050)	0.107 (0.015)		
MP10_12	MP72196	—	0.032	1.376 (0.096)	0.544 (0.046)			
MP10_13	MP72237	—	0.054	0.963 (0.136)	-0.045 (0.125)	0.194 (0.054)		
MP10_14	MP72232	—	0.049	0.787 (0.059)	-0.072 (0.056)			
MP10_15	MP72206	—	0.024	1.330 (0.120)	1.289 (0.079)			
MP12_01	MP72001	—	0.021	1.523 (0.109)	0.611 (0.046)			
MP12_02	MP72019	—	0.030	1.726 (0.118)	0.391 (0.037)			
MP12_03	MP72189	—	0.051	0.993 (0.162)	0.246 (0.120)	0.262 (0.046)		
MP12_04	MP72024	—	0.044	0.899 (0.069)	0.616 (0.068)			
MP12_05	MP72043	—	0.022	2.286 (0.337)	0.759 (0.050)	0.171 (0.016)		
MP12_06	MP72221	—	0.041	1.207 (0.173)	0.331 (0.084)	0.219 (0.034)		
MP12_07	MP72220	—	0.023	1.330 (0.259)	1.153 (0.097)	0.202 (0.021)		
MP12_08	MP72225	—	0.027	1.263 (0.088)	0.450 (0.048)			
MP12_09A	MP72110A	—	0.025	1.493 (0.107)	0.586 (0.046)			
MP12_09B	MP72110B	—	0.018	1.649 (0.130)	0.873 (0.051)			
MP12_10	MP72150	—	0.019	1.827 (0.346)	0.413 (0.083)	0.481 (0.027)		

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
MP12_11	MP72139	—	0.019	1.155 (0.093)	0.995 (0.072)			
MP12_12	MP72229	—	0.013	0.966 (0.067)	1.433 (0.069)		-1.025 (0.143)	1.025 (0.170)
MP12_13	MP72171	—	0.026	1.437 (0.099)	0.405 (0.043)			
MP12_14A	MP72211A	—	0.022	1.497 (0.213)	0.472 (0.068)	0.220 (0.027)		
MP14_01	MP72005	—	0.039	0.704 (0.113)	0.125 (0.169)	0.100 (0.068)		
MP14_02	MP72021	—	0.036	0.916 (0.065)	0.241 (0.054)			
MP14_03	MP72026	—	0.057	0.651 (0.055)	0.615 (0.085)			
MP14_04A	MP72041A	—	0.024	1.268 (0.083)	0.104 (0.040)			
MP14_04B	MP72041B	—	0.035	1.471 (0.098)	0.364 (0.040)			
MP14_05	MP72223	—	0.019	1.948 (0.294)	0.663 (0.058)	0.250 (0.021)		
MP14_06	MP72094	—	0.053	1.172 (0.077)	-0.033 (0.041)			
MP14_07	MP72059	—	0.024	1.363 (0.096)	0.616 (0.048)			
MP14_08	MP72080	—	0.016	1.587 (0.217)	0.874 (0.061)	0.118 (0.017)		
MP14_09	MP72081	—	0.028	0.961 (0.075)	0.861 (0.072)			
MP14_10	MP72140	—	0.031	0.837 (0.062)	0.344 (0.060)			
MP14_11	MP72120	—	0.022	1.146 (0.085)	0.779 (0.060)			
MP14_12	MP72131	—	0.018	1.349 (0.119)	1.286 (0.076)			
MP14_13	MP72147	—	0.013	1.697 (0.149)	1.172 (0.060)			
MP14_14	MP72154	—	0.034	1.325 (0.164)	0.106 (0.075)	0.189 (0.035)		
MP14_15	MP72192	—	0.032	1.009 (0.157)	0.444 (0.104)	0.209 (0.040)		
MP14_16	MP72161	—	0.035	1.164 (0.084)	0.618 (0.054)			

Appendix 12D: Science Item Parameters from the paperTIMSS 2019 Concurrent Calibration—Grade 8

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
Items Released in 2015:								
S01_01	S042258	0.018	—	0.791 (0.098)	1.025 (0.078)	0.186 (0.026)		
S01_02	S042005	0.024	—	0.353 (0.012)	0.568 (0.043)		-2.479 (0.126)	2.479 (0.131)
S01_03	S042016	0.019	—	1.022 (0.115)	1.220 (0.060)	0.135 (0.017)		
S01_04A	S042300A	0.030	—	1.349 (0.056)	0.064 (0.024)			
S01_04B	S042300B	0.056	—	0.549 (0.042)	1.743 (0.116)			
S01_04C	S042300C	0.026	—	1.132 (0.049)	0.132 (0.027)			
S01_05	S042319	0.022	—	1.345 (0.063)	0.762 (0.028)			
S01_06	S042068	0.020	—	1.305 (0.142)	1.022 (0.049)	0.220 (0.017)		
S01_07	S042216	0.026	—	1.045 (0.110)	0.414 (0.078)	0.338 (0.029)		
S01_08	S042249	0.024	—	0.771 (0.076)	0.474 (0.083)	0.163 (0.032)		
S01_09	S042094	0.024	—	0.832 (0.044)	0.761 (0.041)			
S01_10A	S042293A	0.040	—	0.917 (0.042)	-0.393 (0.035)			
S01_10B	S042293B	0.015	—	0.905 (0.065)	1.813 (0.092)			
S01_11	S042195	0.015	—	0.617 (0.047)	1.856 (0.118)			
S01_12	S042400	0.019	—	1.017 (0.053)	0.976 (0.040)			
S01_14	S042164	0.023	—	1.015 (0.087)	0.503 (0.056)	0.154 (0.024)		
S02_01	S062189	0.034	—	0.450 (0.022)	0.004 (0.038)		0.311 (0.069)	-0.311 (0.068)
S02_02	S062094	0.023	—	0.981 (0.087)	0.444 (0.063)	0.188 (0.026)		
S02_03	S062118	0.050	—	0.886 (0.041)	-0.004 (0.032)			
S02_04A	S062103A	0.022	—	1.125 (0.109)	0.562 (0.060)	0.265 (0.024)		
S02_04B	S062103B	0.027	—	0.723 (0.033)	1.006 (0.036)		0.218 (0.043)	-0.218 (0.060)
S02_05	S062010	0.028	—	0.513 (0.034)	0.830 (0.065)			
S02_06	S062253	0.024	—	0.876 (0.083)	0.852 (0.058)	0.115 (0.021)		
S02_07	S062051	0.023	—	0.905 (0.046)	0.776 (0.038)			
S02_08	S062044	0.019	—	1.091 (0.124)	1.326 (0.061)	0.121 (0.015)		
S02_09	S062046	0.032	—	0.896 (0.042)	0.166 (0.032)			
S02_10	S062149	0.029	—	0.442 (0.032)	0.908 (0.078)			
S02_11	S062268	0.035	—	0.997 (0.080)	-0.354 (0.091)	0.253 (0.039)		
S02_12	S062170	0.030	—	0.697 (0.088)	0.247 (0.146)	0.336 (0.044)		
S02_13	S062234	0.050	—	0.811 (0.033)	0.605 (0.027)		0.677 (0.037)	-0.677 (0.047)
S02_14	S062271	0.018	—	0.743 (0.110)	1.028 (0.101)	0.284 (0.031)		
S03_01	S052261	0.021	—	0.936 (0.096)	0.705 (0.066)	0.214 (0.025)		
S03_02Z	S052092Z	0.068	—	0.364 (0.019)	0.733 (0.054)		0.998 (0.077)	-0.998 (0.095)
S03_03A	S052263A	0.031	—	1.419 (0.077)	1.222 (0.037)			
S03_03B	S052263B	0.029	—	1.637 (0.080)	0.972 (0.027)			
S03_04	S052265	0.039	—	0.787 (0.043)	0.904 (0.047)			
S03_05	S052280	0.026	—	0.994 (0.095)	0.418 (0.071)	0.259 (0.028)		

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
S03_06	S052256	0.024	—	1.175 (0.104)	0.694 (0.048)	0.185 (0.020)		
S03_07Z	S052043Z	0.024	—	0.531 (0.035)	1.089 (0.074)			
S03_08	S052194	0.023	—	1.174 (0.113)	0.771 (0.051)	0.218 (0.021)		
S03_09	S052179	0.018	—	0.931 (0.114)	1.060 (0.069)	0.225 (0.023)		
S03_10	S052233	0.022	—	0.711 (0.048)	1.601 (0.086)			
S03_11	S052159	0.035	—	0.483 (0.077)	0.299 (0.255)	0.321 (0.061)		
S03_12A	S052289A	0.066	—	0.840 (0.067)	-0.998 (0.134)	0.226 (0.055)		
S03_12B	S052289B	0.027	—	0.658 (0.081)	0.848 (0.098)	0.174 (0.033)		
S03_12C	S052289C	0.035	—	0.847 (0.044)	0.729 (0.040)			
S05_01	S042053	0.026	—	1.216 (0.092)	-0.167 (0.064)	0.243 (0.030)		
S05_02	S042408	0.019	—	0.740 (0.040)	0.630 (0.042)			
S05_03	S042015	0.024	—	0.902 (0.094)	0.629 (0.072)	0.223 (0.027)		
S05_04	S042309	0.041	—	0.369 (0.062)	1.045 (0.231)	0.166 (0.054)		
S05_05A	S042049A	0.050	—	1.048 (0.047)	-0.596 (0.035)			
S05_05B	S042049B	0.033	—	1.187 (0.052)	0.220 (0.026)			
S05_06	S042182	0.043	—	0.660 (0.060)	-0.466 (0.144)	0.186 (0.050)		
S05_07	S042402	0.017	—	0.909 (0.051)	1.126 (0.050)			
S05_08A	S042228A	0.018	—	1.465 (0.077)	1.100 (0.033)			
S05_08B	S042228B	0.023	—	1.336 (0.057)	0.012 (0.024)			
S05_08C	S042228C	0.022	—	1.542 (0.068)	0.504 (0.022)			
S05_09	S042126	0.020	—	0.806 (0.099)	0.214 (0.129)	0.402 (0.039)		
S05_10	S042210	0.021	—	0.985 (0.185)	1.587 (0.112)	0.312 (0.020)		
S05_11	S042176	0.023	—	1.069 (0.051)	0.650 (0.032)			
S05_12	S042211	0.022	—	0.885 (0.042)	0.110 (0.032)			
S05_13	S042135	0.030	—	0.791 (0.039)	-0.238 (0.038)			
S05_14	S042257	0.016	—	0.543 (0.106)	1.313 (0.164)	0.304 (0.040)		
S06_01	S052003	0.022	—	0.911 (0.100)	0.122 (0.108)	0.393 (0.036)		
S06_02	S052071	0.018	—	1.310 (0.102)	0.469 (0.043)	0.172 (0.020)		
S06_03	S052246	0.019	—	0.909 (0.103)	0.850 (0.070)	0.227 (0.025)		
S06_04	S052276	0.032	—	0.739 (0.070)	-0.025 (0.112)	0.212 (0.040)		
S06_05A	S052303A	0.030	—	0.631 (0.070)	0.012 (0.150)	0.239 (0.048)		
S06_05B	S052303B	0.021	—	0.795 (0.041)	0.611 (0.039)			
S06_06	S052125	0.028	—	0.751 (0.125)	0.913 (0.123)	0.422 (0.032)		
S06_07	S052145	0.022	—	1.201 (0.053)	0.389 (0.026)			
S06_08	S052049	0.037	—	0.690 (0.032)	0.861 (0.034)		0.456 (0.043)	-0.456 (0.059)
S06_09	S052063	0.027	—	0.639 (0.074)	0.524 (0.112)	0.189 (0.038)		
S06_10	S052192	0.020	—	1.421 (0.088)	0.206 (0.035)	0.098 (0.017)		
S06_11	S052232	0.020	—	0.472 (0.093)	1.664 (0.178)	0.200 (0.039)		
S06_12	S052141	0.016	—	1.278 (0.062)	0.876 (0.031)			
S06_13	S052096	0.025	—	0.948 (0.084)	-0.089 (0.091)	0.277 (0.036)		
S06_14	S052116	0.029	—	0.870 (0.033)	0.205 (0.022)		0.172 (0.039)	-0.172 (0.039)

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
S06_15	S052110	0.019	—	0.861 (0.049)	1.084 (0.050)			
S07_01	S042042	0.046	—	0.638 (0.070)	-0.431 (0.192)	0.304 (0.058)		
S07_02	S042030	0.016	—	0.864 (0.049)	1.098 (0.050)			
S07_03	S042003	0.021	—	0.685 (0.108)	1.079 (0.112)	0.287 (0.033)		
S07_04	S042110	0.052	—	0.596 (0.054)	-0.549 (0.157)	0.160 (0.052)		
S07_05A	S042222A	0.018	—	0.961 (0.055)	1.238 (0.052)			
S07_05B	S042222B	0.017	—	0.957 (0.049)	0.842 (0.038)			
S07_05C	S042222C	0.030	—	0.823 (0.074)	-0.133 (0.105)	0.235 (0.040)		
S07_06	S042065	0.050	—	0.724 (0.072)	-0.925 (0.194)	0.335 (0.064)		
S07_07	S042280	0.022	—	1.268 (0.090)	0.202 (0.046)	0.155 (0.022)		
S07_08	S042088	0.030	—	0.666 (0.035)	0.108 (0.041)			
S07_09	S042218	0.016	—	1.339 (0.114)	0.453 (0.049)	0.246 (0.022)		
S07_10	S042104	0.025	—	0.862 (0.048)	1.065 (0.049)			
S07_11	S042064	0.024	—	0.765 (0.041)	0.712 (0.043)			
S07_12	S042273	0.024	—	1.171 (0.051)	0.243 (0.026)			
S07_13	S042301	0.027	—	0.820 (0.040)	0.102 (0.034)			
S07_14	S042312	0.045	—	0.372 (0.050)	-0.414 (0.352)	0.219 (0.077)		
S07_15	S042217	0.022	—	1.769 (0.158)	0.717 (0.036)	0.246 (0.016)		
S07_16	S042406	0.018	—	1.060 (0.052)	0.710 (0.033)			
Items Common in 2015 and 2019:								
SP01_01	SP52006	0.048	0.050	0.635 (0.017)	-0.098 (0.019)		0.620 (0.034)	-0.620 (0.030)
SP01_02	SP52069	0.023	0.029	0.984 (0.072)	0.601 (0.051)	0.325 (0.018)		
SP01_03	SP52012	0.020	0.026	0.947 (0.051)	0.342 (0.042)	0.163 (0.018)		
SP01_04	SP52021	0.019	0.028	1.029 (0.031)	0.638 (0.020)			
SP01_05Z	SP52095Z	0.040	0.038	0.505 (0.020)	-0.198 (0.035)			
SP01_06	SP52134	0.024	0.033	2.121 (0.201)	1.373 (0.029)	0.296 (0.009)		
SP01_07	SP52054	0.047	0.038	0.749 (0.024)	-0.380 (0.027)			
SP01_08	SP52150	0.020	0.030	0.787 (0.067)	1.170 (0.051)	0.181 (0.017)		
SP01_09A	SP52243A	0.028	0.029	0.624 (0.022)	0.373 (0.028)			
SP01_09B	SP52243B	0.032	0.025	0.769 (0.025)	0.394 (0.024)			
SP01_09C	SP52243C	0.029	0.027	0.671 (0.061)	1.026 (0.065)	0.200 (0.022)		
SP01_10	SP52206	0.022	0.022	1.127 (0.063)	0.478 (0.036)	0.207 (0.016)		
SP01_11A	SP52112A	0.031	0.037	0.672 (0.046)	-0.042 (0.095)	0.221 (0.033)		
SP01_11B	SP52112B	0.026	0.028	0.992 (0.031)	0.764 (0.022)			
SP01_12	SP52294	0.034	0.026	1.085 (0.054)	-0.084 (0.045)	0.206 (0.021)		
SP03_01	SP62055	0.040	0.032	0.962 (0.067)	-0.088 (0.079)	0.438 (0.026)		
SP03_02	SP62007	0.022	0.022	1.176 (0.064)	0.457 (0.034)	0.205 (0.015)		
SP03_03	SP62275	0.046	0.023	0.888 (0.029)	0.786 (0.024)			
SP03_04	SP62225	0.022	0.022	1.004 (0.098)	1.334 (0.050)	0.259 (0.014)		
SP03_05	SP62111	0.033	0.034	0.587 (0.016)	0.516 (0.020)		0.033 (0.034)	-0.033 (0.038)
SP03_06A	SP62116A	0.027	0.025	1.164 (0.034)	0.529 (0.017)			

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SP03_06B	SP62116B	0.019	0.029	1.319 (0.041)	0.859 (0.018)			
SP03_06C	SP62116C	0.021	0.027	0.946 (0.035)	1.247 (0.033)			
SP03_07	SP62262	0.017	0.023	0.891 (0.080)	1.063 (0.051)	0.277 (0.017)		
SP03_08	SP62035	0.023	0.022	1.076 (0.077)	1.029 (0.036)	0.199 (0.013)		
SP03_09	SP62144	0.061	0.062	0.725 (0.040)	-0.600 (0.090)	0.163 (0.035)		
SP03_10	SP62162	0.023	0.025	0.777 (0.027)	0.813 (0.028)			
SP03_11	SP62233	0.021	0.015	0.927 (0.077)	0.753 (0.055)	0.343 (0.019)		
SP03_13	SP62171	0.039	0.031	0.384 (0.048)	0.825 (0.188)	0.185 (0.047)		
SP05_01	SP52076	0.030	0.034	0.934 (0.059)	0.343 (0.052)	0.257 (0.021)		
SP05_02	SP52272	0.037	0.044	1.130 (0.031)	-0.074 (0.018)			
SP05_03A	SP52085A	0.020	0.024	1.038 (0.036)	1.164 (0.028)			
SP05_03B	SP52085B	0.045	0.045	1.034 (0.029)	-0.059 (0.019)			
SP05_04	SP52094	0.026	0.028	0.614 (0.024)	0.963 (0.038)			
SP05_05	SP52248	0.021	0.021	1.188 (0.148)	1.547 (0.061)	0.364 (0.012)		
SP05_06	SP52146	0.031	0.026	1.023 (0.030)	0.343 (0.019)			
SP05_07	SP52282	0.028	0.028	0.828 (0.059)	0.790 (0.048)	0.185 (0.018)		
SP05_08	SP52299	0.027	0.025	1.224 (0.072)	0.325 (0.041)	0.309 (0.017)		
SP05_09	SP52144	0.016	0.026	1.160 (0.072)	0.642 (0.036)	0.249 (0.015)		
SP05_10	SP52214	0.032	0.028	0.996 (0.029)	0.288 (0.019)			
SP05_12	SP52101	0.037	0.026	0.563 (0.023)	0.975 (0.041)			
SP05_13	SP52113	0.027	0.020	1.565 (0.089)	0.529 (0.029)	0.292 (0.013)		
SP05_14	SP52107	0.022	0.021	1.000 (0.084)	1.260 (0.043)	0.197 (0.013)		
SP06_01	SP62090	0.038	0.038	1.011 (0.061)	0.112 (0.055)	0.304 (0.022)		
SP06_02	SP62274	0.050	0.050	0.577 (0.015)	0.811 (0.024)		1.149 (0.032)	-1.149 (0.044)
SP06_03	SP62284	0.047	0.052	0.375 (0.042)	0.410 (0.211)	0.172 (0.050)		
SP06_04A	SP62098A	0.036	0.035	0.639 (0.016)	0.432 (0.018)		-0.050 (0.033)	0.050 (0.035)
SP06_04B	SP62098B	0.016	0.023	0.798 (0.023)	1.269 (0.024)		-0.091 (0.029)	0.091 (0.041)
SP06_05	SP62032	0.042	0.037	1.742 (0.171)	1.436 (0.036)	0.287 (0.009)		
SP06_06	SP62043	0.033	0.016	0.907 (0.031)	0.914 (0.026)			
SP06_07	SP62158	0.034	0.032	0.697 (0.062)	0.610 (0.082)	0.299 (0.026)		
SP06_08	SP62159	0.035	0.027	0.983 (0.056)	0.333 (0.044)	0.204 (0.019)		
SP06_09	SP62005	0.026	0.032	1.250 (0.036)	0.598 (0.017)			
SP06_10	SP62075	0.019	0.030	0.990 (0.074)	0.702 (0.049)	0.314 (0.018)		
SP06_11	SP62004	0.022	0.025	1.806 (0.095)	0.817 (0.020)	0.173 (0.009)		
SP06_12	SP62175	0.054	0.043	0.739 (0.025)	0.607 (0.026)			
SP06_13A	SP62173A	0.032	0.035	0.702 (0.024)	0.266 (0.025)			
SP06_13B	SP62173B	0.020	0.020	0.808 (0.100)	1.794 (0.086)	0.203 (0.014)		
SP07_01A	SP52090A	0.042	0.036	0.494 (0.062)	0.472 (0.186)	0.393 (0.041)		
SP07_01B	SP52090B	0.041	0.022	0.609 (0.030)	1.894 (0.075)			
SP07_02	SP52262	0.030	0.026	0.694 (0.060)	0.843 (0.066)	0.227 (0.023)		
SP07_03	SP52267	0.024	0.029	0.988 (0.064)	0.695 (0.041)	0.216 (0.016)		

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SP07_04	SP52273	0.035	0.035	0.638 (0.018)	0.866 (0.022)		0.174 (0.031)	-0.174 (0.039)
SP07_05Z	SP52015Z	0.043	0.043	0.830 (0.025)	-0.301 (0.024)			
SP07_06	SP52051	0.028	0.026	1.005 (0.032)	0.748 (0.021)			
SP07_07	SP52026	0.048	0.048	0.587 (0.059)	0.400 (0.129)	0.350 (0.034)		
SP07_08	SP52130	0.020	0.026	0.909 (0.075)	1.134 (0.045)	0.215 (0.015)		
SP07_09	SP52028	0.030	0.027	0.858 (0.063)	0.552 (0.058)	0.282 (0.021)		
SP07_10	SP52189	0.028	0.045	1.041 (0.030)	0.382 (0.018)			
SP07_11	SP52217	0.025	0.022	0.722 (0.070)	0.991 (0.068)	0.283 (0.022)		
SP07_12	SP52038	0.020	0.019	0.994 (0.079)	0.909 (0.045)	0.290 (0.016)		
SP07_13	SP52099	0.032	0.026	0.947 (0.031)	0.817 (0.023)			
SP07_14	SP52118	0.017	0.025	0.766 (0.030)	1.225 (0.038)			
SP09_01	SP62099	0.031	0.028	0.842 (0.047)	0.256 (0.050)	0.146 (0.021)		
SP09_02	SP62095	0.024	0.030	0.501 (0.015)	0.683 (0.024)		-0.076 (0.039)	0.076 (0.046)
SP09_03	SP62106	0.056	0.055	0.750 (0.037)	-0.721 (0.078)	0.116 (0.032)		
SP09_04	SP62064	0.048	0.033	0.879 (0.026)	-0.356 (0.023)			
SP09_05	SP62132	0.021	0.029	0.992 (0.063)	0.332 (0.052)	0.289 (0.020)		
SP09_06	SP62163	0.016	0.026	1.196 (0.043)	1.308 (0.028)			
SP09_07	SP62153	0.015	0.025	1.278 (0.089)	0.853 (0.035)	0.294 (0.013)		
SP09_08	SP62018	0.028	0.032	0.520 (0.015)	1.485 (0.038)		-0.653 (0.046)	0.653 (0.061)
SP09_09	SP62143	0.025	0.017	0.850 (0.037)	1.704 (0.052)			
SP09_10	SP62276	0.030	0.027	0.718 (0.027)	0.995 (0.034)			
SP09_11	SP62050	0.039	0.022	0.920 (0.031)	1.006 (0.027)			
SP09_12	SP62205	0.024	0.023	1.100 (0.066)	0.825 (0.032)	0.158 (0.013)		
SP09_13	SP62190	0.035	0.031	0.883 (0.045)	0.023 (0.051)	0.140 (0.022)		
SP09_14A	SP62024A	0.027	0.023	0.605 (0.059)	0.876 (0.085)	0.226 (0.028)		
SP09_14B	SP62024B	0.018	0.020	0.801 (0.032)	1.446 (0.044)			
SP11_01	SP62279	0.046	0.049	1.185 (0.055)	0.007 (0.037)	0.187 (0.017)		
SP11_02	SP62112	0.058	0.039	0.534 (0.020)	0.216 (0.032)			
SP11_03	SP62119	0.028	0.025	1.214 (0.063)	0.158 (0.038)	0.249 (0.017)		
SP11_04	SP62093	0.048	0.036	0.630 (0.017)	0.063 (0.018)		0.306 (0.033)	-0.306 (0.032)
SP11_05	SP62089	0.015	0.030	1.347 (0.078)	0.934 (0.026)	0.153 (0.010)		
SP11_06	SP62006	0.033	0.036	0.953 (0.028)	0.362 (0.020)			
SP11_07	SP62067	0.036	0.026	0.823 (0.026)	0.365 (0.022)			
SP11_08	SP62247	0.030	0.030	0.977 (0.090)	1.232 (0.048)	0.268 (0.014)		
SP11_09	SP62177	0.021	0.019	0.711 (0.062)	1.008 (0.060)	0.207 (0.020)		
SP11_10	SP62186	0.023	0.026	1.545 (0.119)	1.177 (0.030)	0.263 (0.010)		
SP11_11A	SP62211A	0.024	0.026	0.814 (0.026)	0.346 (0.022)			
SP11_11B	SP62211B	0.017	0.020	0.868 (0.045)	2.081 (0.075)			
SP11_13	SP62033	0.035	0.021	1.106 (0.034)	0.694 (0.019)			
SP11_14	SP62037	0.027	0.028	0.747 (0.062)	0.564 (0.074)	0.305 (0.024)		
SP11_15	SP62242	0.080	0.072	0.786 (0.026)	-1.200 (0.038)			

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SP13_01A	SP62091A	0.034	0.042	0.958 (0.053)	-0.706 (0.081)	0.304 (0.035)		
SP13_01B	SP62091B	0.054	0.051	0.587 (0.035)	-1.185 (0.152)	0.167 (0.054)		
SP13_02	SP62100	0.024	0.024	0.898 (0.027)	0.336 (0.021)			
SP13_03	SP62097	0.021	0.029	0.909 (0.049)	0.266 (0.046)	0.147 (0.020)		
SP13_04	SP62101	0.027	0.026	0.668 (0.018)	0.179 (0.017)		0.287 (0.031)	-0.287 (0.031)
SP13_06	SP62128	0.027	0.028	0.867 (0.026)	-0.024 (0.021)			
SP13_07	SP62047	0.048	0.058	0.497 (0.021)	0.592 (0.037)			
SP13_08	SP62042	0.027	0.024	0.710 (0.025)	0.639 (0.028)			
SP13_09	SP62250	0.051	0.033	0.580 (0.024)	1.200 (0.047)			
SP13_10	SP62246	0.020	0.022	0.924 (0.088)	1.189 (0.051)	0.288 (0.016)		
SP13_11	SP62056	0.022	0.027	1.147 (0.033)	0.428 (0.017)			
SP13_12	SP62235	0.027	0.023	0.765 (0.060)	0.854 (0.056)	0.195 (0.020)		
SP13_13	SP62180	0.019	0.023	1.210 (0.062)	0.259 (0.036)	0.211 (0.017)		
SP13_14	SP62022	0.025	0.029	0.562 (0.022)	0.621 (0.034)			
SP13_15	SP62243	0.027	0.031	0.664 (0.015)	-0.015 (0.017)		-0.331 (0.036)	0.331 (0.034)
Items Introduced in 2019:								
SP02_01	SP72072	—	0.031	0.824 (0.145)	0.518 (0.125)	0.216 (0.046)		
SP02_02	SP72029	—	0.030	1.324 (0.310)	1.057 (0.102)	0.364 (0.027)		
SP02_03	SP72902	—	0.048	1.017 (0.071)	0.145 (0.046)			
SP02_04	SP72077	—	0.032	0.685 (0.150)	0.395 (0.203)	0.300 (0.064)		
SP02_05A	SP72900A	—	0.037	0.959 (0.079)	0.884 (0.069)			
SP02_05B	SP72900B	—	0.022	0.954 (0.093)	1.360 (0.104)			
SP02_06	SP72103	—	0.048	0.500 (0.049)	-0.078 (0.084)			
SP02_07	SP72110	—	0.026	0.773 (0.069)	0.982 (0.089)			
SP02_08	SP72130	—	0.029	0.559 (0.057)	0.992 (0.118)			
SP02_09	SP72148	—	0.030	0.679 (0.153)	1.158 (0.145)	0.132 (0.042)		
SP02_10	SP72200	—	0.029	0.854 (0.129)	0.672 (0.092)	0.103 (0.034)		
SP02_11	SP72232	—	0.042	1.433 (0.096)	0.257 (0.036)			
SP02_12	SP72275	—	0.057	1.016 (0.108)	-0.521 (0.106)	0.117 (0.050)		
SP02_13	SP72244	—	0.030	0.950 (0.072)	0.497 (0.055)			
SP02_14	SP72301	—	0.020	0.936 (0.217)	1.199 (0.127)	0.220 (0.032)		
SP02_15	SP72721	—	0.033	1.028 (0.130)	0.253 (0.084)	0.137 (0.036)		
SP02_16	SP72335	—	0.029	0.859 (0.147)	0.552 (0.115)	0.199 (0.043)		
SP04_01	SP72002	—	0.031	1.393 (0.172)	0.239 (0.068)	0.212 (0.031)		
SP04_03	SP72021	—	0.035	0.896 (0.140)	0.336 (0.115)	0.221 (0.044)		
SP04_04	SP72082	—	0.057	0.960 (0.069)	0.291 (0.050)			
SP04_05	SP72066	—	0.032	0.837 (0.128)	0.591 (0.099)	0.123 (0.037)		
SP04_06	SP72063	—	0.026	0.582 (0.246)	1.996 (0.389)	0.200 (0.047)		
SP04_07	SP72102	—	0.043	0.482 (0.049)	0.544 (0.102)			
SP04_08A	SP72141A	—	0.024	1.069 (0.086)	0.876 (0.063)			
SP04_08B	SP72141B	—	0.032	0.731 (0.045)	0.601 (0.047)		-0.141 (0.075)	0.141 (0.090)

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SP04_09	SP72921	—	0.021	0.766 (0.077)	1.371 (0.122)			
SP04_10	SP72234	—	0.029	1.141 (0.277)	1.472 (0.139)	0.167 (0.022)		
SP04_11	SP72251	—	0.020	1.064 (0.189)	0.855 (0.090)	0.208 (0.030)		
SP04_12	SP72284	—	0.046	0.786 (0.058)	-0.058 (0.055)			
SP04_13	SP72345	—	0.048	0.860 (0.054)	0.372 (0.040)		0.552 (0.056)	-0.552 (0.068)
SP04_14	SP72349	—	0.037	1.086 (0.134)	0.083 (0.089)	0.178 (0.039)		
SP04_15	SP72363	—	0.049	0.613 (0.102)	0.073 (0.188)	0.101 (0.067)		
SP08_01	SP72070	—	0.057	0.568 (0.111)	-0.261 (0.322)	0.207 (0.099)		
SP08_02	SP72400	—	0.038	0.859 (0.061)	-0.012 (0.051)			
SP08_03	SP72024	—	0.046	0.891 (0.105)	-0.094 (0.107)	0.113 (0.045)		
SP08_04	SP72462	—	0.036	0.490 (0.138)	0.724 (0.286)	0.198 (0.086)		
SP08_05	SP72443	—	0.024	1.165 (0.254)	0.983 (0.103)	0.334 (0.030)		
SP08_06	SP72903	—	0.039	0.796 (0.048)	0.754 (0.046)		-0.090 (0.067)	0.090 (0.084)
SP08_07	SP72145	—	0.018	0.949 (0.091)	1.373 (0.102)			
SP08_08	SP72100	—	0.039	0.560 (0.132)	0.579 (0.227)	0.195 (0.073)		
SP08_10	SP72137	—	0.037	0.836 (0.132)	0.367 (0.122)	0.194 (0.046)		
SP08_11	SP72298	—	0.033	0.814 (0.064)	0.558 (0.063)			
SP08_12	SP72215	—	0.023	0.515 (0.033)	0.963 (0.072)		-0.538 (0.104)	0.538 (0.130)
SP08_13	SP72260	—	0.032	0.671 (0.056)	0.451 (0.071)			
SP08_14	SP72265	—	0.041	0.708 (0.057)	0.249 (0.063)			
SP08_15	SP72347	—	0.031	1.061 (0.208)	1.117 (0.099)	0.186 (0.028)		
SP08_16	SP72351	—	0.029	0.847 (0.072)	0.930 (0.077)			
SP08_17	SP72367	—	0.029	1.114 (0.159)	0.638 (0.076)	0.156 (0.030)		
SP10_01	SP72033	—	0.033	0.649 (0.035)	0.355 (0.044)		-0.436 (0.084)	0.436 (0.094)
SP10_02	SP72440	—	0.043	0.670 (0.053)	-0.347 (0.063)			
SP10_03	SP72032	—	0.029	1.540 (0.315)	1.001 (0.083)	0.315 (0.024)		
SP10_04	SP72031	—	0.025	0.655 (0.139)	0.941 (0.143)	0.137 (0.047)		
SP10_05	SP72086	—	0.038	0.556 (0.049)	-0.161 (0.073)			
SP10_06	SP72005	—	0.038	1.030 (0.065)	0.729 (0.040)		0.248 (0.050)	-0.248 (0.070)
SP10_08	SP72123	—	0.033	0.551 (0.125)	-0.003 (0.329)	0.249 (0.095)		
SP10_09	SP72116	—	0.026	0.574 (0.180)	1.172 (0.213)	0.198 (0.060)		
SP10_10	SP72920	—	0.060	0.599 (0.036)	0.920 (0.061)		1.334 (0.071)	-1.334 (0.128)
SP10_11	SP72294	—	0.033	0.914 (0.066)	0.207 (0.051)			
SP10_12	SP72231	—	0.029	1.257 (0.239)	0.923 (0.088)	0.265 (0.027)		
SP10_13	SP72261	—	0.043	0.671 (0.053)	-0.379 (0.064)			
SP10_14	SP72220	—	0.041	1.761 (0.627)	1.732 (0.166)	0.210 (0.017)		
SP10_15	SP72348	—	0.074	0.805 (0.059)	-0.844 (0.065)			
SP10_16	SP72720	—	0.030	0.412 (0.179)	1.745 (0.333)	0.135 (0.090)		
SP12_01	SP72078	—	0.029	1.019 (0.074)	0.458 (0.052)			
SP12_02	SP72460	—	0.022	0.962 (0.178)	0.710 (0.107)	0.254 (0.036)		
SP12_03	SP72000	—	0.030	0.717 (0.042)	0.318 (0.042)		-0.024 (0.070)	0.024 (0.080)

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SP12_05	SP72901	—	0.022	0.612 (0.185)	1.121 (0.202)	0.273 (0.057)		
SP12_06	SP72038	—	0.039	0.487 (0.108)	0.297 (0.277)	0.103 (0.089)		
SP12_07	SP72120	—	0.043	0.441 (0.104)	-0.046 (0.410)	0.092 (0.121)		
SP12_08	SP72143	—	0.036	0.731 (0.058)	0.314 (0.064)			
SP12_09	SP72523	—	0.049	0.663 (0.043)	0.319 (0.047)		0.309 (0.071)	-0.309 (0.084)
SP12_10	SP72168	—	0.032	1.195 (0.152)	0.319 (0.075)	0.176 (0.031)		
SP12_11	SP72205	—	0.025	1.159 (0.214)	0.881 (0.090)	0.244 (0.029)		
SP12_12	SP72293	—	0.029	0.959 (0.078)	0.858 (0.069)			
SP12_13A	SP72280A	—	0.027	1.309 (0.098)	0.755 (0.050)			
SP12_13B	SP72280B	—	0.029	1.433 (0.202)	-0.062 (0.095)	0.387 (0.039)		
SP12_14	SP72370	—	0.023	1.461 (0.214)	0.419 (0.073)	0.289 (0.030)		
SP14_01	SP72011	—	0.031	1.602 (0.170)	0.059 (0.057)	0.165 (0.029)		
SP14_02	SP72905	—	0.053	0.687 (0.053)	-0.340 (0.062)			
SP14_03	SP72049	—	0.030	0.805 (0.162)	0.616 (0.139)	0.270 (0.047)		
SP14_04	SP72016	—	0.027	0.782 (0.045)	0.560 (0.042)		-0.167 (0.069)	0.167 (0.082)
SP14_05	SP72451	—	0.047	1.084 (0.072)	-0.162 (0.043)			
SP14_06	SP72074	—	0.033	0.785 (0.061)	0.344 (0.060)			
SP14_07	SP72091	—	0.025	1.170 (0.198)	0.763 (0.084)	0.233 (0.030)		
SP14_08	SP72109	—	0.036	0.551 (0.054)	0.836 (0.108)			
SP14_09	SP72140	—	0.024	0.906 (0.206)	0.981 (0.125)	0.279 (0.037)		
SP14_10	SP72132	—	0.018	0.853 (0.096)	1.693 (0.151)			
SP14_11	SP72209	—	0.024	1.207 (0.200)	0.640 (0.085)	0.268 (0.032)		
SP14_12	SP72210	—	0.064	0.484 (0.038)	1.244 (0.088)		0.992 (0.087)	-0.992 (0.154)
SP14_13	SP72249	—	0.022	1.008 (0.170)	0.929 (0.089)	0.143 (0.028)		
SP14_14	SP72323	—	0.028	0.697 (0.169)	0.723 (0.179)	0.295 (0.056)		
SP14_15	SP72368	—	0.024	1.197 (0.191)	0.488 (0.090)	0.286 (0.034)		
SP14_16	SP72303	—	0.021	1.205 (0.255)	1.065 (0.102)	0.210 (0.026)		

Appendix 12E: Mathematics Item Parameters from the TIMSS 2019 Grade 4 Less Difficult Calibration

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
Items Released in 2015:								
N1_01	M011135	0.026	—	0.868 (0.091)	-0.272 (0.124)	0.211 (0.051)		
N1_02	M011114	0.014	—	1.386 (0.081)	0.655 (0.034)			
N1_03	M011216	0.030	—	1.473 (0.079)	0.130 (0.029)			
N1_04	M011255	0.018	—	1.079 (0.109)	0.119 (0.079)	0.191 (0.035)		
N1_05	M011027	0.022	—	1.127 (0.062)	-0.100 (0.035)			
N1_06	M011259	0.020	—	1.703 (0.172)	0.599 (0.043)	0.173 (0.020)		
N1_07	M011031	0.026	—	0.873 (0.081)	-0.304 (0.103)	0.144 (0.044)		
N1_08	M011227	0.026	—	0.613 (0.050)	1.246 (0.098)			
N1_09	M011267	0.042	—	0.711 (0.052)	-1.472 (0.094)			
N1_10	M011042	0.022	—	0.693 (0.046)	-0.114 (0.051)			
N1_11	M011184	0.014	—	0.870 (0.054)	0.311 (0.044)			
N1_12	M011190	0.021	—	1.236 (0.075)	0.717 (0.039)			
N1_13	M011193	0.019	—	1.728 (0.257)	1.068 (0.057)	0.280 (0.019)		
N4_01	M061272	0.018	—	1.024 (0.076)	1.283 (0.068)			
N4_02	M061243	0.019	—	0.663 (0.030)	0.739 (0.039)		-0.566 (0.073)	0.566 (0.084)
N4_03	M061029	0.025	—	1.430 (0.157)	0.665 (0.051)	0.183 (0.022)		
N4_04	M061031	0.017	—	1.411 (0.177)	1.361 (0.065)	0.070 (0.012)		
N4_05	M061050	0.019	—	1.442 (0.244)	1.448 (0.081)	0.186 (0.017)		
N4_06	M061167	0.027	—	0.975 (0.057)	-0.030 (0.039)			
N4_07	M061206	0.018	—	1.327 (0.247)	1.625 (0.103)	0.164 (0.016)		
N4_08A	M061265A	0.034	—	0.869 (0.090)	2.083 (0.155)			
N4_08B	M061265B	0.022	—	1.255 (0.324)	2.019 (0.185)	0.175 (0.015)		
N4_09	M061185	0.030	—	1.392 (0.146)	0.552 (0.053)	0.175 (0.023)		
N4_10	M061239	0.017	—	1.422 (0.083)	0.558 (0.032)			
Items Common in 2015 and 2019:								
MN01_01	MN11128	0.032	0.077	0.998 (0.047)	0.388 (0.032)			
MN01_02	MN11022	0.027	0.047	1.278 (0.056)	-0.440 (0.029)			
MN01_03	MN11010	0.027	0.035	1.239 (0.056)	0.447 (0.027)			
MN01_04A	MN11278A	0.039	0.065	1.161 (0.077)	-0.603 (0.066)	0.134 (0.034)		
MN01_04B	MN11278B	0.034	0.039	1.576 (0.141)	0.928 (0.037)	0.146 (0.014)		
MN01_05	MN11136	0.029	0.035	0.951 (0.044)	-0.079 (0.032)			
MN01_06	MN11261	0.014	0.047	1.099 (0.055)	0.875 (0.036)			
MN01_07	MN11033	0.042	0.062	0.427 (0.030)	-0.128 (0.063)			
MN01_08	MN11039	0.040	0.060	0.727 (0.057)	-0.824 (0.132)	0.151 (0.053)		

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
MN01_09	MN11040	0.036	0.040	0.425 (0.055)	0.259 (0.226)	0.162 (0.059)		
MN01_10	MN11195	0.037	0.053	0.583 (0.042)	1.607 (0.102)			
MN01_11	MN11188	0.021	0.036	0.603 (0.037)	0.831 (0.059)			
MN01_12	MN11252	0.036	0.060	1.925 (0.169)	0.797 (0.033)	0.185 (0.014)		
MN03_01	MN11055	0.031	0.057	1.031 (0.049)	-0.730 (0.038)			
MN03_02	MN11214	0.024	0.044	1.308 (0.107)	0.214 (0.054)	0.229 (0.025)		
MN03_03A	MN11116A	0.030	0.050	1.102 (0.052)	-0.779 (0.037)			
MN03_03B	MN11116B	0.025	0.047	1.100 (0.052)	0.533 (0.031)			
MN03_04A	MN11066A	0.032	0.042	1.176 (0.057)	0.721 (0.032)			
MN03_04B	MN11066B	0.043	0.051	1.236 (0.065)	1.030 (0.038)			
MN03_05	MN11260	0.025	0.027	1.563 (0.113)	0.129 (0.042)	0.183 (0.022)		
MN03_06	MN11032	0.052	0.070	0.814 (0.061)	-0.270 (0.085)	0.112 (0.035)		
MN03_07	MN11170	0.059	0.095	0.462 (0.066)	0.638 (0.191)	0.169 (0.053)		
MN03_08	MN11068	0.046	0.066	0.599 (0.035)	0.022 (0.046)			
MN03_09	MN11269	0.024	0.038	1.045 (0.048)	-0.403 (0.033)			
MN03_10	MN11001	0.046	0.101	1.046 (0.101)	0.629 (0.060)	0.185 (0.024)		
MN03_11	MN11235	0.022	0.039	0.549 (0.024)	1.432 (0.056)		-0.886 (0.080)	0.886 (0.102)
MN05_01	MN11076	0.024	0.038	0.878 (0.072)	-0.454 (0.106)	0.201 (0.044)		
MN05_02	MN11141	0.016	0.031	1.124 (0.050)	-0.107 (0.029)			
MN05_03	MN11142	0.023	0.038	1.888 (0.136)	0.505 (0.029)	0.128 (0.014)		
MN05_04	MN11005	0.031	0.057	2.191 (0.189)	0.640 (0.031)	0.228 (0.015)		
MN05_05A	MN11256A	0.060	0.065	0.989 (0.045)	-0.507 (0.036)			
MN05_05B	MN11256B	0.044	0.054	0.987 (0.045)	-0.030 (0.031)			
MN05_06	MN11108	0.034	0.046	1.000 (0.055)	1.049 (0.045)			
MN05_07	MN11062	0.046	0.067	0.397 (0.031)	0.685 (0.081)			
MN05_08	MN11174	0.037	0.069	0.814 (0.042)	0.529 (0.039)			
MN05_09	MN11067	0.041	0.052	0.488 (0.064)	-0.054 (0.241)	0.223 (0.065)		
MN05_10	MN11043	0.052	0.099	0.687 (0.045)	-2.155 (0.113)			
MN05_11	MN11268	0.027	0.048	0.782 (0.067)	0.264 (0.076)	0.107 (0.029)		
MN05_12	MN11270	0.027	0.026	1.214 (0.058)	0.622 (0.030)			
MN07_01	MN11023	0.029	0.045	1.527 (0.116)	0.169 (0.045)	0.216 (0.022)		
MN07_02	MN11056	0.029	0.063	1.164 (0.094)	0.258 (0.056)	0.183 (0.025)		
MN07_03	MN11057	0.023	0.060	1.235 (0.054)	-0.274 (0.028)			
MN07_04	MN11113	0.021	0.059	1.045 (0.047)	-0.151 (0.031)			
MN07_05	MN11200	0.045	0.073	0.475 (0.016)	-1.102 (0.048)		-1.648 (0.116)	1.648 (0.102)
MN07_06	MN11129	0.023	0.047	1.313 (0.108)	0.466 (0.046)	0.180 (0.021)		
MN07_07	MN11218	0.024	0.070	0.854 (0.042)	-0.770 (0.045)			
MN07_08	MN11036	0.020	0.037	1.373 (0.136)	1.041 (0.045)	0.156 (0.015)		
MN07_09	MN11225	0.054	0.074	0.644 (0.040)	1.090 (0.066)			
MN07_10	MN11041	0.020	0.058	0.913 (0.095)	0.269 (0.094)	0.281 (0.035)		
MN07_11	MN11179	0.028	0.082	0.946 (0.048)	0.710 (0.038)			

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
MN07_12	MN11303	0.030	0.076	1.113 (0.062)	1.156 (0.044)			
MN07_13	MN11305	0.020	0.041	1.136 (0.161)	1.256 (0.071)	0.279 (0.020)		
MN09_01	MN11019	0.027	0.048	0.953 (0.088)	0.237 (0.078)	0.219 (0.031)		
MN09_02	MN11145	0.035	0.069	1.060 (0.048)	-0.379 (0.032)			
MN09_03	MN11211	0.034	0.023	1.858 (0.135)	0.286 (0.034)	0.179 (0.018)		
MN09_04	MN11014	0.020	0.074	1.196 (0.054)	0.395 (0.028)			
MN09_05	MN11300	0.018	0.027	1.148 (0.055)	0.620 (0.031)			
MN09_06	MN11028	0.031	0.041	1.383 (0.060)	0.040 (0.025)			
MN09_07	MN11231	0.029	0.040	1.327 (0.201)	1.610 (0.079)	0.190 (0.014)		
MN09_08	MN11061	0.064	0.096	0.660 (0.051)	-1.040 (0.146)	0.138 (0.054)		
MN09_09	MN11045	0.029	0.072	0.999 (0.081)	0.075 (0.069)	0.165 (0.030)		
MN09_10	MN11265	0.038	0.091	0.851 (0.069)	-1.083 (0.141)	0.215 (0.061)		
MN09_11	MN11154	0.034	0.042	0.685 (0.025)	0.425 (0.028)		-0.377 (0.055)	0.377 (0.059)
MN09_12	MN11240	0.020	0.043	1.100 (0.148)	1.214 (0.069)	0.252 (0.020)		
MN11_01	MN11009	0.034	0.039	0.978 (0.084)	0.073 (0.078)	0.210 (0.033)		
MN11_02	MN11024	0.028	0.063	1.072 (0.048)	0.172 (0.029)			
MN11_03	MN11134	0.025	0.051	1.272 (0.116)	0.430 (0.056)	0.262 (0.024)		
MN11_04	MN11212	0.034	0.060	0.873 (0.042)	-0.178 (0.035)			
MN11_05	MN11253	0.025	0.044	0.960 (0.080)	0.079 (0.075)	0.179 (0.032)		
MN11_06	MN11221	0.035	0.049	2.127 (0.177)	0.767 (0.029)	0.161 (0.013)		
MN11_07	MN11146	0.052	0.078	0.760 (0.042)	0.820 (0.048)			
MN11_08	MN11177	0.019	0.032	1.337 (0.067)	0.910 (0.032)			
MN11_09	MN11158	0.045	0.081	0.675 (0.037)	0.273 (0.044)			
MN11_10	MN11002	0.029	0.054	1.288 (0.143)	1.083 (0.052)	0.204 (0.017)		
MN11_11A	MN11182A	0.037	0.078	0.987 (0.072)	-1.076 (0.105)	0.172 (0.051)		
MN11_11B	MN11182B	0.045	0.070	0.859 (0.064)	-0.548 (0.096)	0.142 (0.042)		
MN11_12	MN11272	0.019	0.028	0.766 (0.043)	1.859 (0.071)		-0.185 (0.060)	0.185 (0.104)
MN13_01	MN11017	0.042	0.060	0.786 (0.040)	-0.936 (0.052)			
MN13_02	MN11125	0.042	0.080	0.894 (0.044)	0.322 (0.034)			
MN13_03	MN11077	0.025	0.020	1.141 (0.058)	0.844 (0.035)			
MN13_04A	MN11047A	0.053	0.075	1.045 (0.072)	-0.753 (0.083)	0.147 (0.040)		
MN13_04B	MN11047B	0.051	0.063	1.100 (0.083)	-0.407 (0.078)	0.198 (0.037)		
MN13_05	MN11223	0.022	0.021	1.154 (0.056)	0.703 (0.032)			
MN13_06	MN11034	0.047	0.070	0.915 (0.113)	1.123 (0.070)	0.173 (0.022)		
MN13_07	MN11175	0.039	0.030	1.049 (0.050)	0.473 (0.031)			
MN13_08	MN11262	0.022	0.035	0.975 (0.105)	0.825 (0.063)	0.191 (0.024)		
MN13_09	MN11239	0.023	0.027	0.800 (0.123)	1.408 (0.097)	0.202 (0.024)		
MN13_10	MN11202	0.025	0.026	0.910 (0.043)	-0.152 (0.034)			
MN13_11	MN11299	0.027	0.083	1.349 (0.061)	0.321 (0.025)			
MP03_01	MP61026	0.025	0.034	1.112 (0.093)	0.459 (0.053)	0.158 (0.023)		
MP03_02	MP61273	0.041	0.043	1.083 (0.110)	0.953 (0.054)	0.156 (0.019)		

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
MP03_03	MP61034	0.016	0.040	1.221 (0.077)	1.498 (0.056)			
MP03_04	MP61040	0.029	0.037	1.722 (0.226)	1.443 (0.054)	0.180 (0.012)		
MP03_05	MP61228	0.026	0.032	0.808 (0.046)	1.885 (0.072)		-0.178 (0.059)	0.178 (0.105)
MP03_06	MP61166	0.021	0.030	1.194 (0.058)	0.709 (0.031)			
MP03_07	MP61171	0.020	0.041	1.404 (0.129)	0.605 (0.047)	0.236 (0.020)		
MP03_08	MP61080	0.021	0.030	0.677 (0.048)	1.668 (0.097)			
MP03_09	MP61222	0.026	0.038	0.880 (0.115)	0.952 (0.081)	0.253 (0.027)		
MP03_10	MP61076	0.035	0.030	0.522 (0.035)	0.790 (0.066)			
MP03_11	MP61084	0.035	0.027	0.961 (0.068)	1.725 (0.083)			
Items Introduced in 2019:								
MN04_01	MN21061	—	0.027	1.821 (0.197)	-0.071 (0.061)	0.142 (0.034)		
MN04_02	MN21067	—	0.023	1.223 (0.181)	0.332 (0.099)	0.254 (0.043)		
MN04_03	MN21046	—	0.024	1.023 (0.084)	0.635 (0.054)			
MN04_04	MN21023	—	0.034	1.193 (0.088)	-0.450 (0.053)			
MN04_05	MN21018	—	0.046	0.894 (0.081)	0.935 (0.071)			
MN04_06	MN21020	—	0.035	1.318 (0.153)	0.384 (0.065)	0.096 (0.030)		
MN04_07	MN21069	—	0.031	1.658 (0.200)	-0.065 (0.076)	0.214 (0.041)		
MN04_08	MN21040	—	0.031	1.990 (0.249)	0.399 (0.054)	0.193 (0.029)		
MN04_09	MN21070	—	0.027	1.154 (0.217)	1.135 (0.091)	0.179 (0.030)		
MN04_10	MN21037	—	0.027	1.675 (0.198)	0.389 (0.057)	0.137 (0.028)		
MN04_11	MN21033	—	0.035	0.719 (0.074)	1.183 (0.102)			
MN04_12	MN21001	—	0.028	1.039 (0.185)	0.809 (0.101)	0.210 (0.039)		
MN04_13	MN21060	—	0.034	0.487 (0.027)	-0.219 (0.059)		-1.214 (0.150)	1.214 (0.142)
MN04_14	MN21003	—	0.062	0.518 (0.041)	-0.174 (0.063)		0.244 (0.116)	-0.244 (0.104)
MN12_01	MN21066	—	0.037	1.003 (0.128)	-0.385 (0.143)	0.157 (0.067)		
MN12_02	MN21045	—	0.036	0.606 (0.132)	0.530 (0.231)	0.151 (0.080)		
MN12_03	MN21064	—	0.024	1.686 (0.207)	0.663 (0.052)	0.114 (0.023)		
MN12_04	MN21051	—	0.026	1.568 (0.112)	0.224 (0.038)			
MN12_05	MN21054	—	0.043	1.363 (0.180)	-0.662 (0.132)	0.279 (0.069)		
MN12_06	MN21025	—	0.027	0.850 (0.045)	0.179 (0.037)		-0.499 (0.084)	0.499 (0.084)
MN12_07	MN21038	—	0.025	1.156 (0.179)	0.359 (0.109)	0.260 (0.045)		
MN12_08	MN21043	—	0.026	1.226 (0.097)	0.699 (0.049)			
MN12_09	MN21030	—	0.028	0.933 (0.156)	0.665 (0.110)	0.163 (0.044)		
MN12_10	MN21032	—	0.034	0.665 (0.059)	-0.335 (0.077)			
MN12_11	MN21053	—	0.029	1.107 (0.155)	-0.160 (0.136)	0.248 (0.061)		
MN12_12A	MN21010A	—	0.031	0.808 (0.150)	-0.086 (0.236)	0.302 (0.083)		
MN12_12B	MN21010B	—	0.040	0.893 (0.120)	2.081 (0.188)			
MN12_13	MN21059	—	0.027	1.166 (0.094)	0.678 (0.051)			
MN14_01	MN21049	—	0.041	0.442 (0.038)	0.355 (0.070)		0.666 (0.116)	-0.666 (0.125)
MN14_02	MN21050	—	0.047	0.512 (0.053)	0.210 (0.089)			
MN14_03	MN21065	—	0.031	1.794 (0.295)	0.817 (0.066)	0.283 (0.028)		

Item		RMSD 2015	RMSD 2019	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
MN14_04	MN21014	—	0.055	0.910 (0.073)	-0.823 (0.075)			
MN14_05	MN21019	—	0.031	1.210 (0.168)	-0.037 (0.118)	0.266 (0.053)		
MN14_06	MN21024	—	0.025	1.498 (0.223)	1.023 (0.062)	0.121 (0.023)		
MN14_07	MN21035	—	0.020	1.261 (0.190)	0.368 (0.099)	0.270 (0.043)		
MN14_08	MN21039	—	0.028	1.190 (0.162)	0.069 (0.110)	0.231 (0.050)		
MN14_09	MN21062	—	0.046	0.517 (0.053)	-0.136 (0.091)			
MN14_10	MN21057	—	0.025	0.650 (0.064)	0.828 (0.088)			
MN14_11	MN21063	—	0.030	1.225 (0.155)	-0.017 (0.103)	0.193 (0.049)		
MN14_12	MN21005	—	0.026	0.919 (0.077)	0.581 (0.058)			
MN14_13A	MN21012A	—	0.053	0.767 (0.064)	-0.319 (0.071)			
MN14_13B	MN21012B	—	0.032	1.605 (0.252)	0.848 (0.066)	0.213 (0.028)		
MP02_01	MP71219	—	0.035	0.995 (0.140)	0.107 (0.122)	0.174 (0.053)		
MP02_02	MP71021	—	0.033	1.692 (0.240)	0.892 (0.057)	0.140 (0.023)		
MP02_03	MP71167	—	0.042	1.084 (0.119)	1.638 (0.111)			
MP02_04	MP71041	—	0.033	0.973 (0.157)	0.810 (0.095)	0.129 (0.037)		
MP02_05	MP71162	—	0.024	0.540 (0.045)	1.972 (0.140)		-0.968 (0.153)	0.968 (0.215)
MP02_06	MP71078	—	0.027	0.788 (0.070)	0.573 (0.068)			
MP02_07	MP71090	—	0.029	0.788 (0.182)	1.355 (0.140)	0.139 (0.040)		
MP02_08	MP71151	—	0.027	0.590 (0.044)	1.819 (0.107)		-1.918 (0.219)	1.918 (0.253)
MP02_09	MP71119	—	0.041	0.649 (0.059)	-0.203 (0.077)			
MP02_10A	MP71217A	—	0.030	0.711 (0.065)	0.411 (0.071)			
MP02_11	MP71142	—	0.031	1.207 (0.094)	0.440 (0.047)			
MP02_12	MP71204	—	0.027	1.112 (0.115)	1.418 (0.091)			
MP08_01	MP71018	—	0.024	1.343 (0.235)	1.140 (0.078)	0.175 (0.026)		
MP08_02	MP71009	—	0.039	1.361 (0.116)	1.019 (0.053)			
MP08_03	MP71037	—	0.052	0.761 (0.071)	0.961 (0.082)			
MP08_04	MP71051	—	0.032	1.049 (0.132)	1.973 (0.150)			
MP08_05	MP71064	—	0.027	1.016 (0.257)	1.590 (0.142)	0.195 (0.030)		
MP08_06	MP71169	—	0.030	1.379 (0.135)	1.391 (0.071)			
MP08_07	MP71083	—	0.033	1.708 (0.432)	1.590 (0.103)	0.237 (0.022)		
MP08_09	MP71184	—	0.027	2.331 (0.847)	1.883 (0.120)	0.244 (0.018)		
MP08_10	MP71141	—	0.050	0.585 (0.089)	2.430 (0.293)			
MP08_11	MP71194	—	0.044	0.750 (0.064)	-0.224 (0.071)			
MP08_12	MP71193	—	0.033	0.596 (0.043)	1.289 (0.077)		-0.565 (0.114)	0.565 (0.143)
MP08_13	MP71192	—	0.025	0.498 (0.037)	1.885 (0.120)		-2.221 (0.247)	2.221 (0.283)
MP13_01	MP61240	—	0.045	0.572 (0.081)	2.083 (0.241)			
MP13_02	MP61254	—	0.062	0.678 (0.069)	1.111 (0.102)			
MP13_03	MP61244	—	0.025	1.382 (0.236)	0.802 (0.083)	0.257 (0.033)		
MP13_04	MP61041	—	0.029	0.729 (0.267)	1.997 (0.286)	0.195 (0.040)		
MP13_05	MP61173	—	0.038	0.774 (0.067)	0.458 (0.066)			
MP13_06	MP61252	—	0.027	1.748 (0.395)	1.650 (0.097)	0.127 (0.017)		

Item		RMSD		Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
		2015	2019					
MP13_07	MP61261	—	0.032	1.562 (0.137)	1.084 (0.051)			
MP13_08	MP61224	—	0.023	0.821 (0.083)	1.321 (0.101)			
MP13_09	MP61077	—	0.027	1.151 (0.232)	1.438 (0.105)	0.126 (0.025)		
MP13_10A	MP61069A	—	0.035	0.716 (0.063)	0.290 (0.069)			
MP13_10B	MP61069B	—	0.033	0.726 (0.070)	0.874 (0.084)			

Appendix 12F: Science Item Parameters from the TIMSS 2019 Grade 4 Less Difficult Calibration

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
Trend Items*:							
SP01_01	SP51054	0.114	0.934	-0.419	0.261		
SP01_02	SP51024	0.058	0.612	0.674			
SP01_03A	SP51132A	0.033	0.881	1.254			
SP01_03B	SP51132B	0.059	0.810	1.065			
SP01_04	SP51040	0.126	0.453	0.606			
SP01_05	SP51193	0.064	0.940	-0.126	0.274		
SP01_06	SP51063	0.034	1.148	0.754	0.222		
SP01_07	SP51012	0.049	0.989	0.268	0.253		
SP01_08	SP51115	0.033	1.090	0.146			
SP01_09	SP51180	0.062	0.880	0.057	0.360		
SP01_10	SP51106	0.062	1.024	0.721	0.215		
SP01_11	SP51148	0.067	1.049	0.043	0.241		
SP03_01	SP61141	0.097	1.235	0.519	0.300		
SP03_02	SP61023	0.039	0.770	0.015			
SP03_03	SP61054	0.075	0.479	0.643		1.489	-1.489
SP03_04	SP61007	0.082	0.647	-0.209	0.163		
SP03_05	SP61006	0.118	0.785	-0.650			
SP03_06	SP61108	0.058	1.050	0.233	0.352		
SP03_07	SP61109	0.064	0.583	0.710	0.235		
SP03_08	SP61080	0.056	0.968	0.297	0.264		
SP03_09	SP61088	0.051	0.672	1.417			
SP03_10	SP61151	0.069	0.952	0.440			
SP03_11	SP61150	0.090	0.624	0.408			
SP03_12	SP61169	0.037	1.077	0.079	0.268		
SP05_01	SP51044	0.119	0.503	0.201			
SP05_03	SP51003	0.061	0.711	-0.122	0.104		
SP05_04	SP51168	0.179	0.704	-0.475			
SP05_05	SP51010	0.060	0.766	0.076			
SP05_06	SP51035	0.035	1.249	1.298	0.236		
SP05_07	SP51059	0.064	0.584	0.104			
SP05_08	SP51142	0.048	0.802	0.598	0.199		
SP05_09A	SP51131A	0.054	1.014	-0.089	0.193		
SP05_09B	SP51131B	0.029	0.988	0.576	0.197		
SP05_10	SP51151	0.122	0.918	-1.120			
SP05_11	SP51157	0.065	0.739	0.999	0.190		

* Item parameters for trend items were fixed from the paperTIMSS fourth grade science concurrent calibration.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
Trend Items*:							
SP06_01	SP61071	0.124	0.335	-1.372	0.197		
SP06_02	SP61138	0.119	0.616	0.002			
SP06_03A	SP61016A	0.049	0.926	0.365	0.216		
SP06_03B	SP61016B	0.026	0.990	0.509			
SP06_04	SP61011	0.103	0.733	-0.536			
SP06_06	SP61083	0.100	0.726	-1.025			
SP06_07	SP61034	0.038	0.788	1.088			
SP06_08	SP61044	0.052	0.740	0.551			
SP06_09A	SP61142A	0.056	0.623	0.351			
SP06_09B	SP61142B	0.052	0.788	1.034			
SP06_10A	SP61115A	0.062	1.468	0.346	0.264		
SP06_10B	SP61115B	0.072	1.345	0.662	0.328		
SP07_01	SP51161	0.109	0.488	1.007	0.217		
SP07_02	SP51051	0.121	1.391	1.370	0.281		
SP07_03Z	SP51138Z	0.055	0.583	0.313			
SP07_04	SP51194	0.017	0.970	1.014			
SP07_05	SP51029	0.040	0.518	1.220	0.202		
SP07_06	SP51077	0.079	0.747	-0.167			
SP07_07	SP51200	0.129	0.679	1.196			
SP07_08	SP51075	0.135	0.670	-0.586			
SP07_09	SP51065	0.084	0.870	-0.215	0.333		
SP07_10	SP51191	0.051	1.342	0.578	0.205		
SP07_11	SP51099	0.044	0.868	0.332	0.216		
SP07_12	SP51175	0.024	0.978	0.968			
SP09_01	SP61135	0.102	0.758	-0.598	0.268		
SP09_02	SP61069	0.120	0.400	-0.481			
SP09_03	SP61134	0.086	0.651	0.181	0.126		
SP09_04	SP61140	0.040	1.039	0.601	0.296		
SP09_05	SP61019	0.031	0.887	0.943			
SP09_06	SP61022	0.085	0.656	0.183	0.241		
SP09_07	SP61036	0.095	0.951	0.903			
SP09_08	SP61160	0.108	0.761	-0.954			
SP09_09	SP61159	0.114	0.826	-0.788			
SP09_10	SP61091	0.036	0.452	1.170		-0.176	0.176
SP09_11	SP61118	0.034	1.001	0.542	0.217		
SP09_12	SP61097	0.036	0.798	0.517	0.275		
SP11_01	SP61132	0.090	0.710	0.539	0.213		
SP11_02	SP61120	0.048	0.884	0.333	0.197		
SP11_03	SP61025	0.079	0.531	-0.366			

* Item parameters for trend items were fixed from the paperTIMSS fourth grade science concurrent calibration.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
Trend Items*:							
SP11_04A	SP61133A	0.061	1.370	0.245	0.326		
SP11_04B	SP61133B	0.046	1.701	0.792	0.114		
SP11_05	SP61074	0.075	0.772	0.219			
SP11_06	SP61093	0.142	0.761	-0.057		0.937	-0.937
SP11_07	SP61161	0.086	0.614	0.664			
SP11_08A	SP61042A	0.031	1.366	0.806	0.239		
SP11_08B	SP61042B	0.040	0.791	0.640	0.150		
SP11_09A	SP61041A	0.055	0.871	0.116			
SP11_09B	SP61041B	0.066	0.719	0.167			
SP11_10	SP61155	0.097	0.735	-0.488	0.286		
SP13_02	SP61014	0.085	0.495	0.425			
SP13_03	SP61056	0.104	0.853	-0.738			
SP13_04	SP61015	0.110	0.692	-0.395			
SP13_05	SP61113	0.101	0.760	0.954			
SP13_06	SP61107	0.075	1.001	0.641	0.180		
SP13_07	SP61046	0.029	1.164	0.804	0.227		
SP13_08	SP61047	0.095	0.751	-0.518	0.313		
SP13_09	SP61048	0.045	1.300	0.509	0.221		
SP13_10	SP61096	0.054	1.100	0.730	0.257		
SP13_11	SP61124	0.031	0.590	1.242			
SP13_12	SP61116	0.092	0.681	0.159			
New Items:							
SP02_01	SP71002	0.103	0.516 (0.047)	-0.414 (0.099)			
SP02_02	SP71402	0.049	1.200 (0.183)	-0.140 (0.086)	0.240 (0.027)		
SP02_03	SP71017	0.048	0.598 (0.065)	0.571 (0.142)			
SP02_04	SP71077	0.038	0.970 (0.089)	0.243 (0.076)			
SP02_05	SP71072	0.030	0.959 (0.240)	0.673 (0.150)	0.246 (0.027)		
SP02_06	SP71054	0.053	0.954 (0.082)	-0.045 (0.067)			
SP02_07	SP71115	0.046	0.703 (0.197)	0.728 (0.202)	0.249 (0.034)		
SP02_08	SP71140	0.046	0.850 (0.180)	0.366 (0.131)	0.231 (0.030)		
SP02_09	SP71128	0.084	0.478 (0.122)	-0.242 (0.285)	0.281 (0.063)		
SP02_10	SP71147	0.086	0.639 (0.125)	-0.285 (0.171)	0.235 (0.045)		
SP02_11A	SP71920A	0.049	0.862 (0.078)	0.113 (0.079)			
SP02_11B	SP71920B	0.029	0.955 (0.096)	0.458 (0.091)			
SP02_12	SP71268	0.024	1.089 (0.425)	1.318 (0.292)	0.213 (0.022)		
SP04_01	SP71013	0.086	0.959 (0.140)	-0.850 (0.124)	0.272 (0.038)		
SP04_02	SP71902	0.058	0.343 (0.043)	0.537 (0.215)			
SP04_03	SP71076	0.050	1.053 (0.155)	-0.254 (0.092)	0.210 (0.029)		

* Item parameters for trend items were fixed from the paperTIMSS fourth grade science concurrent calibration.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
New Items:							
SP04_04	SP71041	0.030	0.659 (0.063)	1.056 (0.125)		-0.010 (0.105)	0.010 (0.183)
SP04_05	SP71046	0.039	0.645 (0.069)	0.627 (0.137)			
SP04_06	SP71095	0.048	0.558 (0.060)	0.567 (0.147)			
SP04_07	SP71129	0.091	0.818 (0.134)	-0.904 (0.165)	0.308 (0.045)		
SP04_08	SP71102	0.031	0.811 (0.088)	0.656 (0.122)			
SP04_09	SP71124	0.031	0.900 (0.232)	0.762 (0.167)	0.223 (0.027)		
SP04_10	SP71112	0.138	0.686 (0.099)	-1.453 (0.211)	0.218 (0.058)		
SP04_11	SP71265	0.053	0.209 (0.099)	1.461 (0.883)	0.273 (0.097)		
SP04_12	SP71223	0.118	0.482 (0.078)	-1.707 (0.350)	0.238 (0.080)		
SP08_02	SP71033	0.059	0.320 (0.131)	0.812 (0.461)	0.275 (0.080)		
SP08_03	SP71065	0.092	0.428 (0.042)	-0.391 (0.119)			
SP08_04	SP71025	0.081	0.353 (0.094)	-0.043 (0.366)	0.215 (0.072)		
SP08_05	SP71081	0.023	0.937 (0.323)	1.328 (0.300)	0.167 (0.022)		
SP08_06	SP71056	0.025	0.610 (0.082)	1.382 (0.243)			
SP08_07	SP71145	0.075	0.493 (0.114)	0.010 (0.214)	0.198 (0.051)		
SP08_08	SP71104	0.115	0.741 (0.057)	-0.780 (0.071)			
SP08_09	SP71144	0.065	0.492 (0.131)	0.547 (0.222)	0.180 (0.045)		
SP08_10	SP71150	0.065	0.742 (0.065)	-0.104 (0.081)			
SP08_11	SP71201	0.042	1.069 (0.180)	-0.046 (0.098)	0.248 (0.029)		
SP08_12	SP71237	0.044	1.097 (0.097)	0.096 (0.065)			
SP08_13	SP71260	0.031	0.664 (0.237)	1.404 (0.381)	0.170 (0.028)		
SP10_01	SP71009	0.126	0.521 (0.036)	-0.386 (0.073)		1.209 (0.101)	-1.209 (0.126)
SP10_02	SP71093	0.081	0.731 (0.059)	-0.493 (0.073)			
SP10_03	SP71069	0.038	0.842 (0.281)	0.955 (0.240)	0.288 (0.029)		
SP10_04	SP71051	0.036	0.687 (0.073)	0.630 (0.134)			
SP10_05	SP71039	0.045	0.985 (0.161)	0.083 (0.096)	0.173 (0.026)		
SP10_06	SP71080	0.028	0.819 (0.473)	1.799 (0.682)	0.236 (0.026)		
SP10_07	SP71137	0.066	0.720 (0.062)	-0.213 (0.080)			
SP10_08	SP71103	0.043	0.833 (0.189)	0.323 (0.142)	0.271 (0.032)		
SP10_09	SP71106	0.057	0.422 (0.054)	1.021 (0.242)			
SP10_10	SP71100	0.067	0.697 (0.146)	-0.203 (0.167)	0.288 (0.042)		
SP10_12	SP71220	0.034	0.783 (0.221)	0.832 (0.206)	0.222 (0.030)		
SP10_13	SP71254	0.041	0.546 (0.067)	0.967 (0.200)			
SP12_01	SP71031	0.066	0.320 (0.044)	1.046 (0.295)			
SP12_02	SP71090	0.061	0.726 (0.065)	0.024 (0.088)			
SP12_03	SP71048	0.027	2.262 (0.784)	1.004 (0.127)	0.224 (0.017)		
SP12_04	SP71071	0.025	0.732 (0.093)	1.110 (0.186)			
SP12_05	SP71011	0.070	1.081 (0.154)	-0.350 (0.089)	0.203 (0.028)		
SP12_06	SP71142	0.095	0.387 (0.101)	-0.138 (0.331)	0.221 (0.069)		

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
New Items:							
SP12_07	SP71138	0.113	0.702 (0.055)	-0.752 (0.074)			
SP12_08	SP71127	0.068	0.711 (0.153)	-0.064 (0.157)	0.285 (0.039)		
SP12_10	SP71500	0.049	0.816 (0.153)	0.232 (0.120)	0.170 (0.028)		
SP12_11	SP71257	0.034	1.649 (0.762)	1.030 (0.207)	0.429 (0.022)		
SP12_12	SP71222	0.044	0.985 (0.092)	0.194 (0.076)			
SP12_13	SP71252	0.043	0.968 (0.200)	0.200 (0.119)	0.266 (0.030)		
SP14_01	SP71063	0.075	0.409 (0.045)	0.160 (0.153)			
SP14_02	SP71900	0.043	0.965 (0.206)	0.083 (0.127)	0.351 (0.032)		
SP14_04	SP71043	0.026	0.742 (0.096)	1.179 (0.193)			
SP14_05	SP71005	0.101	0.845 (0.065)	-0.691 (0.063)			
SP14_06	SP71118	0.033	1.178 (0.233)	0.473 (0.101)	0.185 (0.023)		
SP14_07	SP71139	0.106	0.538 (0.106)	-0.701 (0.243)	0.236 (0.059)		
SP14_08	SP71114	0.086	0.713 (0.059)	-0.481 (0.074)			
SP14_09	SP71131	0.077	0.566 (0.053)	-0.097 (0.102)			
SP14_10	SP71152	0.040	0.862 (0.197)	0.361 (0.138)	0.271 (0.031)		
SP14_11	SP71218	0.120	0.600 (0.102)	-0.922 (0.212)	0.211 (0.054)		
SP14_12	SP71214	0.040	0.923 (0.169)	0.218 (0.108)	0.185 (0.027)		
SP14_13	SP71213	0.030	0.939 (0.106)	0.721 (0.116)			

Appendix 12G: Mathematics Item Parameters from the eTIMSS 2019 Non-Invariant Model Calibration—Grade 4

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
ME01_01	ME51043	0.030	0.445 (0.035)	0.016 (0.064)			
ME01_02	ME51040	0.017	1.224 (0.134)	-0.026 (0.097)	0.415 (0.040)		
ME01_03	ME51008	0.088	1.309 (0.075)	1.185 (0.041)			
ME01_04A	ME51031A	0.067	1.565 (0.073)	0.257 (0.023)			
ME01_04B	ME51031B	0.058	1.787 (0.083)	0.290 (0.021)			
ME01_05	ME51508	0.084	1.262 (0.061)	0.324 (0.027)			
ME01_06A	ME51216A	0.038	1.199 (0.136)	0.649 (0.066)	0.290 (0.028)		
ME01_06B	ME51216B	0.103	0.741 (0.090)	-0.089 (0.168)	0.280 (0.060)		
ME01_07	ME51221	0.030	0.645 (0.071)	-0.717 (0.223)	0.251 (0.077)		
ME01_08	ME51115	0.037	0.641 (0.114)	1.749 (0.138)	0.118 (0.028)		
ME01_09A	ME51507A	0.040	0.759 (0.044)	-0.487 (0.049)			
ME01_09B	ME51507B	0.029	1.114 (0.061)	0.851 (0.037)			
ME02_01	ME71219	0.028	0.751 (0.067)	-0.865 (0.157)	0.186 (0.064)		
ME02_02	ME71021	0.034	1.106 (0.087)	0.103 (0.061)	0.132 (0.030)		
ME02_03	ME71167	0.100	1.565 (0.084)	1.052 (0.032)			
ME02_04	ME71041	0.039	1.145 (0.086)	-0.267 (0.070)	0.147 (0.037)		
ME02_05	ME71162	0.030	0.558 (0.027)	1.359 (0.055)		-0.492 (0.070)	0.492 (0.093)
ME02_06	ME71078	0.124	0.449 (0.036)	-0.884 (0.098)			
ME02_07	ME71090	0.034	1.197 (0.112)	0.338 (0.064)	0.226 (0.031)		
ME02_08	ME71151	0.038	0.743 (0.027)	0.783 (0.029)		-0.785 (0.062)	0.785 (0.069)
ME02_09	ME71119	0.072	0.666 (0.043)	-0.880 (0.070)			
ME02_10A	ME71217A	0.075	0.934 (0.053)	-0.921 (0.055)			
ME02_11	ME71142	0.030	1.114 (0.057)	-0.411 (0.036)			
ME02_12	ME71204	0.029	1.483 (0.072)	0.442 (0.025)			
ME03_01	ME61026	0.038	0.924 (0.075)	-0.637 (0.108)	0.166 (0.051)		
ME03_02	ME61273	0.031	0.793 (0.085)	0.374 (0.099)	0.174 (0.039)		
ME03_03	ME61034	0.030	1.202 (0.061)	0.661 (0.031)			
ME03_04	ME61040	0.049	1.636 (0.147)	0.737 (0.039)	0.176 (0.018)		
ME03_05	ME61228	0.080	0.739 (0.034)	1.210 (0.042)		-0.164 (0.053)	0.164 (0.072)
ME03_06	ME61166	0.031	1.182 (0.058)	-0.425 (0.033)			
ME03_07	ME61171	0.054	1.210 (0.109)	-0.143 (0.081)	0.269 (0.040)		
ME03_08	ME61080	0.039	0.686 (0.042)	0.505 (0.047)			
ME03_09	ME61222	0.060	0.828 (0.107)	0.295 (0.132)	0.319 (0.047)		
ME03_10	ME61076	0.073	0.443 (0.037)	-1.207 (0.116)			
ME03_11	ME61084	0.035	1.058 (0.056)	0.632 (0.034)			
ME04_01	ME71013	0.035	1.028 (0.100)	-0.210 (0.105)	0.275 (0.047)		
ME04_02	ME71026	0.027	1.177 (0.057)	0.142 (0.028)			
ME04_03	ME71036	0.019	0.923 (0.049)	-0.574 (0.042)			

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
ME04_04	ME71040	0.058	1.496 (0.111)	0.483 (0.036)	0.100 (0.017)		
ME04_05	ME71068	0.046	0.491 (0.072)	0.735 (0.170)	0.148 (0.051)		
ME04_06A	ME71075A	0.034	1.266 (0.061)	0.353 (0.027)			
ME04_06B	ME71075B	0.067	1.360 (0.070)	0.823 (0.031)			
ME04_07	ME71080	0.038	1.534 (0.149)	0.508 (0.050)	0.271 (0.024)		
ME04_08	ME71211	0.035	0.625 (0.040)	0.139 (0.047)			
ME04_09	ME71178	0.040	0.848 (0.048)	0.576 (0.041)			
ME04_10B	ME71135B	0.080	0.790 (0.045)	-0.220 (0.042)			
ME04_11	ME71201	0.128	0.743 (0.055)	1.599 (0.092)			
ME04_12	ME71175	0.039	0.790 (0.037)	-0.097 (0.030)		0.675 (0.050)	-0.675 (0.044)
ME05_01	ME51206	0.028	0.567 (0.039)	-0.819 (0.074)			
ME05_02	ME51052	0.045	0.775 (0.091)	0.041 (0.144)	0.262 (0.053)		
ME05_03	ME51049	0.051	1.491 (0.118)	0.218 (0.046)	0.175 (0.025)		
ME05_04	ME51045	0.071	1.148 (0.056)	0.084 (0.029)			
ME05_05	ME51098	0.033	1.065 (0.106)	0.818 (0.056)	0.136 (0.023)		
ME05_06	ME51030	0.029	1.006 (0.060)	1.165 (0.050)			
ME05_07	ME51502	0.043	0.997 (0.119)	1.215 (0.066)	0.130 (0.020)		
ME05_08	ME51224	0.031	0.837 (0.092)	0.003 (0.128)	0.256 (0.050)		
ME05_09	ME51207	0.035	0.786 (0.129)	0.860 (0.124)	0.333 (0.040)		
ME05_10	ME51427	0.020	1.144 (0.112)	0.728 (0.055)	0.160 (0.024)		
ME05_11	ME51533	0.038	1.085 (0.054)	0.181 (0.030)			
ME05_12	ME51080	0.121	1.104 (0.057)	0.231 (0.031)			
ME06_01	ME61018	0.029	0.944 (0.049)	0.106 (0.034)			
ME06_02	ME61274	0.028	0.743 (0.081)	-0.456 (0.180)	0.267 (0.067)		
ME06_03	ME61248	0.032	0.910 (0.040)	0.408 (0.025)		0.443 (0.038)	-0.443 (0.042)
ME06_04	ME61039	0.021	1.103 (0.055)	0.238 (0.030)			
ME06_05	ME61079	0.100	1.253 (0.066)	0.917 (0.035)			
ME06_06	ME61179	0.055	1.284 (0.107)	0.197 (0.057)	0.191 (0.029)		
ME06_07	ME61052	0.048	1.101 (0.086)	0.215 (0.057)	0.116 (0.027)		
ME06_08	ME61207	0.048	1.595 (0.115)	0.428 (0.035)	0.103 (0.017)		
ME06_09	ME61236	0.070	0.789 (0.045)	0.432 (0.040)			
ME06_10	ME61266	0.033	0.494 (0.020)	0.716 (0.040)		-0.892 (0.081)	0.892 (0.090)
ME06_11	ME61106	0.043	1.040 (0.106)	0.095 (0.094)	0.276 (0.041)		
ME07_01	ME51401	0.046	0.786 (0.045)	0.583 (0.042)			
ME07_02	ME51075	0.058	1.318 (0.187)	1.202 (0.065)	0.295 (0.020)		
ME07_03	ME51402	0.013	0.922 (0.049)	0.398 (0.035)			
ME07_04	ME51226	0.039	1.362 (0.142)	0.715 (0.054)	0.257 (0.023)		
ME07_05	ME51131	0.027	0.716 (0.042)	0.098 (0.042)			
ME07_06	ME51103	0.019	1.279 (0.120)	0.180 (0.068)	0.277 (0.032)		
ME07_07	ME51217	0.020	1.202 (0.061)	0.625 (0.031)			
ME07_08	ME51079	0.119	0.839 (0.048)	0.683 (0.042)			

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
ME07_09	ME51211	0.031	0.811 (0.096)	-0.234 (0.168)	0.333 (0.060)		
ME07_10	ME51102	0.031	1.038 (0.105)	0.793 (0.059)	0.146 (0.024)		
ME07_11	ME51009	0.022	0.829 (0.045)	-0.017 (0.038)			
ME07_12	ME51100	0.033	0.853 (0.109)	0.345 (0.124)	0.318 (0.045)		
ME08_01	ME71018	0.036	1.311 (0.109)	0.250 (0.052)	0.177 (0.026)		
ME08_02	ME71009	0.023	1.153 (0.056)	0.145 (0.029)			
ME08_03	ME71037	0.068	0.742 (0.042)	-0.031 (0.041)			
ME08_04	ME71051	0.030	1.220 (0.066)	0.955 (0.037)			
ME08_05	ME71064	0.040	0.697 (0.083)	0.630 (0.103)	0.148 (0.038)		
* ME08_06	ME71176	—	0.719 (0.068)	-1.067 (0.190)	0.223 (0.075)		
ME08_07	ME71169	0.062	1.197 (0.061)	0.662 (0.032)			
ME08_08	ME71083	0.045	1.177 (0.111)	0.396 (0.062)	0.213 (0.029)		
ME08_10	ME71184	0.036	2.115 (0.240)	1.037 (0.038)	0.238 (0.015)		
ME08_11	ME71141	0.148	0.871 (0.047)	0.271 (0.036)			
ME08_12	ME71194	0.050	0.732 (0.046)	-1.189 (0.074)			
ME08_13	ME71193	0.032	0.679 (0.024)	0.383 (0.028)		-0.779 (0.065)	0.779 (0.068)
ME08_14	ME71192	0.051	0.598 (0.023)	1.060 (0.040)		-1.329 (0.090)	1.329 (0.101)
ME09_01	ME61275	0.022	0.765 (0.077)	-0.543 (0.161)	0.234 (0.063)		
ME09_02	ME61027	0.037	0.795 (0.044)	-0.483 (0.046)			
ME09_03	ME61255	0.028	0.838 (0.034)	0.560 (0.026)		-0.131 (0.045)	0.131 (0.051)
ME09_04	ME61021	0.181	0.829 (0.055)	1.365 (0.069)			
ME09_05	ME61043	0.019	1.215 (0.059)	0.321 (0.028)			
ME09_06	ME61151	0.045	1.293 (0.100)	0.015 (0.055)	0.158 (0.029)		
ME09_07	ME61172	0.026	1.643 (0.140)	0.771 (0.036)	0.130 (0.016)		
ME09_08	ME61223	0.068	0.771 (0.072)	-0.299 (0.125)	0.170 (0.050)		
ME09_09	ME61269	0.037	0.816 (0.082)	-0.358 (0.141)	0.233 (0.057)		
ME09_10A	ME61081A	0.026	1.010 (0.056)	0.828 (0.040)			
ME09_10B	ME61081B	0.061	0.975 (0.059)	1.165 (0.053)			
ME10_02	ME71016	0.031	0.856 (0.046)	-0.119 (0.038)			
ME10_03	ME71163	0.018	1.875 (0.157)	0.956 (0.032)	0.084 (0.011)		
ME10_04	ME71045	0.028	1.288 (0.116)	0.383 (0.055)	0.207 (0.027)		
ME10_05	ME71213	0.030	0.835 (0.047)	0.423 (0.038)			
ME10_06	ME71070	0.036	0.431 (0.076)	0.228 (0.324)	0.251 (0.079)		
ME10_07	ME71181	0.042	0.892 (0.049)	0.506 (0.037)			
ME10_08	ME71179	0.051	1.079 (0.063)	1.052 (0.045)			
ME10_09	ME71067	0.031	0.561 (0.021)	1.046 (0.041)		-1.606 (0.100)	1.606 (0.110)
ME10_10A	ME71147A	0.017	1.327 (0.066)	-0.471 (0.032)			
ME10_10B	ME71147B	0.063	1.039 (0.053)	0.129 (0.031)			
ME10_11	ME71189	0.023	0.903 (0.055)	-1.301 (0.070)			

* eTIMSS items without a paperTIMSS counterpart

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
ME10_12A	ME71187A	0.073	0.720 (0.044)	-0.692 (0.058)			
ME10_12B	ME71187B	0.085	0.660 (0.042)	-0.042 (0.047)			
ME11_01	ME61178	0.033	0.952 (0.050)	-0.015 (0.034)			
ME11_02	ME61246	0.047	1.166 (0.094)	0.237 (0.054)	0.138 (0.027)		
ME11_03	ME61271	0.039	0.623 (0.040)	-0.549 (0.058)			
ME11_04	ME61256	0.035	0.953 (0.050)	0.168 (0.033)			
ME11_05	ME61182	0.044	1.180 (0.070)	1.216 (0.047)			
ME11_06	ME61049	0.030	1.029 (0.108)	-0.420 (0.129)	0.354 (0.053)		
ME11_07	ME61232	0.031	0.840 (0.125)	0.774 (0.108)	0.312 (0.037)		
ME11_08	ME61095	0.039	0.908 (0.048)	-0.005 (0.035)			
ME11_09	ME61264	0.031	0.568 (0.026)	0.447 (0.034)		-0.177 (0.063)	0.177 (0.068)
ME11_10	ME61108	0.048	0.587 (0.097)	0.871 (0.152)	0.210 (0.049)		
ME11_11A	ME61211A	0.071	1.234 (0.061)	0.340 (0.028)			
ME11_11B	ME61211B	0.052	1.376 (0.156)	0.794 (0.055)	0.273 (0.023)		
ME12_01	ME71001	0.029	0.815 (0.076)	-0.897 (0.164)	0.234 (0.069)		
ME12_02	ME71010	0.037	0.585 (0.038)	-0.299 (0.054)			
ME12_03	ME71062	0.039	1.412 (0.179)	1.248 (0.056)	0.179 (0.017)		
ME12_04A	ME71216A	0.099	1.422 (0.066)	-0.113 (0.026)			
ME12_04B	ME71216B	0.076	1.024 (0.053)	0.460 (0.033)			
ME12_05	ME71117	0.083	0.673 (0.040)	-0.066 (0.045)			
ME12_06	ME71071	0.043	0.883 (0.113)	0.480 (0.106)	0.296 (0.039)		
ME12_07	ME71098	0.045	0.698 (0.035)	0.928 (0.038)		0.178 (0.049)	-0.178 (0.065)
* ME12_08	ME71069	—	1.096 (0.057)	0.565 (0.032)			
ME12_09A	ME71134A	0.104	1.785 (0.124)	0.218 (0.033)	0.115 (0.019)		
ME12_09B	ME71134B	0.125	1.486 (0.073)	0.533 (0.025)			
ME12_10	ME71202	0.044	0.558 (0.038)	-0.527 (0.065)			
ME12_11	ME71190	0.028	1.011 (0.052)	-0.141 (0.034)			
ME12_12	ME71218	0.032	1.322 (0.078)	1.177 (0.042)			
ME13_01	ME61240	0.056	0.724 (0.044)	0.735 (0.049)			
ME13_02	ME61254	0.055	0.926 (0.048)	0.169 (0.034)			
ME13_03	ME61244	0.026	0.994 (0.097)	-0.084 (0.101)	0.258 (0.044)		
ME13_04	ME61041	0.021	1.261 (0.148)	0.935 (0.057)	0.243 (0.022)		
ME13_05	ME61173	0.033	0.633 (0.039)	-0.390 (0.053)			
ME13_06	ME61252	0.029	1.184 (0.102)	0.676 (0.046)	0.113 (0.020)		
ME13_07	ME61261	0.070	1.336 (0.063)	0.289 (0.026)			
ME13_08	ME61224	0.107	0.973 (0.055)	0.884 (0.042)			
ME13_09	ME61077	0.039	0.881 (0.079)	-0.190 (0.104)	0.179 (0.045)		
ME13_10A	ME61069A	0.034	0.737 (0.043)	-0.659 (0.054)			
ME13_10B	ME61069B	0.039	0.712 (0.042)	0.040 (0.043)			

* eTIMSS items without a paperTIMSS counterpart

Item		RMSD	Slope (a_i)		Location (b_i)		Guessing (c_i)		Step 1 (d_{i1})	Step 2 (d_{i2})
ME14_01	ME71024	0.035	0.877	(0.047)	0.266	(0.036)				
ME14_02	ME71008	0.028	1.132	(0.085)	-0.182	(0.065)	0.137	(0.033)		
ME14_03	ME71165	0.050	1.212	(0.103)	0.309	(0.055)	0.167	(0.026)		
ME14_04	ME71049	0.106	0.749	(0.042)	0.007	(0.041)				
ME14_05	ME71063	0.042	0.845	(0.046)	0.239	(0.037)				
ME14_06	ME71079	0.050	1.303	(0.142)	0.887	(0.053)	0.216	(0.021)		
ME14_07	ME71081	0.032	1.012	(0.051)	-0.036	(0.032)				
ME14_08	ME71094	0.048	0.697	(0.097)	0.563	(0.133)	0.238	(0.046)		
ME14_09	ME71177	0.057	0.521	(0.037)	0.201	(0.055)				
ME14_10	ME71206	0.049	0.592	(0.066)	-0.329	(0.193)	0.195	(0.064)		
ME14_11A	ME71138A	0.025	0.753	(0.043)	0.007	(0.041)				
ME14_11B	ME71138B	0.079	1.013	(0.058)	0.994	(0.045)				
ME14_12	ME71203	0.044	0.911	(0.135)	1.242	(0.084)	0.207	(0.026)		
ME14_13	ME71205	0.064	1.201	(0.061)	0.531	(0.030)				

Appendix 12H: Science Item Parameters from the eTIMSS 2019 Non-Invariant Model Calibration— Grade 4

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SE01_01	SE51054	0.022	0.899 (0.090)	-0.467 (0.136)	0.276 (0.057)		
SE01_02	SE51024	0.032	0.503 (0.041)	0.711 (0.066)			
SE01_03A	SE51132A	0.025	0.723 (0.057)	1.451 (0.088)			
SE01_03B	SE51132B	0.024	0.844 (0.058)	1.171 (0.060)			
SE01_04	SE51040	0.024	0.458 (0.039)	0.688 (0.072)			
SE01_05	SE51193	0.023	0.976 (0.096)	-0.125 (0.105)	0.257 (0.047)		
SE01_06	SE51063	0.040	1.095 (0.135)	0.847 (0.063)	0.217 (0.028)		
SE01_07	SE51012	0.033	1.212 (0.121)	0.301 (0.070)	0.268 (0.034)		
SE01_08	SE51115	0.036	1.113 (0.058)	0.239 (0.029)			
SE01_09	SE51180	0.039	0.795 (0.094)	-0.070 (0.147)	0.279 (0.055)		
SE01_10	SE51106	0.031	1.066 (0.122)	0.743 (0.063)	0.192 (0.029)		
SE01_11	SE51148	0.023	1.147 (0.104)	-0.011 (0.079)	0.231 (0.039)		
SE02_01	SE71002	0.079	0.615 (0.042)	0.352 (0.047)			
SE02_02	SE71402	0.035	1.133 (0.107)	-0.188 (0.093)	0.293 (0.044)		
SE02_03	SE71017	0.030	0.671 (0.043)	0.153 (0.044)			
SE02_04	SE71077	0.031	1.005 (0.053)	0.155 (0.031)			
SE02_05	SE71072	0.073	0.895 (0.120)	0.679 (0.092)	0.264 (0.037)		
SE02_06	SE71054	0.053	0.808 (0.047)	0.024 (0.039)			
SE02_07	SE71115	0.019	0.993 (0.134)	0.783 (0.079)	0.273 (0.032)		
SE02_08	SE71140	0.045	0.925 (0.090)	-0.100 (0.104)	0.226 (0.046)		
SE02_09	SE71128	0.035	0.978 (0.119)	0.183 (0.113)	0.360 (0.044)		
SE02_10	SE71147	0.029	1.069 (0.100)	-0.194 (0.096)	0.260 (0.045)		
SE02_11A	SE71920A	0.124	0.535 (0.043)	0.877 (0.070)			
SE02_11B	SE71920B	0.041	0.811 (0.051)	0.709 (0.043)			
SE02_12	SE71268	0.081	0.904 (0.114)	0.834 (0.075)	0.179 (0.031)		
SE03_01	SE61141	0.026	1.281 (0.136)	0.509 (0.062)	0.284 (0.030)		
SE03_02	SE61023	0.094	0.745 (0.045)	-0.353 (0.049)			
SE03_03	SE61054	0.039	0.508 (0.021)	0.825 (0.046)		1.507 (0.062)	-1.507 (0.087)
SE03_04	SE61007	0.021	0.722 (0.081)	-0.081 (0.146)	0.218 (0.055)		
SE03_05	SE61006	0.022	0.761 (0.046)	-0.701 (0.059)			
SE03_06	SE61108	0.038	0.928 (0.114)	0.265 (0.113)	0.318 (0.044)		
SE03_07	SE61109	0.030	0.614 (0.100)	0.727 (0.150)	0.220 (0.051)		
SE03_08	SE61080	0.035	1.161 (0.121)	0.391 (0.071)	0.264 (0.034)		
SE03_09	SE61088	0.027	0.711 (0.057)	1.517 (0.092)			
SE03_10	SE61151	0.031	0.940 (0.054)	0.516 (0.035)			
SE03_11	SE61150	0.017	0.590 (0.042)	0.385 (0.050)			
SE03_12	SE61169	0.074	1.234 (0.128)	0.331 (0.071)	0.285 (0.034)		

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SE04_01	SE71013	0.037	1.128 (0.104)	-0.646 (0.116)	0.305 (0.056)		
SE04_02	SE71902	0.090	0.338 (0.041)	1.971 (0.220)			
SE04_03	SE71076	0.046	1.022 (0.086)	-0.605 (0.105)	0.207 (0.051)		
SE04_04	SE71041	0.068	0.862 (0.042)	0.827 (0.029)		0.288 (0.039)	-0.288 (0.051)
SE04_05	SE71046	0.051	0.907 (0.051)	0.272 (0.034)			
SE04_06	SE71095	0.036	0.555 (0.040)	0.142 (0.052)			
SE04_07	SE71129	0.028	0.806 (0.090)	-0.581 (0.181)	0.322 (0.067)		
SE04_08	SE71102	0.069	1.019 (0.058)	0.685 (0.035)			
SE04_09	SE71124	0.038	0.963 (0.107)	0.398 (0.086)	0.234 (0.038)		
SE04_10	SE71112	0.044	0.703 (0.074)	-0.915 (0.207)	0.257 (0.074)		
SE04_11	SE71265	0.080	0.355 (0.082)	0.211 (0.502)	0.317 (0.095)		
SE04_12	SE71223	0.059	0.423 (0.056)	-2.841 (0.458)	0.234 (0.092)		
SE05_01	SE51044	0.062	0.427 (0.038)	0.516 (0.071)			
SE05_03	SE51003	0.099	0.616 (0.073)	0.403 (0.119)	0.133 (0.042)		
SE05_04	SE51168	0.022	0.699 (0.044)	-0.686 (0.062)			
SE05_05	SE51010	0.022	0.785 (0.047)	0.098 (0.039)			
SE05_06	SE51035	0.028	1.335 (0.191)	1.175 (0.060)	0.233 (0.021)		
SE05_07	SE51059	0.033	0.488 (0.039)	0.130 (0.058)			
SE05_08	SE51142	0.044	0.669 (0.102)	0.605 (0.141)	0.235 (0.050)		
SE05_09A	SE51131A	0.041	0.964 (0.091)	0.064 (0.088)	0.193 (0.040)		
SE05_09B	SE51131B	0.060	0.848 (0.099)	0.697 (0.077)	0.151 (0.033)		
SE05_10	SE51151	0.045	0.849 (0.052)	-0.949 (0.065)			
SE05_11	SE51157	0.022	0.816 (0.130)	1.035 (0.093)	0.214 (0.035)		
SE06_01	SE61071	0.033	0.245 (0.047)	-1.669 (0.687)	0.248 (0.095)		
SE06_02	SE61138	0.026	0.558 (0.040)	0.079 (0.052)			
SE06_03A	SE61016A	0.054	0.949 (0.104)	0.535 (0.076)	0.191 (0.034)		
SE06_03B	SE61016B	0.048	0.996 (0.056)	0.669 (0.035)			
SE06_04	SE61011	0.027	0.735 (0.045)	-0.463 (0.052)			
SE06_06	SE61083	0.021	0.730 (0.047)	-0.980 (0.073)			
SE06_07	SE61034	0.028	0.733 (0.053)	1.219 (0.070)			
SE06_08	SE61044	0.023	0.759 (0.048)	0.545 (0.042)			
SE06_09A	SE61142A	0.050	0.662 (0.045)	0.518 (0.047)			
SE06_09B	SE61142B	0.047	0.811 (0.057)	1.203 (0.064)			
SE06_10A	SE61115A	0.037	1.583 (0.153)	0.482 (0.049)	0.270 (0.026)		
SE06_10B	SE61115B	0.082	1.292 (0.145)	0.736 (0.055)	0.237 (0.026)		
SE07_01	SE51161	0.023	0.494 (0.108)	1.148 (0.197)	0.224 (0.059)		
SE07_02	SE51051	0.025	1.336 (0.264)	1.505 (0.088)	0.281 (0.020)		
SE07_03Z	SE51138Z	0.030	0.516 (0.040)	0.378 (0.056)			
SE07_04	SE51194	0.078	1.038 (0.059)	0.771 (0.035)			
SE07_05	SE51029	0.026	0.654 (0.126)	1.161 (0.129)	0.234 (0.044)		
SE07_06	SE51077	0.030	0.750 (0.045)	-0.267 (0.047)			

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SE07_07	SE51200	0.064	0.827 (0.061)	1.353 (0.072)			
SE07_08	SE51075	0.045	0.551 (0.040)	-0.511 (0.070)			
SE07_09	SE51065	0.023	0.835 (0.092)	-0.363 (0.158)	0.299 (0.061)		
SE07_10	SE51191	0.036	1.383 (0.145)	0.668 (0.052)	0.241 (0.026)		
SE07_11	SE51099	0.020	0.914 (0.102)	0.379 (0.093)	0.218 (0.040)		
SE07_12	SE51175	0.097	1.047 (0.070)	1.288 (0.055)			
SE08_02	SE71033	0.039	0.471 (0.094)	0.490 (0.291)	0.287 (0.074)		
SE08_03	SE71065	0.101	0.623 (0.041)	-0.236 (0.053)			
SE08_04	SE71025	0.035	0.474 (0.102)	0.708 (0.271)	0.285 (0.071)		
SE08_05	SE71081	0.065	0.956 (0.138)	1.005 (0.077)	0.231 (0.030)		
SE08_06	SE71056	0.038	0.511 (0.043)	1.064 (0.083)			
SE08_07	SE71145	0.051	0.563 (0.079)	-0.294 (0.261)	0.284 (0.076)		
SE08_08	SE71104	0.092	0.618 (0.042)	-0.621 (0.066)			
SE08_09	SE71144	0.038	0.641 (0.075)	0.199 (0.137)	0.166 (0.049)		
SE08_10	SE71150	0.043	1.051 (0.055)	-0.295 (0.036)			
SE08_11	SE71201	0.038	0.972 (0.103)	-0.088 (0.116)	0.306 (0.049)		
SE08_12	SE71237	0.040	1.061 (0.056)	0.280 (0.030)			
SE08_13	SE71260	0.035	0.707 (0.098)	1.089 (0.090)	0.122 (0.032)		
SE09_01	SE61135	0.040	0.831 (0.082)	-0.411 (0.135)	0.231 (0.055)		
SE09_02	SE61069	0.022	0.361 (0.035)	-0.550 (0.101)			
SE09_03	SE61134	0.037	0.743 (0.075)	0.226 (0.099)	0.138 (0.039)		
SE09_04	SE61140	0.029	1.131 (0.144)	0.714 (0.072)	0.299 (0.031)		
SE09_05	SE61019	0.047	0.798 (0.051)	0.835 (0.047)			
SE09_06	SE61022	0.024	0.618 (0.080)	0.045 (0.180)	0.225 (0.060)		
SE09_07	SE61036	0.027	0.993 (0.061)	0.984 (0.044)			
SE09_08	SE61160	0.028	0.871 (0.052)	-0.909 (0.061)			
SE09_09	SE61159	0.078	0.662 (0.045)	-1.239 (0.092)			
SE09_10	SE61091	0.021	0.479 (0.028)	1.134 (0.058)		-0.190 (0.071)	0.190 (0.092)
SE09_11	SE61118	0.047	1.171 (0.130)	0.665 (0.061)	0.225 (0.029)		
SE09_12	SE61097	0.029	0.766 (0.097)	0.436 (0.116)	0.216 (0.045)		
SE10_01	SE71009	0.025	0.572 (0.026)	-0.508 (0.041)		1.156 (0.078)	-1.156 (0.054)
SE10_02	SE71093	0.050	0.591 (0.041)	-0.661 (0.070)			
SE10_03	SE71069	0.049	1.534 (0.223)	1.107 (0.055)	0.306 (0.020)		
SE10_04	SE71051	0.052	0.705 (0.048)	0.822 (0.053)			
SE10_05	SE71039	0.036	0.837 (0.078)	0.178 (0.085)	0.137 (0.037)		
SE10_06	SE71080	0.034	0.703 (0.110)	0.886 (0.113)	0.209 (0.042)		
SE10_07	SE71137	0.035	0.742 (0.045)	-0.355 (0.049)			
SE10_08	SE71103	0.058	1.130 (0.116)	0.329 (0.074)	0.257 (0.035)		
SE10_09	SE71106	0.047	0.706 (0.046)	0.546 (0.045)			
SE10_10	SE71100	0.055	0.801 (0.102)	-0.089 (0.165)	0.346 (0.057)		

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
* SE10_11	SE71921	—	1.641 (0.279)	1.361 (0.064)	0.262 (0.018)		
SE10_12	SE71220	0.040	0.852 (0.122)	0.815 (0.093)	0.238 (0.037)		
SE10_13	SE71254	0.091	0.673 (0.044)	0.290 (0.044)			
SE11_01	SE61132	0.053	0.785 (0.121)	0.605 (0.127)	0.312 (0.045)		
SE11_02	SE61120	0.028	0.896 (0.099)	0.430 (0.086)	0.198 (0.037)		
SE11_03	SE61025	0.056	0.668 (0.043)	-0.037 (0.046)			
SE11_04A	SE61133A	0.051	1.625 (0.147)	0.286 (0.050)	0.284 (0.028)		
SE11_04B	SE61133B	0.072	1.492 (0.130)	0.947 (0.036)	0.074 (0.013)		
SE11_05	SE61074	0.033	0.719 (0.045)	0.359 (0.042)			
SE11_06	SE61093	0.016	0.764 (0.032)	-0.054 (0.031)		0.935 (0.053)	-0.935 (0.044)
SE11_07	SE61161	0.034	0.508 (0.041)	0.626 (0.063)			
SE11_08A	SE61042A	0.034	1.399 (0.155)	0.843 (0.048)	0.213 (0.022)		
SE11_08B	SE61042B	0.042	0.931 (0.116)	0.808 (0.074)	0.187 (0.031)		
SE11_09A	SE61041A	0.021	0.869 (0.050)	0.171 (0.036)			
SE11_09B	SE61041B	0.041	0.678 (0.045)	0.328 (0.044)			
SE11_10	SE61155	0.035	0.830 (0.092)	-0.348 (0.155)	0.289 (0.060)		
SE12_01	SE71031	0.024	0.617 (0.041)	-0.027 (0.049)			
SE12_02	SE71090	0.063	0.784 (0.047)	0.207 (0.038)			
SE12_03	SE71048	0.033	1.564 (0.196)	1.094 (0.047)	0.219 (0.018)		
SE12_04	SE71071	0.063	1.055 (0.065)	1.054 (0.044)			
SE12_05	SE71011	0.025	1.109 (0.093)	-0.464 (0.093)	0.218 (0.047)		
SE12_06	SE71142	0.035	0.949 (0.126)	0.423 (0.108)	0.349 (0.041)		
SE12_07	SE71138	0.031	0.640 (0.043)	-0.778 (0.071)			
SE12_08	SE71127	0.062	1.156 (0.114)	0.173 (0.078)	0.272 (0.037)		
SE12_10	SE71500	0.059	1.090 (0.114)	0.569 (0.064)	0.199 (0.030)		
SE12_11	SE71257	0.046	1.824 (0.325)	1.224 (0.060)	0.411 (0.018)		
SE12_12	SE71222	0.035	1.040 (0.056)	0.267 (0.030)			
SE12_13	SE71252	0.030	0.989 (0.115)	0.296 (0.098)	0.296 (0.041)		
SE13_02	SE61014	0.047	0.493 (0.040)	0.639 (0.065)			
SE13_03	SE61056	0.031	0.726 (0.046)	-0.837 (0.069)			
SE13_04	SE61015	0.051	0.519 (0.038)	-0.408 (0.068)			
SE13_05	SE61113	0.020	0.798 (0.054)	1.052 (0.056)			
SE13_06	SE61107	0.027	1.035 (0.123)	0.701 (0.070)	0.228 (0.031)		
SE13_07	SE61046	0.058	1.515 (0.173)	0.923 (0.046)	0.231 (0.021)		
SE13_08	SE61047	0.029	0.726 (0.091)	-0.333 (0.198)	0.331 (0.066)		
SE13_09	SE61048	0.031	1.556 (0.156)	0.564 (0.049)	0.278 (0.025)		
SE13_10	SE61096	0.041	1.172 (0.142)	0.681 (0.068)	0.288 (0.030)		
SE13_11	SE61124	0.037	0.606 (0.050)	1.376 (0.093)			
SE13_12	SE61116	0.037	0.718 (0.046)	0.307 (0.042)			

* eTIMSS items without a paperTIMSS counterpart

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SE14_01	SE71063	0.071	0.420 (0.036)	0.078 (0.067)			
SE14_02	SE71900	0.072	0.517 (0.078)	-0.138 (0.271)	0.269 (0.074)		
SE14_04	SE71043	0.030	0.693 (0.054)	1.435 (0.088)			
SE14_05	SE71005	0.038	0.818 (0.047)	-0.588 (0.052)			
SE14_06	SE71118	0.057	1.170 (0.142)	0.992 (0.056)	0.187 (0.023)		
SE14_07	SE71139	0.030	1.011 (0.114)	0.020 (0.112)	0.354 (0.045)		
SE14_08	SE71114	0.044	0.711 (0.044)	-0.596 (0.058)			
SE14_09	SE71131	0.051	0.454 (0.037)	0.090 (0.062)			
SE14_10	SE71152	0.034	1.080 (0.130)	0.534 (0.079)	0.302 (0.034)		
SE14_11	SE71218	0.062	0.742 (0.076)	-1.147 (0.208)	0.265 (0.077)		
SE14_12	SE71214	0.021	1.078 (0.100)	0.163 (0.075)	0.209 (0.036)		
SE14_13	SE71213	0.068	0.769 (0.055)	1.179 (0.065)			

Appendix 12I: Mathematics Item Parameters from the eTIMSS 2019 Non-Invariant Model Calibration—Grade 8

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
ME01_01	ME52024	0.059	1.440 (0.157)	0.620 (0.062)	0.231 (0.029)		
ME01_02A	ME52058A	0.050	1.181 (0.071)	-0.248 (0.039)			
ME01_02B	ME52058B	0.059	1.368 (0.084)	1.044 (0.039)			
ME01_03	ME52125	0.045	1.166 (0.118)	0.772 (0.061)	0.124 (0.025)		
ME01_04	ME52229	0.129	1.224 (0.070)	0.379 (0.034)			
ME01_05	ME52063	0.043	1.674 (0.170)	0.634 (0.049)	0.196 (0.024)		
ME01_06	ME52072	0.050	1.229 (0.116)	0.170 (0.072)	0.165 (0.037)		
ME01_07A	ME52146A	0.068	0.865 (0.055)	0.437 (0.045)			
ME01_07B	ME52146B	0.070	1.625 (0.110)	1.364 (0.043)			
ME01_08	ME52092	0.042	1.379 (0.226)	1.661 (0.077)	0.186 (0.018)		
ME01_09	ME52046	0.071	0.946 (0.162)	1.762 (0.106)	0.143 (0.023)		
ME01_10	ME52083	0.044	1.328 (0.163)	1.046 (0.063)	0.200 (0.024)		
ME01_11	ME52082	0.031	1.207 (0.124)	0.285 (0.078)	0.203 (0.038)		
ME01_12	ME52161	0.037	1.034 (0.104)	-0.147 (0.108)	0.200 (0.052)		
ME01_13A	ME52418A	0.063	1.636 (0.161)	0.809 (0.045)	0.139 (0.020)		
ME01_13B	ME52418B	0.071	1.487 (0.172)	0.773 (0.061)	0.245 (0.027)		
ME02_01	ME72007	0.049	0.760 (0.038)	1.016 (0.040)		-0.143 (0.061)	0.143 (0.075)
ME02_02	ME72025	0.038	1.636 (0.175)	0.678 (0.053)	0.223 (0.025)		
ME02_03	ME72017	0.080	1.377 (0.091)	1.223 (0.044)			
ME02_04	ME72190	0.057	0.969 (0.059)	-0.005 (0.042)			
ME02_05	ME72068	0.039	1.357 (0.140)	0.131 (0.078)	0.250 (0.039)		
ME02_06	ME72076	0.052	1.297 (0.140)	0.766 (0.061)	0.175 (0.026)		
ME02_07	ME72056	0.033	1.288 (0.074)	0.555 (0.034)			
ME02_08	ME72098	0.038	1.906 (0.111)	0.765 (0.028)			
ME02_09	ME72103	0.025	1.280 (0.135)	0.688 (0.062)	0.170 (0.027)		
ME02_10	ME72121	0.073	1.141 (0.067)	-0.092 (0.038)			
ME02_11	ME72180	0.103	0.505 (0.043)	0.313 (0.070)			
ME02_12	ME72198	0.026	1.353 (0.078)	0.621 (0.034)			
ME02_13	ME72227	0.037	1.371 (0.078)	0.499 (0.033)			
ME02_14	ME72170	0.042	0.785 (0.052)	0.129 (0.049)			
ME02_15	ME72209	0.040	0.976 (0.073)	1.541 (0.072)			
ME03_01	ME62005	0.082	1.000 (0.146)	0.824 (0.104)	0.294 (0.037)		
ME03_02	ME62139	0.065	0.977 (0.061)	0.803 (0.045)			
ME03_03	ME62164	0.073	1.514 (0.143)	0.320 (0.057)	0.183 (0.030)		
ME03_04	ME62142	0.037	0.859 (0.055)	-0.160 (0.048)			
ME03_05	ME62084	0.026	1.715 (0.236)	1.486 (0.055)	0.142 (0.015)		
ME03_06	ME62351	0.034	0.843 (0.171)	1.541 (0.120)	0.251 (0.033)		
ME03_07	ME62223	0.048	1.633 (0.154)	0.028 (0.061)	0.212 (0.035)		

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
ME03_08	ME62027	0.037	0.739 (0.051)	0.722 (0.055)			
ME03_09	ME62174	0.034	1.733 (0.239)	0.976 (0.061)	0.342 (0.023)		
ME03_10	ME62244	0.050	1.131 (0.066)	0.575 (0.038)			
ME03_11	ME62261	0.028	2.219 (0.315)	1.516 (0.046)	0.140 (0.013)		
ME03_12	ME62300	0.033	0.748 (0.033)	0.443 (0.032)		-0.368 (0.066)	0.368 (0.069)
ME03_13	ME62254	0.032	0.852 (0.063)	1.461 (0.075)			
ME03_14A	ME62132A	0.045	1.084 (0.066)	-0.180 (0.041)			
ME03_14B	ME62132B	0.054	0.896 (0.140)	0.927 (0.115)	0.270 (0.040)		
ME04_01	ME72178	0.070	1.230 (0.073)	0.805 (0.038)			
ME04_02	ME72234	0.087	1.454 (0.186)	1.065 (0.061)	0.240 (0.023)		
ME04_03	ME72020	0.039	0.751 (0.035)	0.059 (0.033)		-0.188 (0.066)	0.188 (0.062)
ME04_04	ME72027	0.027	1.325 (0.129)	0.224 (0.070)	0.199 (0.036)		
ME04_05	ME72052	0.039	1.291 (0.086)	1.334 (0.049)			
ME04_06	ME72067	0.084	1.665 (0.180)	0.355 (0.062)	0.292 (0.032)		
ME04_07A	ME72083A	0.029	1.173 (0.069)	-0.101 (0.038)			
ME04_07B	ME72083B	0.089	0.873 (0.098)	0.883 (0.080)	0.111 (0.030)		
ME04_08A	ME72108A	0.094	0.797 (0.052)	0.324 (0.048)			
ME04_08B	ME72108B	0.148	1.128 (0.070)	0.967 (0.044)			
ME04_09	ME72181	0.120	0.984 (0.064)	1.047 (0.051)			
ME04_10	ME72126	0.036	0.742 (0.032)	0.967 (0.037)		-0.744 (0.078)	0.744 (0.087)
ME04_11	ME72164	0.092	0.657 (0.054)	1.488 (0.094)			
ME04_12A	ME72185A	0.138	1.421 (0.084)	0.813 (0.035)			
ME04_12B	ME72185B	0.119	1.289 (0.077)	0.742 (0.037)			
ME05_01	ME52413	0.145	1.322 (0.131)	0.369 (0.067)	0.193 (0.033)		
ME05_02	ME52134	0.027	1.348 (0.122)	-0.136 (0.074)	0.176 (0.041)		
ME05_03	ME52078	0.044	0.910 (0.126)	0.987 (0.094)	0.196 (0.035)		
ME05_04	ME52034	0.052	1.120 (0.136)	0.625 (0.086)	0.250 (0.036)		
ME05_05A	ME52174A	0.041	0.964 (0.059)	0.317 (0.041)			
ME05_05B	ME52174B	0.047	1.066 (0.070)	1.184 (0.051)			
ME05_06	ME52130	0.035	1.415 (0.155)	0.973 (0.054)	0.159 (0.022)		
ME05_07	ME52073	0.080	1.417 (0.141)	0.706 (0.054)	0.157 (0.025)		
ME05_08	ME52110	0.050	1.586 (0.092)	0.785 (0.031)			
ME05_09	ME52105	0.023	1.154 (0.082)	1.482 (0.059)			
ME05_10	ME52407	0.061	1.143 (0.138)	0.358 (0.099)	0.297 (0.043)		
ME05_11	ME52036	0.121	0.815 (0.056)	0.926 (0.055)			
ME05_12	ME52502	0.072	1.047 (0.063)	-0.036 (0.040)			
ME05_13	ME52117	0.060	0.638 (0.065)	2.338 (0.174)			
ME05_14	ME52426	0.058	0.862 (0.094)	-0.342 (0.154)	0.227 (0.066)		
ME06_01	ME62150	0.033	1.062 (0.065)	-0.231 (0.042)			
ME06_02	ME62335	0.030	1.439 (0.133)	-0.025 (0.069)	0.192 (0.039)		
ME06_03	ME62219	0.030	2.183 (0.237)	0.888 (0.040)	0.198 (0.018)		

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
ME06_04	ME62002	0.072	0.454 (0.043)	0.839 (0.089)			
ME06_05	ME62149	0.085	1.263 (0.125)	0.760 (0.055)	0.121 (0.024)		
ME06_06	ME62241	0.037	1.568 (0.090)	0.729 (0.031)			
ME06_08	ME62105	0.026	0.765 (0.031)	0.842 (0.033)		-1.536 (0.112)	1.536 (0.117)
ME06_09	ME62040	0.051	0.680 (0.134)	1.270 (0.148)	0.235 (0.047)		
ME06_10	ME62288	0.022	0.804 (0.036)	1.172 (0.038)		-0.914 (0.085)	0.914 (0.095)
ME06_11	ME62173	0.030	1.231 (0.075)	0.904 (0.040)			
ME06_12	ME62133	0.045	1.219 (0.140)	0.749 (0.069)	0.197 (0.030)		
ME06_13A	ME62123A	0.041	1.531 (0.165)	0.406 (0.065)	0.264 (0.032)		
ME06_13B	ME62123B	0.040	1.472 (0.160)	0.831 (0.054)	0.178 (0.024)		
ME07_01	ME52079	0.032	1.140 (0.161)	0.671 (0.099)	0.352 (0.037)		
ME07_02	ME52204	0.025	1.100 (0.129)	0.541 (0.086)	0.229 (0.037)		
ME07_03	ME52364	0.049	1.256 (0.074)	-0.219 (0.037)			
ME07_04	ME52215	0.111	0.938 (0.058)	0.133 (0.042)			
ME07_05	ME52147	0.038	1.580 (0.197)	0.890 (0.059)	0.277 (0.024)		
ME07_06	ME52067	0.063	1.340 (0.139)	0.204 (0.076)	0.243 (0.038)		
ME07_07	ME52068	0.034	1.400 (0.160)	1.269 (0.054)	0.105 (0.016)		
ME07_08	ME52087	0.018	1.774 (0.115)	1.163 (0.035)			
ME07_09	ME52048	0.169	0.779 (0.053)	0.660 (0.052)			
ME07_10	ME52039	0.055	1.323 (0.075)	0.431 (0.033)			
ME07_11	ME52208	0.047	2.078 (0.195)	1.179 (0.035)	0.050 (0.009)		
ME07_12A	ME52419A	0.088	0.882 (0.079)	0.028 (0.088)	0.106 (0.037)		
ME07_12B	ME52419B	0.069	1.289 (0.117)	-0.397 (0.085)	0.178 (0.048)		
ME07_13	ME52115	0.037	1.637 (0.139)	0.475 (0.042)	0.103 (0.021)		
ME07_14	ME52421	0.045	0.736 (0.052)	0.729 (0.056)			
ME08_01	ME72002	0.096	1.378 (0.081)	0.792 (0.035)			
ME08_02	ME72188	0.035	1.141 (0.120)	0.808 (0.062)	0.129 (0.026)		
ME08_03	ME72035	0.080	1.198 (0.072)	0.752 (0.038)			
ME08_04	ME72055	0.087	1.306 (0.080)	0.957 (0.039)			
ME08_05	ME72222	0.073	1.133 (0.128)	0.623 (0.075)	0.197 (0.033)		
ME08_06	ME72090	0.049	1.931 (0.230)	0.946 (0.047)	0.243 (0.020)		
ME08_07	ME72233	0.050	1.117 (0.178)	0.917 (0.099)	0.369 (0.034)		
ME08_08A	ME72106A	0.096	0.967 (0.059)	-0.030 (0.042)			
ME08_08B	ME72106B	0.130	1.588 (0.095)	0.879 (0.033)			
ME08_08C	ME72106C	0.129	1.839 (0.119)	1.140 (0.034)			
ME08_09A	ME72128A	0.045	1.008 (0.062)	0.664 (0.042)			
ME08_09B	ME72128B	0.060	0.900 (0.045)	0.886 (0.033)		-0.022 (0.052)	0.022 (0.062)
ME08_10	ME72119	0.053	1.075 (0.064)	0.352 (0.038)			
ME08_11A	ME72153A	0.041	1.327 (0.076)	0.352 (0.033)			
ME08_11B	ME72153B	0.043	1.457 (0.101)	1.359 (0.046)			
ME08_12	ME72172	0.042	1.065 (0.098)	0.292 (0.072)	0.117 (0.032)		

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
ME09_01	ME62329	0.045	0.683 (0.077)	-0.925 (0.223)	0.233 (0.082)		
ME09_02	ME62151	0.047	1.256 (0.075)	0.848 (0.038)			
ME09_03	ME62346	0.079	0.997 (0.063)	0.868 (0.046)			
ME09_04	ME62212	0.047	1.379 (0.172)	1.262 (0.056)	0.139 (0.018)		
ME09_05	ME62056	0.022	1.366 (0.088)	1.170 (0.042)			
ME09_06	ME62317	0.090	1.378 (0.086)	1.070 (0.039)			
ME09_07	ME62350	0.027	1.380 (0.199)	1.575 (0.068)	0.120 (0.016)		
ME09_08	ME62078	0.034	1.568 (0.089)	0.683 (0.030)			
ME09_09	ME62284	0.060	0.724 (0.127)	0.750 (0.170)	0.300 (0.054)		
ME09_10	ME62245	0.065	1.449 (0.158)	0.795 (0.055)	0.191 (0.024)		
ME09_11	ME62287	0.045	1.369 (0.101)	1.528 (0.055)			
ME09_12A	ME62345A	0.059	0.731 (0.041)	0.669 (0.037)		0.400 (0.057)	-0.400 (0.066)
ME09_13	ME62115	0.023	1.572 (0.223)	1.382 (0.058)	0.186 (0.018)		
ME10_01	ME72187	0.117	0.923 (0.057)	0.014 (0.043)			
ME10_02	ME72022	0.055	1.508 (0.213)	1.210 (0.062)	0.253 (0.021)		
ME10_04	ME72045	0.042	1.294 (0.073)	0.525 (0.034)			
ME10_05	ME72049	0.121	0.797 (0.053)	-0.347 (0.055)			
ME10_06	ME72069	0.119	1.640 (0.091)	0.178 (0.028)			
ME10_07	ME72074	0.065	1.543 (0.094)	0.984 (0.035)			
ME10_08	ME72013	0.064	1.325 (0.126)	0.731 (0.052)	0.111 (0.022)		
ME10_09	ME72095	0.079	1.258 (0.074)	0.740 (0.037)			
ME10_10	ME72109	0.080	1.664 (0.112)	1.286 (0.040)			
ME10_11	ME72125	0.038	2.069 (0.183)	0.718 (0.035)	0.099 (0.015)		
ME10_12	ME72196	0.057	1.311 (0.076)	0.679 (0.035)			
ME10_13	ME72237	0.051	0.729 (0.095)	0.081 (0.178)	0.237 (0.064)		
ME10_14	ME72232	0.057	0.592 (0.046)	-0.101 (0.065)			
ME10_15	ME72206	0.040	1.307 (0.090)	1.360 (0.050)			
ME11_01	ME62271	0.063	1.526 (0.150)	0.573 (0.053)	0.171 (0.025)		
ME11_02	ME62152	0.049	1.116 (0.065)	0.459 (0.037)			
ME11_03	ME62215	0.042	0.889 (0.041)	0.721 (0.031)		-0.129 (0.054)	0.129 (0.061)
ME11_04	ME62143	0.026	1.545 (0.091)	0.851 (0.033)			
ME11_05	ME62230	0.033	1.506 (0.236)	1.430 (0.067)	0.240 (0.019)		
ME11_06	ME62095	0.029	1.620 (0.167)	0.597 (0.053)	0.207 (0.025)		
ME11_07	ME62076	0.027	1.703 (0.186)	0.292 (0.063)	0.301 (0.032)		
ME11_08	ME62030	0.041	0.516 (0.043)	0.227 (0.069)			
ME11_09	ME62171	0.078	0.887 (0.097)	0.313 (0.106)	0.169 (0.044)		
ME11_10	ME62301	0.034	1.031 (0.067)	1.125 (0.052)			
ME11_11	ME62194	0.041	0.864 (0.103)	-0.352 (0.175)	0.287 (0.072)		
ME11_12	ME62344	0.025	0.966 (0.064)	1.114 (0.055)			
ME11_13	ME62320	0.056	1.771 (0.153)	0.626 (0.039)	0.097 (0.018)		
ME11_14	ME62296	0.037	1.144 (0.067)	0.113 (0.037)			

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
ME12_01	ME72001	0.058	1.774 (0.102)	0.723 (0.029)			
ME12_02	ME72019	0.053	1.489 (0.084)	0.522 (0.031)			
ME12_03	ME72189	0.109	1.408 (0.152)	0.507 (0.065)	0.236 (0.030)		
ME12_04	ME72024	0.089	1.222 (0.072)	0.722 (0.037)			
ME12_05	ME72043	0.045	2.027 (0.208)	0.702 (0.042)	0.197 (0.020)		
ME12_06	ME72221	0.082	1.695 (0.189)	0.635 (0.055)	0.263 (0.026)		
ME12_07	ME72220	0.078	1.500 (0.160)	1.061 (0.049)	0.120 (0.018)		
ME12_08	ME72225	0.034	1.250 (0.072)	0.521 (0.035)			
ME12_09A	ME72110A	0.079	1.377 (0.081)	0.788 (0.035)			
ME12_09B	ME72110B	0.102	1.837 (0.118)	1.132 (0.034)			
ME12_10	ME72150	0.063	1.561 (0.197)	0.404 (0.078)	0.391 (0.033)		
ME12_11	ME72139	0.031	1.255 (0.079)	1.050 (0.043)			
ME12_12	ME72229	0.018	0.973 (0.048)	1.444 (0.039)		-1.088 (0.104)	1.088 (0.114)
ME12_13	ME72171	0.033	1.282 (0.074)	0.435 (0.034)			
ME12_14A	ME72211A	0.076	1.656 (0.178)	0.325 (0.063)	0.276 (0.032)		
ME13_01	ME62001	0.090	0.964 (0.146)	1.043 (0.097)	0.257 (0.035)		
ME13_02	ME62214	0.052	1.028 (0.062)	0.551 (0.040)			
ME13_03	ME62146	0.038	1.408 (0.141)	0.832 (0.051)	0.133 (0.022)		
ME13_04	ME62154	0.044	1.395 (0.079)	0.033 (0.032)			
ME13_05	ME62067	0.042	0.992 (0.122)	0.129 (0.130)	0.303 (0.053)		
ME13_06	ME62341	0.036	0.892 (0.172)	1.619 (0.113)	0.225 (0.029)		
ME13_07	ME62242	0.025	1.204 (0.119)	0.269 (0.076)	0.185 (0.037)		
ME13_08A	ME62250A	0.057	1.103 (0.064)	0.303 (0.037)			
ME13_08B	ME62250B	0.056	1.399 (0.085)	0.978 (0.037)			
ME13_09	ME62170	0.051	0.582 (0.037)	0.978 (0.050)		0.644 (0.066)	-0.644 (0.087)
ME13_10	ME62192	0.031	1.129 (0.073)	1.145 (0.048)			
ME13_11	ME62072	0.035	0.980 (0.059)	0.189 (0.041)			
ME13_13	ME62120	0.067	1.284 (0.146)	0.765 (0.066)	0.206 (0.028)		
ME14_01	ME72005	0.052	0.927 (0.103)	0.424 (0.099)	0.174 (0.042)		
ME14_02	ME72021	0.106	1.200 (0.070)	0.470 (0.035)			
ME14_03	ME72026	0.027	0.788 (0.053)	0.627 (0.050)			
ME14_04A	ME72041A	0.094	1.129 (0.066)	0.330 (0.036)			
ME14_04B	ME72041B	0.097	1.099 (0.066)	0.608 (0.038)			
ME14_05	ME72223	0.052	2.059 (0.221)	0.582 (0.046)	0.265 (0.024)		
ME14_06	ME72094	0.035	1.255 (0.073)	-0.103 (0.036)			
ME14_07	ME72059	0.033	1.419 (0.081)	0.537 (0.031)			
ME14_08	ME72080	0.050	2.159 (0.212)	0.909 (0.036)	0.132 (0.016)		
ME14_09	ME72081	0.091	1.106 (0.071)	1.119 (0.048)			
ME14_10	ME72140	0.066	0.769 (0.052)	0.149 (0.049)			
ME14_11	ME72120	0.118	1.440 (0.089)	1.085 (0.038)			
ME14_12	ME72131	0.033	1.610 (0.107)	1.332 (0.041)			

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
ME14_13	ME72147	0.031	1.398 (0.090)	1.214 (0.042)			
ME14_14	ME72154	0.041	1.166 (0.123)	0.098 (0.094)	0.235 (0.046)		
ME14_15	ME72192	0.034	0.899 (0.110)	0.418 (0.117)	0.220 (0.047)		
ME14_16	ME72161	0.064	1.083 (0.067)	0.814 (0.042)			

Appendix 12J:Science Item Parameters from the eTIMSS 2019 Non-Invariant Model Calibration Grade 8

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SE01_01	SE52006	0.034	0.585 (0.036)	-0.234 (0.047)		0.706 (0.084)	-0.706 (0.066)
SE01_02	SE52069	0.047	1.004 (0.173)	0.846 (0.110)	0.359 (0.040)		
SE01_03	SE52012	0.034	1.025 (0.120)	0.360 (0.095)	0.223 (0.042)		
SE01_04	SE52021	0.033	0.998 (0.066)	0.713 (0.042)			
SE01_05Z	SE52095Z	0.037	0.462 (0.044)	-0.059 (0.080)			
SE01_07	SE52054	0.037	0.739 (0.053)	-0.250 (0.059)			
SE01_08	SE52150	0.030	0.937 (0.158)	1.170 (0.091)	0.213 (0.033)		
SE01_09A	SE52243A	0.061	0.664 (0.052)	0.612 (0.057)			
SE01_09B	SE52243B	0.049	0.781 (0.056)	0.602 (0.049)			
SE01_09C	SE52243C	0.064	0.615 (0.111)	1.215 (0.134)	0.149 (0.043)		
SE01_10	SE52206	0.050	1.077 (0.110)	0.319 (0.078)	0.162 (0.036)		
SE01_11A	SE52112A	0.038	0.761 (0.096)	0.058 (0.153)	0.221 (0.057)		
SE01_11B	SE52112B	0.022	1.020 (0.068)	0.760 (0.042)			
SE01_12	SE52294	0.025	1.032 (0.109)	-0.048 (0.107)	0.221 (0.049)		
SE02_01	SE72072	0.027	0.909 (0.119)	0.517 (0.106)	0.222 (0.044)		
SE02_02	SE72029	0.051	1.336 (0.192)	0.802 (0.076)	0.339 (0.032)		
SE02_03	SE72902	0.067	1.230 (0.073)	0.236 (0.033)			
SE02_04	SE72077	0.067	0.709 (0.104)	0.441 (0.154)	0.224 (0.055)		
SE02_05A	SE72900A	0.045	0.865 (0.062)	0.875 (0.051)			
SE02_05B	SE72900B	0.094	0.678 (0.070)	1.953 (0.147)			
SE02_06	SE72103	0.081	0.715 (0.052)	0.160 (0.052)			
SE02_07	SE72110	0.037	0.883 (0.062)	0.800 (0.048)			
SE02_08	SE72130	0.066	0.752 (0.058)	0.963 (0.061)			
SE02_09	SE72148	0.069	1.290 (0.159)	0.939 (0.059)	0.179 (0.026)		
SE02_10	SE72200	0.029	0.906 (0.118)	0.786 (0.087)	0.164 (0.036)		
SE02_11	SE72232	0.040	1.517 (0.086)	0.303 (0.029)			
SE02_12	SE72275	0.034	1.064 (0.101)	-0.333 (0.101)	0.182 (0.049)		
SE02_13	SE72244	0.030	1.010 (0.065)	0.513 (0.039)			
SE02_14	SE72301	0.068	0.646 (0.128)	1.091 (0.151)	0.220 (0.050)		
SE02_15	SE72721	0.068	1.186 (0.116)	0.097 (0.080)	0.190 (0.040)		
SE02_16	SE72335	0.038	1.145 (0.164)	0.747 (0.087)	0.301 (0.036)		
SE03_01	SE62055	0.045	1.071 (0.141)	0.026 (0.135)	0.394 (0.052)		
SE03_02	SE62007	0.023	1.335 (0.138)	0.450 (0.063)	0.194 (0.031)		
SE03_03	SE62275	0.053	0.937 (0.068)	1.026 (0.053)			
SE03_04	SE62225	0.033	1.255 (0.257)	1.424 (0.088)	0.290 (0.025)		
SE03_05	SE62111	0.024	0.545 (0.033)	0.533 (0.042)		-0.040 (0.078)	0.040 (0.082)
SE03_06A	SE62116A	0.026	1.256 (0.076)	0.588 (0.033)			
SE03_06B	SE62116B	0.082	1.309 (0.088)	1.068 (0.041)			

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SE03_06C	SE62116C	0.040	1.059 (0.080)	1.308 (0.060)			
SE03_07	SE62262	0.040	1.052 (0.182)	1.124 (0.087)	0.270 (0.032)		
SE03_08	SE62035	0.035	1.064 (0.156)	1.038 (0.076)	0.200 (0.030)		
SE03_09	SE62144	0.045	0.878 (0.093)	-0.312 (0.133)	0.205 (0.056)		
SE03_10	SE62162	0.032	0.717 (0.056)	0.909 (0.062)			
SE03_11	SE62233	0.045	1.072 (0.178)	0.904 (0.095)	0.327 (0.036)		
SE03_13	SE62171	0.035	0.427 (0.109)	1.157 (0.285)	0.232 (0.072)		
SE04_01	SE72002	0.046	1.019 (0.105)	0.127 (0.093)	0.188 (0.042)		
* SE04_02	SE72403	—	0.615 (0.048)	0.088 (0.060)			
SE04_03	SE72021	0.046	1.289 (0.144)	0.372 (0.076)	0.262 (0.036)		
SE04_04	SE72082	0.068	0.695 (0.051)	0.091 (0.054)			
SE04_05	SE72066	0.071	1.048 (0.119)	0.458 (0.083)	0.196 (0.037)		
SE04_06	SE72063	0.039	0.731 (0.190)	1.811 (0.176)	0.194 (0.035)		
SE04_07	SE72102	0.078	0.760 (0.054)	0.370 (0.049)			
SE04_08A	SE72141A	0.040	1.010 (0.070)	0.985 (0.048)			
SE04_08B	SE72141B	0.095	0.712 (0.036)	0.907 (0.038)		-0.336 (0.066)	0.336 (0.076)
SE04_09	SE72921	0.073	1.209 (0.090)	1.315 (0.054)			
SE04_10	SE72234	0.063	1.325 (0.175)	1.159 (0.057)	0.151 (0.021)		
SE04_11	SE72251	0.055	1.099 (0.172)	1.081 (0.078)	0.238 (0.030)		
SE04_12	SE72284	0.060	0.587 (0.047)	-0.037 (0.065)			
SE04_13	SE72345	0.038	0.803 (0.045)	0.485 (0.034)		0.638 (0.053)	-0.638 (0.056)
SE04_14	SE72349	0.025	1.125 (0.116)	0.145 (0.087)	0.210 (0.041)		
SE04_15	SE72363	0.051	0.740 (0.103)	0.224 (0.159)	0.245 (0.057)		
SE05_01	SE52076	0.045	0.865 (0.132)	0.509 (0.135)	0.311 (0.049)		
SE05_02	SE52272	0.017	1.059 (0.065)	-0.079 (0.041)			
SE05_03A	SE52085A	0.044	1.058 (0.079)	1.280 (0.058)			
SE05_03B	SE52085B	0.008	0.974 (0.062)	-0.077 (0.044)			
SE05_04	SE52094	0.030	0.648 (0.053)	0.927 (0.068)			
SE05_05	SE52248	0.034	0.947 (0.256)	1.631 (0.142)	0.329 (0.032)		
SE05_06	SE52146	0.022	1.161 (0.070)	0.402 (0.035)			
SE05_07	SE52282	0.052	0.701 (0.123)	0.822 (0.148)	0.244 (0.051)		
SE05_08	SE52299	0.047	1.111 (0.149)	0.323 (0.111)	0.364 (0.045)		
SE05_09	SE52144	0.037	1.301 (0.155)	0.680 (0.066)	0.234 (0.031)		
SE05_10	SE52214	0.028	1.033 (0.065)	0.375 (0.038)			
SE05_12	SE52101	0.031	0.538 (0.050)	1.010 (0.086)			
SE05_13	SE52113	0.062	1.515 (0.181)	0.706 (0.059)	0.265 (0.029)		
SE05_14	SE52107	0.045	0.769 (0.174)	1.482 (0.133)	0.233 (0.039)		
SE06_01	SE62090	0.047	1.084 (0.129)	0.265 (0.100)	0.272 (0.044)		
SE06_02	SE62274	0.031	0.575 (0.029)	0.794 (0.050)		1.266 (0.069)	-1.266 (0.089)

* eTIMSS items without a paperTIMSS counterpart

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SE06_03	SE62284	0.040	0.384 (0.081)	0.394 (0.381)	0.242 (0.084)		
SE06_04A	SE62098A	0.019	0.626 (0.036)	0.477 (0.037)		0.011 (0.069)	-0.011 (0.072)
SE06_04B	SE62098B	0.027	0.956 (0.056)	1.309 (0.042)		0.016 (0.050)	-0.016 (0.071)
SE06_05	SE62032	0.030	1.610 (0.261)	1.305 (0.061)	0.251 (0.021)		
SE06_06	SE62043	0.042	0.962 (0.069)	1.075 (0.054)			
SE06_07	SE62158	0.050	0.696 (0.134)	0.845 (0.165)	0.290 (0.054)		
SE06_08	SE62159	0.061	1.086 (0.132)	0.593 (0.082)	0.218 (0.036)		
SE06_09	SE62005	0.055	1.258 (0.078)	0.741 (0.035)			
SE06_10	SE62075	0.039	0.966 (0.157)	0.852 (0.106)	0.302 (0.040)		
SE06_11	SE62004	0.021	2.042 (0.220)	0.848 (0.040)	0.195 (0.020)		
SE06_12	SE62175	0.039	0.625 (0.051)	0.605 (0.060)			
SE06_13A	SE62173A	0.027	0.631 (0.050)	0.247 (0.057)			
SE06_13B	SE62173B	0.044	0.847 (0.279)	2.133 (0.254)	0.217 (0.029)		
SE07_01A	SE52090A	0.042	0.396 (0.080)	-0.015 (0.431)	0.280 (0.092)		
SE07_01B	SE52090B	0.023	0.644 (0.068)	1.975 (0.153)			
SE07_02	SE52262	0.031	0.647 (0.129)	1.028 (0.159)	0.242 (0.052)		
SE07_03	SE52267	0.023	0.959 (0.139)	0.772 (0.098)	0.249 (0.039)		
SE07_04	SE52273	0.027	0.578 (0.038)	0.956 (0.050)		0.160 (0.070)	-0.160 (0.087)
SE07_05Z	SE52015Z	0.041	0.833 (0.057)	-0.427 (0.059)			
SE07_06	SE52051	0.048	0.924 (0.062)	0.651 (0.043)			
SE07_07	SE52026	0.030	0.774 (0.134)	0.489 (0.175)	0.353 (0.056)		
SE07_08	SE52130	0.027	0.953 (0.159)	1.167 (0.090)	0.215 (0.033)		
SE07_09	SE52028	0.029	0.770 (0.124)	0.642 (0.142)	0.266 (0.051)		
SE07_10	SE52189	0.043	0.995 (0.064)	0.476 (0.039)			
SE07_11	SE52217	0.026	0.643 (0.122)	0.842 (0.172)	0.254 (0.056)		
SE07_12	SE52038	0.046	1.262 (0.202)	1.077 (0.075)	0.299 (0.029)		
SE07_13	SE52099	0.023	0.953 (0.066)	0.899 (0.048)			
SE07_14	SE52118	0.017	0.785 (0.063)	1.250 (0.073)			
SE08_01	SE72070	0.079	0.905 (0.147)	0.541 (0.138)	0.363 (0.048)		
SE08_02	SE72400	0.045	0.864 (0.057)	-0.180 (0.050)			
SE08_03	SE72024	0.041	1.207 (0.126)	0.107 (0.086)	0.239 (0.042)		
SE08_04	SE72462	0.045	0.596 (0.114)	0.911 (0.174)	0.220 (0.056)		
SE08_05	SE72443	0.121	0.965 (0.119)	0.304 (0.110)	0.251 (0.046)		
SE08_06	SE72903	0.046	0.662 (0.036)	0.875 (0.040)		-0.216 (0.067)	0.216 (0.078)
SE08_07	SE72145	0.023	1.035 (0.078)	1.313 (0.061)			
SE08_08	SE72100	0.051	0.814 (0.181)	1.070 (0.145)	0.376 (0.045)		
SE08_10	SE72137	0.059	0.976 (0.099)	0.158 (0.090)	0.160 (0.040)		
SE08_11	SE72298	0.034	0.889 (0.059)	0.440 (0.043)			
SE08_12	SE72215	0.051	0.450 (0.022)	0.922 (0.052)		-1.332 (0.116)	1.332 (0.127)
SE08_13	SE72260	0.029	0.656 (0.050)	0.335 (0.055)			
SE08_14	SE72265	0.032	0.684 (0.051)	0.167 (0.054)			

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SE08_15	SE72347	0.046	0.862 (0.166)	1.337 (0.106)	0.216 (0.035)		
SE08_16	SE72351	0.048	0.738 (0.060)	1.146 (0.071)			
SE08_17	SE72367	0.050	0.998 (0.131)	0.821 (0.080)	0.183 (0.034)		
SE09_01	SE62099	0.023	0.947 (0.108)	0.342 (0.098)	0.192 (0.042)		
SE09_02	SE62095	0.033	0.480 (0.031)	0.790 (0.052)		-0.064 (0.085)	0.064 (0.098)
SE09_03	SE62106	0.025	0.785 (0.085)	-0.450 (0.151)	0.185 (0.059)		
SE09_04	SE62064	0.027	0.829 (0.056)	-0.331 (0.055)			
SE09_05	SE62132	0.031	0.886 (0.109)	0.263 (0.120)	0.229 (0.049)		
SE09_06	SE62163	0.019	1.139 (0.085)	1.356 (0.059)			
SE09_07	SE62153	0.032	1.424 (0.209)	0.958 (0.068)	0.316 (0.028)		
SE09_08	SE62018	0.021	0.573 (0.034)	1.448 (0.066)		-0.568 (0.086)	0.568 (0.112)
SE09_09	SE62143	0.061	0.910 (0.093)	2.025 (0.132)			
SE09_10	SE62276	0.024	0.776 (0.059)	0.960 (0.060)			
SE09_11	SE62050	0.047	0.905 (0.069)	1.242 (0.064)			
SE09_12	SE62205	0.035	1.252 (0.143)	0.884 (0.057)	0.144 (0.025)		
SE09_13	SE62190	0.032	0.921 (0.095)	0.146 (0.096)	0.155 (0.042)		
SE09_14A	SE62024A	0.025	0.659 (0.118)	0.882 (0.152)	0.219 (0.052)		
SE09_14B	SE62024B	0.047	0.692 (0.068)	1.772 (0.125)			
SE10_01	SE72033	0.043	0.786 (0.034)	0.303 (0.030)		-0.463 (0.066)	0.463 (0.065)
SE10_02	SE72440	0.046	0.647 (0.049)	-0.231 (0.064)			
SE10_03	SE72032	0.032	1.284 (0.192)	0.950 (0.073)	0.299 (0.029)		
SE10_04	SE72031	0.056	0.857 (0.134)	1.131 (0.092)	0.164 (0.033)		
SE10_05	SE72086	0.114	0.622 (0.049)	-0.700 (0.085)			
SE10_06	SE72005	0.024	0.917 (0.050)	0.707 (0.031)		0.230 (0.047)	-0.230 (0.055)
SE10_08	SE72123	0.069	0.724 (0.114)	0.411 (0.169)	0.276 (0.058)		
SE10_09	SE72116	0.029	0.563 (0.121)	1.151 (0.179)	0.213 (0.055)		
SE10_10	SE72920	0.062	0.534 (0.032)	0.967 (0.054)		1.002 (0.071)	-1.002 (0.097)
SE10_11	SE72294	0.099	0.725 (0.053)	0.526 (0.052)			
SE10_12	SE72231	0.075	0.868 (0.146)	1.140 (0.098)	0.203 (0.035)		
SE10_13	SE72261	0.063	0.856 (0.057)	-0.204 (0.050)			
SE10_14	SE72220	0.060	1.429 (0.231)	1.435 (0.069)	0.169 (0.019)		
SE10_15	SE72348	0.065	0.589 (0.049)	-0.876 (0.098)			
SE10_16	SE72720	0.032	0.426 (0.126)	2.036 (0.290)	0.166 (0.055)		
SE11_01	SE62279	0.036	1.172 (0.125)	0.061 (0.091)	0.248 (0.044)		
SE11_02	SE62112	0.022	0.512 (0.046)	0.294 (0.068)			
SE11_03	SE62119	0.032	1.042 (0.117)	0.131 (0.102)	0.242 (0.046)		
SE11_04	SE62093	0.038	0.554 (0.034)	-0.010 (0.044)		0.206 (0.082)	-0.206 (0.074)
SE11_05	SE62089	0.033	1.627 (0.189)	0.962 (0.047)	0.171 (0.021)		
SE11_06	SE62006	0.021	1.069 (0.066)	0.405 (0.037)			
SE11_07	SE62067	0.040	0.755 (0.055)	0.499 (0.050)			
SE11_08	SE62247	0.030	0.881 (0.177)	1.236 (0.112)	0.273 (0.037)		

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SE11_09	SE62177	0.036	0.859 (0.134)	0.978 (0.098)	0.198 (0.037)		
SE11_10	SE62186	0.031	1.462 (0.237)	1.262 (0.065)	0.243 (0.022)		
SE11_11A	SE62211A	0.022	0.789 (0.055)	0.402 (0.047)			
SE11_11B	SE62211B	0.014	0.938 (0.098)	2.053 (0.138)			
SE11_13	SE62033	0.024	1.263 (0.078)	0.724 (0.036)			
SE11_14	SE62037	0.041	0.745 (0.142)	0.798 (0.158)	0.315 (0.052)		
SE11_15	SE62242	0.025	0.864 (0.063)	-1.050 (0.082)			
SE12_01	SE72078	0.025	0.943 (0.061)	0.385 (0.041)			
SE12_02	SE72460	0.049	0.861 (0.140)	0.708 (0.128)	0.298 (0.046)		
SE12_03	SE72000	0.047	0.640 (0.031)	0.313 (0.035)		-0.347 (0.073)	0.347 (0.073)
SE12_05	SE72901	0.042	0.438 (0.105)	1.004 (0.290)	0.239 (0.074)		
SE12_06	SE72038	0.082	0.858 (0.124)	0.769 (0.105)	0.208 (0.041)		
SE12_07	SE72120	0.085	0.723 (0.122)	0.772 (0.144)	0.239 (0.051)		
SE12_08	SE72143	0.149	0.894 (0.062)	0.783 (0.048)			
SE12_09	SE72523	0.029	0.598 (0.037)	0.353 (0.040)		0.386 (0.070)	-0.386 (0.071)
SE12_10	SE72168	0.094	1.265 (0.137)	0.591 (0.063)	0.181 (0.030)		
SE12_11	SE72205	0.076	0.763 (0.114)	0.634 (0.131)	0.219 (0.049)		
SE12_12	SE72293	0.028	1.096 (0.071)	0.821 (0.041)			
SE12_13A	SE72280A	0.062	1.174 (0.076)	0.925 (0.041)			
SE12_13B	SE72280B	0.044	1.155 (0.122)	-0.309 (0.113)	0.276 (0.055)		
SE12_14	SE72370	0.049	1.288 (0.142)	0.234 (0.082)	0.269 (0.040)		
SE13_01A	SE62091A	0.051	1.112 (0.117)	-0.484 (0.126)	0.279 (0.061)		
SE13_01B	SE62091B	0.082	0.678 (0.082)	-0.500 (0.207)	0.229 (0.073)		
SE13_02	SE62100	0.056	0.985 (0.063)	0.526 (0.040)			
SE13_03	SE62097	0.037	0.934 (0.116)	0.396 (0.108)	0.226 (0.045)		
SE13_04	SE62101	0.048	0.562 (0.035)	0.033 (0.043)		0.305 (0.079)	-0.305 (0.072)
SE13_06	SE62128	0.041	0.929 (0.059)	0.117 (0.042)			
SE13_07	SE62047	0.048	0.466 (0.044)	0.368 (0.075)			
SE13_08	SE62042	0.061	0.551 (0.049)	0.861 (0.076)			
SE13_09	SE62250	0.029	0.601 (0.055)	1.366 (0.100)			
SE13_10	SE62246	0.033	1.214 (0.204)	1.163 (0.079)	0.292 (0.028)		
SE13_11	SE62056	0.033	1.263 (0.075)	0.505 (0.033)			
SE13_12	SE62235	0.033	0.780 (0.117)	0.853 (0.112)	0.186 (0.042)		
SE13_13	SE62180	0.028	1.182 (0.121)	0.264 (0.077)	0.192 (0.038)		
SE13_14	SE62022	0.044	0.584 (0.050)	0.771 (0.069)			
SE13_15	SE62243	0.050	0.647 (0.033)	-0.129 (0.039)		-0.227 (0.078)	0.227 (0.069)
SE14_01	SE72011	0.097	1.238 (0.127)	-0.090 (0.093)	0.253 (0.047)		
SE14_02	SE72905	0.090	0.482 (0.044)	-0.097 (0.078)			
SE14_03	SE72049	0.056	0.629 (0.108)	0.434 (0.207)	0.266 (0.065)		
SE14_04	SE72016	0.069	1.031 (0.051)	0.690 (0.027)		0.033 (0.045)	-0.033 (0.050)
SE14_05	SE72451	0.053	0.868 (0.057)	-0.137 (0.049)			

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SE14_06	SE72074	0.094	0.878 (0.060)	0.608 (0.045)			
SE14_07	SE72091	0.038	1.150 (0.160)	0.821 (0.079)	0.264 (0.033)		
SE14_08	SE72109	0.062	0.693 (0.052)	0.487 (0.053)			
SE14_09	SE72140	0.036	0.808 (0.141)	0.915 (0.124)	0.262 (0.045)		
SE14_10	SE72132	0.036	1.007 (0.083)	1.540 (0.077)			
SE14_11	SE72209	0.038	0.940 (0.118)	0.548 (0.097)	0.202 (0.041)		
SE14_12	SE72210	0.043	0.479 (0.035)	1.126 (0.065)		0.604 (0.079)	-0.604 (0.107)
SE14_13	SE72249	0.039	0.788 (0.116)	0.998 (0.098)	0.149 (0.036)		
SE14_14	SE72323	0.051	0.687 (0.115)	0.690 (0.157)	0.234 (0.054)		
SE14_15	SE72368	0.138	0.744 (0.088)	-0.236 (0.167)	0.214 (0.063)		
SE14_16	SE72303	0.078	0.786 (0.144)	1.364 (0.109)	0.165 (0.035)		

Appendix 12K: Mathematics Item Parameters from the Final eTIMSS 2019 Adjusted Model Calibration—Grade 4

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
ME01_01	ME51043	0.045	0.450 (0.035)	0.011 (0.063)			
* ME01_02	ME51040	0.033	1.162	0.051	0.422		
* ME01_03	ME51008	0.050	1.270	1.010			
* ME01_04A	ME51031A	0.032	1.449	0.178			
* ME01_04B	ME51031B	0.029	1.619	0.252			
* ME01_05	ME51508	0.039	1.256	0.190			
* ME01_06A	ME51216A	0.038	1.272	0.592	0.237		
ME01_06B	ME51216B	0.029	0.749 (0.090)	-0.099 (0.167)	0.278 (0.060)		
* ME01_07	ME51221	0.038	0.571	-0.907	0.168		
* ME01_08	ME51115	0.030	0.591	1.706	0.113		
* ME01_09A	ME51507A	0.040	0.704	-0.564			
* ME01_09B	ME51507B	0.024	1.101	0.862			
* ME02_01	ME71219	0.042	0.709	-1.072	0.032		
* ME02_02	ME71021	0.056	1.146	0.191	0.089		
ME02_03	ME71167	0.025	1.557 (0.084)	1.073 (0.032)			
* ME02_04	ME71041	0.026	1.375	-0.220	0.143		
* ME02_05	ME71162	0.050	0.479	1.545		-0.840	0.840
ME02_06	ME71078	0.040	0.456 (0.037)	-0.848 (0.096)			
* ME02_07	ME71090	0.021	1.102	0.277	0.164		
* ME02_08	ME71151	0.057	0.593	0.990		-1.236	1.236
ME02_09	ME71119	0.033	0.675 (0.043)	-0.847 (0.069)			
ME02_10A	ME71217A	0.041	0.946 (0.054)	-0.888 (0.054)			
ME02_11	ME71142	0.030	1.132 (0.058)	-0.385 (0.035)			
* ME02_12	ME71204	0.063	1.334	0.569			
* ME03_01	ME61026	0.041	0.904	-0.740	0.098		
* ME03_02	ME61273	0.025	0.779	0.335	0.138		
* ME03_03	ME61034	0.027	1.187	0.694			
* ME03_04	ME61040	0.030	1.504	0.683	0.174		
* ME03_05	ME61228	0.058	0.734	0.965		-0.255	0.255
* ME03_06	ME61166	0.071	1.106	-0.263			
* ME03_07	ME61171	0.018	1.310	-0.249	0.231		
ME03_08	ME61080	0.024	0.700 (0.043)	0.525 (0.046)			
* ME03_09	ME61222	0.055	0.853	0.576	0.323		
ME03_10	ME61076	0.029	0.454 (0.038)	-1.144 (0.113)			
ME03_11	ME61084	0.019	1.076 (0.057)	0.648 (0.034)			

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
* ME04_01	ME71013	0.036	1.155	-0.166	0.234		
* ME04_02	ME71026	0.056	1.118	0.255			
ME04_03	ME71036	0.044	0.949 (0.051)	-0.543 (0.041)			
* ME04_04	ME71040	0.031	1.391	0.432	0.103		
* ME04_05	ME71068	0.034	0.492	0.513	0.113		
* ME04_06A	ME71075A	0.026	1.256	0.360			
* ME04_06B	ME71075B	0.019	1.471	0.740			
* ME04_07	ME71080	0.063	1.595	0.731	0.303		
ME04_08	ME71211	0.026	0.640 (0.040)	0.153 (0.046)			
ME04_09	ME71178	0.022	0.862 (0.048)	0.582 (0.040)			
ME04_10B	ME71135B	0.025	0.807 (0.046)	-0.199 (0.041)			
ME04_11	ME71201	0.026	0.753 (0.056)	1.592 (0.090)			
* ME04_12	ME71175	0.054	0.801	0.008		0.560	-0.560
* ME05_01	ME51206	0.044	0.591	-0.793			
* ME05_02	ME51052	0.028	0.824	0.084	0.297		
* ME05_03	ME51049	0.021	1.341	0.131	0.143		
* ME05_04	ME51045	0.037	1.066	-0.015			
* ME05_05	ME51098	0.026	0.990	0.753	0.121		
* ME05_06	ME51030	0.034	0.945	1.187			
* ME05_07	ME51502	0.034	0.961	1.192	0.153		
* ME05_08	ME51224	0.032	0.938	0.080	0.301		
* ME05_09	ME51207	0.018	0.799	0.887	0.341		
* ME05_10	ME51427	0.030	1.053	0.752	0.136		
* ME05_11	ME51533	0.022	1.056	0.168			
ME05_12	ME51080	0.026	1.090 (0.056)	0.227 (0.031)			
ME06_01	ME61018	0.025	0.934 (0.048)	0.104 (0.034)			
* ME06_02	ME61274	0.048	0.665	-0.592	0.197		
* ME06_03	ME61248	0.048	0.828	0.439		0.401	-0.401
* ME06_04	ME61039	0.049	1.068	0.327			
ME06_05	ME61079	0.022	1.225 (0.065)	0.932 (0.036)			
* ME06_06	ME61179	0.023	1.141	0.070	0.157		
* ME06_07	ME61052	0.032	0.945	0.116	0.091		
* ME06_08	ME61207	0.034	1.429	0.376	0.113		
ME06_09	ME61236	0.041	0.783 (0.044)	0.435 (0.041)			
* ME06_10	ME61266	0.030	0.466	0.765		-0.844	0.844
* ME06_11	ME61106	0.023	0.974	-0.032	0.219		
* ME07_01	ME51401	0.028	0.784	0.540			
* ME07_02	ME51075	0.045	1.297	1.137	0.326		
* ME07_03	ME51402	0.040	0.917	0.471			

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
* ME07_04	ME51226	0.031	1.302	0.681	0.270		
* ME07_05	ME51131	0.030	0.731	0.062			
* ME07_06	ME51103	0.033	1.258	0.267	0.280		
* ME07_07	ME51217	0.028	1.153	0.670			
ME07_08	ME51079	0.023	0.843 (0.048)	0.712 (0.042)			
* ME07_09	ME51211	0.054	0.783	-0.105	0.274		
* ME07_10	ME51102	0.031	0.948	0.792	0.159		
* ME07_11	ME51009	0.050	0.777	0.061			
* ME07_12	ME51100	0.044	0.642	0.217	0.195		
* ME08_01	ME71018	0.019	1.371	0.271	0.160		
* ME08_02	ME71009	0.066	1.248	0.303			
ME08_03	ME71037	0.047	0.751 (0.043)	0.019 (0.040)			
* ME08_04	ME71051	0.020	1.170	1.006			
* ME08_05	ME71064	0.051	0.724	0.850	0.155		
ME08_06	ME71176	0.044	0.728 (0.069)	-1.007 (0.188)	0.221 (0.075)		
* ME08_07	ME71169	0.021	1.317	0.600			
* ME08_08	ME71083	0.060	1.202	0.600	0.209		
* ME08_10	ME71184	0.037	1.635	1.153	0.244		
ME08_11	ME71141	0.031	0.877 (0.047)	0.317 (0.036)			
ME08_12	ME71194	0.046	0.743 (0.047)	-1.122 (0.073)			
ME08_13	ME71193	0.024	0.684 (0.024)	0.427 (0.027)		-0.773 (0.065)	0.773 (0.067)
ME08_14	ME71192	0.020	0.599 (0.023)	1.103 (0.040)		-1.327 (0.090)	1.327 (0.101)
* ME09_01	ME61275	0.043	0.709	-0.476	0.212		
* ME09_02	ME61027	0.048	0.893	-0.484			
* ME09_03	ME61255	0.029	0.812	0.576		-0.182	0.182
ME09_04	ME61021	0.025	0.827 (0.054)	1.414 (0.069)			
* ME09_05	ME61043	0.032	1.232	0.394			
* ME09_06	ME61151	0.027	1.203	-0.065	0.132		
* ME09_07	ME61172	0.033	1.520	0.849	0.123		
* ME09_08	ME61223	0.064	0.725	-0.633	0.119		
* ME09_09	ME61269	0.039	0.851	-0.370	0.130		
ME09_10A	ME61081A	0.028	1.013 (0.056)	0.873 (0.040)			
ME09_10B	ME61081B	0.027	0.978 (0.059)	1.210 (0.052)			
* ME10_02	ME71016	0.046	0.949	0.044			
* ME10_03	ME71163	0.040	1.762	1.060	0.076		
* ME10_04	ME71045	0.024	1.087	0.351	0.163		
ME10_05	ME71213	0.020	0.838 (0.047)	0.465 (0.038)			
* ME10_06	ME71070	0.023	0.354	-0.516	0.021		
ME10_07	ME71181	0.026	0.892 (0.049)	0.548 (0.037)			

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
ME10_08	ME71179	0.021	1.080 (0.063)	1.095 (0.045)			
* ME10_09	ME71067	0.035	0.543	1.054		-1.542	1.542
* ME10_10A	ME71147A	0.054	1.302	-0.335			
* ME10_10B	ME71147B	0.091	0.886	0.392			
ME10_11	ME71189	0.050	0.910 (0.056)	-1.247 (0.069)			
ME10_12A	ME71187A	0.035	0.724 (0.044)	-0.645 (0.057)			
* ME10_12B	ME71187B	0.069	0.676	-0.261			
* ME11_01	ME61178	0.064	0.829	0.142			
* ME11_02	ME61246	0.029	0.953	0.145	0.090		
* ME11_03	ME61271	0.035	0.618	-0.626			
* ME11_04	ME61256	0.046	0.835	0.218			
* ME11_05	ME61182	0.025	1.210	1.173			
* ME11_06	ME61049	0.040	0.910	-0.389	0.310		
* ME11_07	ME61232	0.019	0.970	0.753	0.321		
ME11_08	ME61095	0.025	0.904 (0.048)	0.009 (0.035)			
ME11_09	ME61264	0.029	0.560 (0.026)	0.465 (0.034)		-0.183 (0.063)	0.183 (0.069)
* ME11_10	ME61108	0.035	0.520	0.647	0.182		
* ME11_11A	ME61211A	0.038	1.222	0.241			
* ME11_11B	ME61211B	0.024	1.512	0.719	0.276		
* ME12_01	ME71001	0.045	0.857	-0.986	0.087		
* ME12_02	ME71010	0.068	0.694	-0.093			
* ME12_03	ME71062	0.045	1.337	1.262	0.129		
* ME12_04A	ME71216A	0.057	1.253	-0.288			
* ME12_04B	ME71216B	0.044	0.831	0.388			
ME12_05	ME71117	0.027	0.676 (0.040)	-0.064 (0.045)			
* ME12_06	ME71071	0.031	1.248	0.610	0.332		
* ME12_07	ME71098	0.035	0.729	0.855		0.060	-0.060
ME12_08	ME71069	0.034	1.088 (0.056)	0.568 (0.033)			
ME12_09A	ME71134A	0.024	1.785 (0.124)	0.215 (0.033)	0.114 (0.019)		
ME12_09B	ME71134B	0.025	1.483 (0.072)	0.535 (0.026)			
ME12_10	ME71202	0.030	0.562 (0.038)	-0.521 (0.064)			
ME12_11	ME71190	0.031	1.009 (0.051)	-0.142 (0.034)			
* ME12_12	ME71218	0.036	1.098	1.289			
ME13_01	ME61240	0.026	0.732 (0.044)	0.739 (0.049)			
ME13_02	ME61254	0.034	0.935 (0.048)	0.176 (0.033)			
* ME13_03	ME61244	0.036	0.931	-0.068	0.220		
* ME13_04	ME61041	0.044	1.209	1.090	0.242		
* ME13_05	ME61173	0.059	0.706	-0.210			
* ME13_06	ME61252	0.026	1.157	0.684	0.113		

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Item			RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
*	ME13_07	ME61261	0.033	1.261	0.208			
	ME13_08	ME61224	0.024	0.974 (0.055)	0.890 (0.042)			
*	ME13_09	ME61077	0.080	0.830	-0.068	0.093		
*	ME13_10A	ME61069A	0.040	0.725	-0.698			
*	ME13_10B	ME61069B	0.032	0.732	-0.021			
*	ME14_01	ME71024	0.023	0.921	0.254			
*	ME14_02	ME71008	0.053	1.118	-0.105	0.128		
*	ME14_03	ME71165	0.025	1.277	0.294	0.190		
	ME14_04	ME71049	0.030	0.770 (0.043)	0.006 (0.039)			
*	ME14_05	ME71063	0.054	1.050	0.314			
*	ME14_06	ME71079	0.023	1.179	0.790	0.192		
*	ME14_07	ME71081	0.034	1.007	-0.012			
*	ME14_08	ME71094	0.045	1.007	0.741	0.280		
	ME14_09	ME71177	0.025	0.531 (0.038)	0.196 (0.054)			
*	ME14_10	ME71206	0.038	0.681	-0.526	0.125		
	ME14_11A	ME71138A	0.025	0.770 (0.044)	0.004 (0.040)			
*	ME14_11B	ME71138B	0.037	0.984	0.841			
*	ME14_12	ME71203	0.055	0.653	1.272	0.106		
*	ME14_13	ME71205	0.030	1.108	0.460			

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Appendix 12L: Science Item Parameters from the Final eTIMSS 2019 Adjusted Model Calibration—Grade 4

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
* SE01_01	SE51054	0.039	0.934	-0.360	0.261		
* SE01_02	SE51024	0.038	0.612	0.733			
* SE01_03A	SE51132A	0.016	0.881	1.313			
* SE01_03B	SE51132B	0.026	0.810	1.124			
* SE01_04	SE51040	0.032	0.453	0.665			
* SE01_05	SE51193	0.036	0.940	-0.067	0.274		
* SE01_06	SE51063	0.022	1.148	0.812	0.222		
* SE01_07	SE51012	0.042	0.989	0.327	0.253		
* SE01_08	SE51115	0.036	1.090	0.205			
* SE01_09	SE51180	0.033	0.880	0.116	0.360		
* SE01_10	SE51106	0.034	1.024	0.780	0.215		
* SE01_11	SE51148	0.047	1.049	0.102	0.241		
* SE02_01	SE71002	0.065	0.572	0.102			
* SE02_02	SE71402	0.025	1.119	-0.194	0.299		
SE02_03	SE71017	0.028	0.683 (0.044)	0.193 (0.043)			
* SE02_04	SE71077	0.054	1.100	0.285			
* SE02_05	SE71072	0.079	1.212	0.845	0.232		
* SE02_06	SE71054	0.081	0.941	0.272			
* SE02_07	SE71115	0.034	0.848	0.856	0.249		
* SE02_08	SE71140	0.055	0.703	-0.012	0.240		
* SE02_09	SE71128	0.023	0.852	0.075	0.330		
* SE02_10	SE71147	0.045	0.883	-0.165	0.241		
SE02_11A	SE71920A	0.033	0.551 (0.043)	0.899 (0.068)			
* SE02_11B	SE71920B	0.018	0.956	0.671			
SE02_12	SE71268	0.020	0.923 (0.117)	0.866 (0.073)	0.180 (0.031)		
* SE03_01	SE61141	0.029	1.235	0.577	0.300		
SE03_02	SE61023	0.030	0.759 (0.046)	-0.304 (0.048)			
* SE03_03	SE61054	0.040	0.479	0.702		1.489	-1.489
* SE03_04	SE61007	0.035	0.647	-0.150	0.163		
* SE03_05	SE61006	0.040	0.785	-0.591			
* SE03_06	SE61108	0.022	1.050	0.292	0.352		
* SE03_07	SE61109	0.032	0.583	0.769	0.235		
* SE03_08	SE61080	0.034	0.968	0.356	0.264		
* SE03_09	SE61088	0.026	0.672	1.476			
* SE03_10	SE61151	0.024	0.952	0.499			
* SE03_11	SE61150	0.041	0.624	0.467			
* SE03_12	SE61169	0.050	1.077	0.138	0.268		

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
* SE04_01	SE71013	0.048	0.852	-0.707	0.278		
SE04_02	SE71902	0.023	0.346 (0.042)	1.980 (0.215)			
* SE04_03	SE71076	0.069	0.860	-0.505	0.134		
* SE04_04	SE71041	0.090	0.778	1.036		0.021	-0.021
SE04_05	SE71046	0.032	0.933 (0.052)	0.318 (0.033)			
* SE04_06	SE71095	0.051	0.654	0.284			
* SE04_07	SE71129	0.029	0.855	-0.559	0.346		
* SE04_08	SE71102	0.062	0.751	0.727			
* SE04_09	SE71124	0.040	1.132	0.569	0.252		
* SE04_10	SE71112	0.052	0.743	-1.124	0.216		
* SE04_11	SE71265	0.071	0.708	0.687	0.341		
* SE04_12	SE71223	0.062	0.548	-1.514	0.298		
* SE05_01	SE51044	0.045	0.503	0.259			
* SE05_03	SE51003	0.083	0.711	-0.063	0.104		
* SE05_04	SE51168	0.080	0.704	-0.416			
* SE05_05	SE51010	0.044	0.766	0.135			
* SE05_06	SE51035	0.048	1.249	1.357	0.236		
* SE05_07	SE51059	0.024	0.584	0.163			
* SE05_08	SE51142	0.052	0.802	0.657	0.199		
* SE05_09A	SE51131A	0.037	1.014	-0.030	0.193		
* SE05_09B	SE51131B	0.036	0.988	0.635	0.197		
SE05_10	SE51151	0.041	0.874 (0.053)	-0.905 (0.062)			
* SE05_11	SE51157	0.031	0.739	1.058	0.190		
* SE06_01	SE61071	0.032	0.335	-1.313	0.197		
* SE06_02	SE61138	0.034	0.616	0.061			
* SE06_03A	SE61016A	0.038	0.926	0.424	0.216		
* SE06_03B	SE61016B	0.035	0.990	0.568			
* SE06_04	SE61011	0.056	0.733	-0.477			
SE06_06	SE61083	0.047	0.762 (0.048)	-0.967 (0.069)			
* SE06_07	SE61034	0.032	0.788	1.147			
* SE06_08	SE61044	0.056	0.740	0.610			
SE06_09A	SE61142A	0.032	0.679 (0.046)	0.477 (0.046)			
* SE06_09B	SE61142B	0.027	0.788	1.093			
* SE06_10A	SE61115A	0.030	1.468	0.405	0.264		
* SE06_10B	SE61115B	0.062	1.345	0.721	0.328		
* SE07_01	SE51161	0.031	0.488	1.066	0.217		
* SE07_02	SE51051	0.024	1.391	1.429	0.281		
SE07_03Z	SE51138Z	0.034	0.528 (0.040)	0.350 (0.055)			
SE07_04	SE51194	0.037	1.049 (0.059)	0.737 (0.035)			

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
* SE07_05	SE51029	0.037	0.518	1.279	0.202		
* SE07_06	SE51077	0.077	0.747	-0.108			
* SE07_07	SE51200	0.050	0.679	1.255			
* SE07_08	SE51075	0.039	0.670	-0.527			
* SE07_09	SE51065	0.050	0.870	-0.156	0.333		
* SE07_10	SE51191	0.045	1.342	0.637	0.205		
* SE07_11	SE51099	0.040	0.868	0.391	0.216		
* SE07_12	SE51175	0.060	0.978	1.026			
* SE08_02	SE71033	0.022	0.544	0.335	0.289		
SE08_03	SE71065	0.033	0.636 (0.042)	-0.218 (0.052)			
* SE08_04	SE71025	0.040	0.270	-0.270			
* SE08_05	SE71081	0.085	0.949	1.110	0.157		
* SE08_06	SE71056	0.021	0.635	0.911			
* SE08_07	SE71145	0.067	0.516	-0.231	0.181		
* SE08_08	SE71104	0.074	0.795	-0.791			
* SE08_09	SE71144	0.041	0.515	-0.028	0.081		
* SE08_10	SE71150	0.044	1.055	-0.343			
* SE08_11	SE71201	0.046	1.048	0.033	0.285		
* SE08_12	SE71237	0.039	1.086	0.272			
* SE08_13	SE71260	0.032	0.735	1.164	0.151		
* SE09_01	SE61135	0.057	0.758	-0.539	0.268		
* SE09_02	SE61069	0.030	0.400	-0.422			
* SE09_03	SE61134	0.044	0.651	0.240	0.126		
* SE09_04	SE61140	0.029	1.039	0.660	0.296		
* SE09_05	SE61019	0.069	0.887	1.002			
* SE09_06	SE61022	0.037	0.656	0.241	0.241		
* SE09_07	SE61036	0.029	0.951	0.962			
SE09_08	SE61160	0.041	0.909 (0.054)	-0.828 (0.058)			
* SE09_09	SE61159	0.076	0.826	-0.729			
* SE09_10	SE61091	0.035	0.452	1.229		-0.176	0.176
* SE09_11	SE61118	0.039	1.001	0.601	0.217		
* SE09_12	SE61097	0.036	0.798	0.576	0.275		
SE10_01	SE71009	0.044	0.594 (0.027)	-0.451 (0.040)		1.114 (0.075)	-1.114 (0.052)
* SE10_02	SE71093	0.059	0.727	-0.350			
* SE10_03	SE71069	0.039	0.946	1.199	0.295		
* SE10_04	SE71051	0.029	0.748	0.681			
* SE10_05	SE71039	0.044	0.766	0.209	0.147		
* SE10_06	SE71080	0.037	0.929	0.987	0.235		
* SE10_07	SE71137	0.062	0.705	-0.224			

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
* SE10_08	SE71103	0.049	0.815	0.333	0.259		
SE10_09	SE71106	0.030	0.720 (0.047)	0.568 (0.044)			
* SE10_10	SE71100	0.070	0.910	0.334	0.374		
SE10_11	SE71921	0.026	1.639 (0.279)	1.373 (0.064)	0.262 (0.018)		
* SE10_12	SE71220	0.023	0.998	0.791	0.232		
SE10_13	SE71254	0.031	0.688 (0.045)	0.317 (0.043)			
* SE11_01	SE61132	0.066	0.710	0.598	0.213		
* SE11_02	SE61120	0.025	0.884	0.392	0.197		
* SE11_03	SE61025	0.065	0.531	-0.307			
* SE11_04A	SE61133A	0.045	1.370	0.303	0.326		
* SE11_04B	SE61133B	0.043	1.701	0.851	0.114		
* SE11_05	SE61074	0.032	0.772	0.278			
* SE11_06	SE61093	0.054	0.761	0.002		0.937	-0.937
* SE11_07	SE61161	0.054	0.614	0.723			
* SE11_08A	SE61042A	0.031	1.366	0.865	0.239		
* SE11_08B	SE61042B	0.023	0.791	0.699	0.150		
* SE11_09A	SE61041A	0.031	0.871	0.175			
* SE11_09B	SE61041B	0.031	0.719	0.226			
* SE11_10	SE61155	0.042	0.735	-0.429	0.286		
SE12_01	SE71031	0.037	0.620 (0.042)	-0.025 (0.049)			
* SE12_02	SE71090	0.040	0.767	0.070			
* SE12_03	SE71048	0.048	1.433	1.250	0.220		
* SE12_04	SE71071	0.034	0.990	0.934			
* SE12_05	SE71011	0.053	1.209	-0.362	0.193		
* SE12_06	SE71142	0.055	0.826	0.552	0.323		
* SE12_07	SE71138	0.049	0.771	-0.560			
* SE12_08	SE71127	0.047	0.920	0.093	0.288		
* SE12_10	SE71500	0.036	0.792	0.392	0.140		
* SE12_11	SE71257	0.037	1.395	1.443	0.431		
* SE12_12	SE71222	0.056	0.906	0.290			
* SE12_13	SE71252	0.049	0.988	0.411	0.290		
* SE13_02	SE61014	0.039	0.495	0.484			
* SE13_03	SE61056	0.050	0.853	-0.679			
* SE13_04	SE61015	0.040	0.692	-0.336			
* SE13_05	SE61113	0.026	0.760	1.013			
* SE13_06	SE61107	0.046	1.001	0.700	0.180		
* SE13_07	SE61046	0.040	1.164	0.863	0.227		
* SE13_08	SE61047	0.032	0.751	-0.459	0.313		
* SE13_09	SE61048	0.053	1.300	0.568	0.221		

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
* SE13_10	SE61096	0.061	1.100	0.789	0.257		
SE13_11	SE61124	0.024	0.620 (0.051)	1.359 (0.091)			
SE13_12	SE61116	0.033	0.743 (0.048)	0.313 (0.041)			
SE14_01	SE71063	0.032	0.445 (0.039)	0.099 (0.063)			
* SE14_02	SE71900	0.037	1.029	0.037	0.373		
* SE14_04	SE71043	0.031	0.644	1.440			
* SE14_05	SE71005	0.045	1.021	-0.525			
* SE14_06	SE71118	0.034	1.130	0.886	0.188		
* SE14_07	SE71139	0.033	0.952	0.066	0.359		
SE14_08	SE71114	0.049	0.756 (0.047)	-0.534 (0.054)			
* SE14_09	SE71131	0.030	0.577	0.030			
* SE14_10	SE71152	0.027	1.235	0.538	0.300		
* SE14_11	SE71218	0.079	0.795	-0.567	0.309		
* SE14_12	SE71214	0.046	1.098	0.178	0.167		
* SE14_13	SE71213	0.022	1.005	1.009			

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Appendix 12M: Mathematics Item Parameters from the Final eTIMSS 2019 Adjusted Model Calibration—Grade 8

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
* ME01_01	ME52024	0.018	1.646	0.551	0.232		
* ME01_02A	ME52058A	0.033	1.281	-0.255			
* ME01_02B	ME52058B	0.013	1.504	0.992			
* ME01_03	ME52125	0.028	1.196	0.684	0.098		
ME01_04	ME52229	0.021	1.237 (0.071)	0.391 (0.034)			
* ME01_05	ME52063	0.037	1.320	0.672	0.196		
* ME01_06	ME52072	0.042	1.009	0.107	0.146		
* ME01_07A	ME52146A	0.046	0.859	0.292			
* ME01_07B	ME52146B	0.035	1.533	1.263			
* ME01_08	ME52092	0.030	1.244	1.624	0.151		
* ME01_09	ME52046	0.053	1.125	1.586	0.188		
* ME01_10	ME52083	0.033	1.501	0.991	0.169		
* ME01_11	ME52082	0.027	1.202	0.271	0.174		
* ME01_12	ME52161	0.037	1.187	-0.100	0.189		
* ME01_13A	ME52418A	0.022	1.908	0.758	0.147		
* ME01_13B	ME52418B	0.036	1.916	0.663	0.250		
ME02_01	ME72007	0.033	0.743 (0.037)	1.042 (0.040)		-0.151 (0.063)	0.151 (0.077)
* ME02_02	ME72025	0.044	1.492	0.739	0.195		
* ME02_03	ME72017	0.043	1.319	1.127			
* ME02_04	ME72190	0.060	0.740	0.072			
* ME02_05	ME72068	0.039	1.285	0.089	0.185		
* ME02_06	ME72076	0.036	0.859	0.660	0.092		
* ME02_07	ME72056	0.055	1.159	0.661			
* ME02_08	ME72098	0.076	1.597	0.923			
* ME02_09	ME72103	0.037	1.249	0.754	0.150		
* ME02_10	ME72121	0.031	1.309	-0.154			
ME02_11	ME72180	0.042	0.499 (0.042)	0.327 (0.071)			
* ME02_12	ME72198	0.042	1.233	0.720			
* ME02_13	ME72227	0.094	1.507	0.688			
ME02_14	ME72170	0.034	0.774 (0.051)	0.140 (0.050)			
* ME02_15	ME72209	0.025	1.057	1.470			
* ME03_01	ME62005	0.058	0.871	0.588	0.304		
* ME03_02	ME62139	0.031	0.986	0.693			
* ME03_03	ME62164	0.032	1.357	0.185	0.172		
ME03_04	ME62142	0.038	0.837 (0.054)	-0.176 (0.049)			
* ME03_05	ME62084	0.039	1.393	1.663	0.144		

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
* ME03_06	ME62351	0.035	0.804	1.515	0.207		
* ME03_07	ME62223	0.033	1.420	-0.054	0.188		
* ME03_08	ME62027	0.021	0.772	0.666			
* ME03_09	ME62174	0.028	1.403	0.972	0.319		
* ME03_10	ME62244	0.034	0.971	0.572			
* ME03_11	ME62261	0.027	1.889	1.570	0.132		
ME03_12	ME62300	0.036	0.721 (0.032)	0.442 (0.034)		-0.386 (0.069)	0.386 (0.072)
* ME03_13	ME62254	0.029	0.744	1.600			
* ME03_14A	ME62132A	0.047	1.185	-0.186			
* ME03_14B	ME62132B	0.032	1.049	0.887	0.263		
ME04_01	ME72178	0.018	1.184 (0.070)	0.793 (0.040)			
* ME04_02	ME72234	0.059	0.959	1.052	0.258		
ME04_03	ME72020	0.032	0.739 (0.035)	0.029 (0.033)		-0.192 (0.067)	0.192 (0.063)
* ME04_04	ME72027	0.076	1.225	0.321	0.154		
* ME04_05	ME72052	0.078	0.814	1.664			
* ME04_06	ME72067	0.034	1.318	0.106	0.218		
* ME04_07A	ME72083A	0.085	1.406	0.019			
* ME04_07B	ME72083B	0.054	0.776	0.579	0.076		
* ME04_08A	ME72108A	0.053	0.728	0.099			
ME04_08B	ME72108B	0.017	1.085 (0.067)	0.963 (0.046)			
ME04_09	ME72181	0.032	0.956 (0.062)	1.043 (0.052)			
* ME04_10	ME72126	0.043	0.679	1.010		-0.811	0.811
ME04_11	ME72164	0.026	0.639 (0.052)	1.498 (0.096)			
ME04_12A	ME72185A	0.022	1.376 (0.081)	0.799 (0.036)			
ME04_12B	ME72185B	0.018	1.251 (0.075)	0.726 (0.038)			
ME05_01	ME52413	0.027	1.276 (0.126)	0.346 (0.069)	0.189 (0.034)		
* ME05_02	ME52134	0.048	1.261	-0.161	0.130		
* ME05_03	ME52078	0.035	0.990	0.993	0.183		
* ME05_04	ME52034	0.029	1.216	0.659	0.279		
* ME05_05A	ME52174A	0.045	1.088	0.323			
* ME05_05B	ME52174B	0.017	1.118	1.130			
* ME05_06	ME52130	0.044	1.232	1.080	0.173		
* ME05_07	ME52073	0.039	1.385	0.583	0.174		
* ME05_08	ME52110	0.022	1.464	0.763			
* ME05_09	ME52105	0.030	1.172	1.538			
* ME05_10	ME52407	0.028	1.344	0.469	0.378		
ME05_11	ME52036	0.031	0.799 (0.054)	0.922 (0.056)			
* ME05_12	ME52502	0.035	1.165	-0.139			

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
* ME05_13	ME52117	0.033	0.625	2.205			
* ME05_14	ME52426	0.056	0.785	-0.687	0.142		
* ME06_01	ME62150	0.045	1.111	-0.193			
* ME06_02	ME62335	0.045	1.377	0.004	0.175		
* ME06_03	ME62219	0.034	2.050	0.961	0.218		
ME06_04	ME62002	0.030	0.447 (0.042)	0.846 (0.091)			
* ME06_05	ME62149	0.043	1.089	0.617	0.111		
* ME06_06	ME62241	0.024	1.708	0.743			
* ME06_08	ME62105	0.058	0.757	0.960		-1.718	1.718
* ME06_09	ME62040	0.039	0.769	1.057	0.224		
* ME06_10	ME62288	0.030	0.776	1.250		-0.880	0.880
* ME06_11	ME62173	0.027	1.119	0.922			
* ME06_12	ME62133	0.018	1.315	0.726	0.214		
* ME06_13A	ME62123A	0.028	1.562	0.464	0.306		
* ME06_13B	ME62123B	0.031	1.444	0.814	0.138		
* ME07_01	ME52079	0.037	0.966	0.534	0.271		
* ME07_02	ME52204	0.038	0.871	0.506	0.180		
ME07_03	ME52364	0.047	1.228 (0.072)	-0.235 (0.038)			
ME07_04	ME52215	0.026	0.911 (0.056)	0.126 (0.043)			
* ME07_05	ME52147	0.021	1.572	0.872	0.275		
* ME07_06	ME52067	0.043	1.063	0.176	0.263		
* ME07_07	ME52068	0.035	1.417	1.374	0.132		
* ME07_08	ME52087	0.031	1.622	1.249			
ME07_09	ME52048	0.033	0.757 (0.051)	0.669 (0.053)			
* ME07_10	ME52039	0.022	1.235	0.382			
* ME07_11	ME52208	0.031	2.264	1.221	0.081		
* ME07_12A	ME52419A	0.053	0.888	-0.264	0.050		
* ME07_12B	ME52419B	0.035	1.372	-0.562	0.104		
* ME07_13	ME52115	0.030	1.738	0.457	0.080		
* ME07_14	ME52421	0.050	0.824	0.751			
* ME08_01	ME72002	0.039	1.517	0.652			
* ME08_02	ME72188	0.044	1.280	0.880	0.138		
* ME08_03	ME72035	0.033	1.132	0.661			
ME08_04	ME72055	0.020	1.260 (0.077)	0.968 (0.040)			
* ME08_05	ME72222	0.093	0.603	0.761	0.098		
* ME08_06	ME72090	0.044	1.211	0.987	0.198		
* ME08_07	ME72233	0.025	1.075	0.802	0.367		
* ME08_08A	ME72106A	0.047	1.068	-0.188			
ME08_08B	ME72106B	0.020	1.537 (0.091)	0.887 (0.034)			

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
ME08_08C	ME72106C	0.014	1.770 (0.114)	1.160 (0.035)			
* ME08_09A	ME72128A	0.025	0.999	0.654			
* ME08_09B	ME72128B	0.082	0.892	1.144		0.042	-0.042
ME08_10	ME72119	0.022	1.038 (0.062)	0.342 (0.039)			
* ME08_11A	ME72153A	0.072	1.021	0.488			
* ME08_11B	ME72153B	0.021	1.548	1.340			
* ME08_12	ME72172	0.037	1.048	0.204	0.060		
ME09_01	ME62329	0.081	0.675 (0.075)	-0.964 (0.224)	0.231 (0.082)		
* ME09_02	ME62151	0.027	1.247	0.826			
* ME09_03	ME62346	0.040	1.185	0.756			
* ME09_04	ME62212	0.013	1.397	1.199	0.124		
* ME09_05	ME62056	0.036	1.244	1.237			
* ME09_06	ME62317	0.033	1.328	0.933			
* ME09_07	ME62350	0.024	1.389	1.648	0.129		
* ME09_08	ME62078	0.047	1.441	0.721			
* ME09_09	ME62284	0.047	0.676	0.522	0.290		
* ME09_10	ME62245	0.031	1.273	0.752	0.204		
ME09_11	ME62287	0.029	1.321 (0.097)	1.545 (0.057)			
* ME09_12A	ME62345A	0.058	0.589	0.557		0.267	-0.267
* ME09_13	ME62115	0.031	1.507	1.468	0.202		
ME10_01	ME72187	0.041	0.909 (0.056)	-0.001 (0.044)			
* ME10_02	ME72022	0.024	1.631	1.180	0.279		
* ME10_04	ME72045	0.038	1.307	0.571			
ME10_05	ME72049	0.051	0.794 (0.053)	-0.361 (0.055)			
ME10_06	ME72069	0.045	1.615 (0.089)	0.163 (0.029)			
* ME10_07	ME72074	0.041	1.162	1.036			
* ME10_08	ME72013	0.032	1.126	0.704	0.120		
* ME10_09	ME72095	0.038	1.416	0.623			
* ME10_10	ME72109	0.039	1.467	1.194			
* ME10_11	ME72125	0.098	2.017	0.930	0.107		
* ME10_12	ME72196	0.021	1.376	0.653			
* ME10_13	ME72237	0.061	0.963	0.065	0.194		
ME10_14	ME72232	0.050	0.593 (0.046)	-0.112 (0.065)			
* ME10_15	ME72206	0.025	1.330	1.399			
* ME11_01	ME62271	0.057	1.536	0.635	0.252		
* ME11_02	ME62152	0.030	1.197	0.458			
* ME11_03	ME62215	0.040	0.889	0.765		-0.188	0.188
* ME11_04	ME62143	0.033	1.655	0.914			
* ME11_05	ME62230	0.030	1.555	1.468	0.224		

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
* ME11_06	ME62095	0.029	1.586	0.660	0.219		
* ME11_07	ME62076	0.041	1.745	0.341	0.291		
* ME11_08	ME62030	0.056	0.536	0.168			
* ME11_09	ME62171	0.063	0.832	-0.035	0.128		
* ME11_10	ME62301	0.020	1.080	1.108			
* ME11_11	ME62194	0.064	1.025	-0.164	0.290		
* ME11_12	ME62344	0.038	0.874	1.202			
* ME11_13	ME62320	0.014	1.899	0.579	0.092		
ME11_14	ME62296	0.038	1.137 (0.067)	0.117 (0.037)			
* ME12_01	ME72001	0.024	1.523	0.721			
ME12_02	ME72019	0.030	1.459 (0.082)	0.529 (0.032)			
ME12_03	ME72189	0.023	1.340 (0.145)	0.510 (0.069)	0.234 (0.031)		
* ME12_04	ME72024	0.042	0.899	0.726			
* ME12_05	ME72043	0.085	2.286	0.869	0.171		
* ME12_06	ME72221	0.047	1.207	0.440	0.219		
* ME12_07	ME72220	0.066	1.330	1.263	0.202		
* ME12_08	ME72225	0.032	1.263	0.559			
* ME12_09A	ME72110A	0.040	1.493	0.696			
ME12_09B	ME72110B	0.016	1.794 (0.115)	1.155 (0.035)			
* ME12_10	ME72150	0.033	1.827	0.523	0.481		
* ME12_11	ME72139	0.021	1.155	1.104			
* ME12_12	ME72229	0.026	0.966	1.543		-1.025	1.025
* ME12_13	ME72171	0.056	1.437	0.515			
ME12_14A	ME72211A	0.029	1.607 (0.172)	0.321 (0.065)	0.273 (0.032)		
* ME13_01	ME62001	0.071	1.007	0.956	0.339		
* ME13_02	ME62214	0.026	1.151	0.499			
* ME13_03	ME62146	0.015	1.444	0.815	0.124		
* ME13_04	ME62154	0.031	1.359	0.024			
* ME13_05	ME62067	0.038	1.159	0.206	0.335		
* ME13_06	ME62341	0.039	0.932	1.753	0.218		
* ME13_07	ME62242	0.036	1.269	0.285	0.171		
* ME13_08A	ME62250A	0.029	1.207	0.248			
* ME13_08B	ME62250B	0.016	1.403	0.927			
* ME13_09	ME62170	0.085	0.535	1.031		0.551	-0.551
* ME13_10	ME62192	0.039	1.044	1.230			
* ME13_11	ME62072	0.059	1.024	0.220			
* ME13_13	ME62120	0.037	1.250	0.575	0.166		
* ME14_01	ME72005	0.038	0.704	0.234	0.100		
ME14_02	ME72021	0.022	1.220 (0.071)	0.485 (0.034)			

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
ME14_03	ME72026	0.045	0.800 (0.054)	0.640 (0.049)			
* ME14_04A	ME72041A	0.045	1.268	0.214			
* ME14_04B	ME72041B	0.052	1.471	0.474			
* ME14_05	ME72223	0.083	1.948	0.773	0.250		
* ME14_06	ME72094	0.085	1.172	0.077			
* ME14_07	ME72059	0.084	1.363	0.726			
* ME14_08	ME72080	0.050	1.587	0.984	0.118		
ME14_09	ME72081	0.033	1.119 (0.072)	1.124 (0.047)			
ME14_10	ME72140	0.029	0.784 (0.053)	0.170 (0.048)			
ME14_11	ME72120	0.018	1.453 (0.090)	1.091 (0.038)			
* ME14_12	ME72131	0.033	1.349	1.395			
* ME14_13	ME72147	0.051	1.697	1.282			
* ME14_14	ME72154	0.065	1.325	0.216	0.189		
* ME14_15	ME72192	0.049	1.009	0.554	0.209		
* ME14_16	ME72161	0.030	1.164	0.728			

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Appendix 12N: Science Item Parameters from the Final eTIMSS 2019 Adjusted Model Calibration—Grade 8

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
* SE01_01	SE52006	0.054	0.635	-0.030		0.620	-0.620
* SE01_02	SE52069	0.029	0.984	0.668	0.325		
* SE01_03	SE52012	0.055	0.947	0.410	0.163		
* SE01_04	SE52021	0.018	1.029	0.706			
SE01_05Z	SE52095Z	0.031	0.466 (0.044)	-0.044 (0.080)			
* SE01_07	SE52054	0.042	0.749	-0.312			
* SE01_08	SE52150	0.033	0.787	1.237	0.181		
* SE01_09A	SE52243A	0.046	0.624	0.441			
* SE01_09B	SE52243B	0.037	0.769	0.461			
* SE01_09C	SE52243C	0.052	0.671	1.093	0.200		
* SE01_10	SE52206	0.063	1.127	0.545	0.207		
* SE01_11A	SE52112A	0.036	0.672	0.026	0.221		
* SE01_11B	SE52112B	0.051	0.992	0.832			
* SE01_12	SE52294	0.029	1.085	-0.017	0.206		
* SE02_01	SE72072	0.035	0.824	0.585	0.216		
* SE02_02	SE72029	0.065	1.324	1.125	0.364		
* SE02_03	SE72902	0.040	1.017	0.213			
* SE02_04	SE72077	0.046	0.685	0.463	0.300		
* SE02_05A	SE72900A	0.042	0.959	0.951			
* SE02_05B	SE72900B	0.067	0.954	1.428			
* SE02_06	SE72103	0.068	0.500	-0.011			
SE02_07	SE72110	0.024	0.868 (0.061)	0.817 (0.049)			
SE02_08	SE72130	0.031	0.720 (0.056)	0.995 (0.064)			
* SE02_09	SE72148	0.074	0.679	1.226	0.132		
* SE02_10	SE72200	0.034	0.854	0.739	0.103		
SE02_11	SE72232	0.030	1.479 (0.084)	0.311 (0.029)			
* SE02_12	SE72275	0.038	1.016	-0.454	0.117		
* SE02_13	SE72244	0.032	0.950	0.565			
* SE02_14	SE72301	0.076	0.936	1.267	0.220		
SE02_15	SE72721	0.029	1.153 (0.113)	0.095 (0.083)	0.189 (0.040)		
* SE02_16	SE72335	0.036	0.859	0.620	0.199		
* SE03_01	SE62055	0.031	0.962	-0.020	0.438		
* SE03_02	SE62007	0.046	1.176	0.525	0.205		
* SE03_03	SE62275	0.042	0.888	0.853			
* SE03_04	SE62225	0.025	1.004	1.402	0.259		
* SE03_05	SE62111	0.039	0.587	0.584		0.033	-0.033
* SE03_06A	SE62116A	0.036	1.164	0.597			
* SE03_06B	SE62116B	0.035	1.319	0.926			

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
* SE03_06C	SE62116C	0.023	0.946	1.315			
* SE03_07	SE62262	0.025	0.891	1.130	0.277		
* SE03_08	SE62035	0.031	1.076	1.097	0.199		
* SE03_09	SE62144	0.054	0.725	-0.533	0.163		
* SE03_10	SE62162	0.030	0.777	0.881			
* SE03_11	SE62233	0.035	0.927	0.820	0.343		
* SE03_13	SE62171	0.035	0.384	0.893	0.185		
* SE04_01	SE72002	0.060	1.393	0.307	0.212		
SE04_02	SE72403	0.033	0.618 (0.048)	0.086 (0.060)			
* SE04_03	SE72021	0.058	0.896	0.404	0.221		
SE04_04	SE72082	0.057	0.704 (0.051)	0.089 (0.053)			
SE04_05	SE72066	0.028	1.053 (0.119)	0.446 (0.083)	0.194 (0.037)		
* SE04_06	SE72063	0.035	0.582	2.063	0.200		
* SE04_07	SE72102	0.087	0.482	0.612			
* SE04_08A	SE72141A	0.026	1.069	0.944			
* SE04_08B	SE72141B	0.065	0.731	0.669		-0.141	0.141
* SE04_09	SE72921	0.053	0.766	1.439			
* SE04_10	SE72234	0.085	1.141	1.540	0.167		
* SE04_11	SE72251	0.024	1.064	0.922	0.208		
* SE04_12	SE72284	0.050	0.786	0.009			
SE04_13	SE72345	0.035	0.823 (0.045)	0.478 (0.033)		0.632 (0.052)	-0.632 (0.055)
* SE04_14	SE72349	0.042	1.086	0.150	0.178		
* SE04_15	SE72363	0.076	0.613	0.140	0.101		
* SE05_01	SE52076	0.035	0.934	0.411	0.257		
* SE05_02	SE52272	0.050	1.130	-0.007			
* SE05_03A	SE52085A	0.016	1.038	1.232			
* SE05_03B	SE52085B	0.054	1.034	0.009			
* SE05_04	SE52094	0.036	0.614	1.030			
* SE05_05	SE52248	0.022	1.188	1.615	0.364		
* SE05_06	SE52146	0.040	1.023	0.411			
* SE05_07	SE52282	0.071	0.828	0.857	0.185		
* SE05_08	SE52299	0.063	1.224	0.392	0.309		
* SE05_09	SE52144	0.032	1.160	0.710	0.249		
* SE05_10	SE52214	0.028	0.996	0.356			
* SE05_12	SE52101	0.033	0.563	1.043			
* SE05_13	SE52113	0.042	1.565	0.597	0.292		
* SE05_14	SE52107	0.044	1.000	1.328	0.197		
* SE06_01	SE62090	0.043	1.011	0.180	0.304		
* SE06_02	SE62274	0.059	0.577	0.879		1.149	-1.149

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
* SE06_03	SE62284	0.061	0.375	0.478	0.172		
* SE06_04A	SE62098A	0.040	0.639	0.500		-0.050	0.050
* SE06_04B	SE62098B	0.030	0.798	1.337		-0.091	0.091
* SE06_05	SE62032	0.052	1.742	1.504	0.287		
* SE06_06	SE62043	0.028	0.907	0.981			
* SE06_07	SE62158	0.031	0.697	0.678	0.299		
* SE06_08	SE62159	0.036	0.983	0.400	0.204		
* SE06_09	SE62005	0.020	1.250	0.666			
* SE06_10	SE62075	0.025	0.990	0.770	0.314		
* SE06_11	SE62004	0.049	1.806	0.885	0.173		
* SE06_12	SE62175	0.059	0.739	0.674			
SE06_13A	SE62173A	0.036	0.647 (0.051)	0.253 (0.056)			
* SE06_13B	SE62173B	0.026	0.808	1.862	0.203		
* SE07_01A	SE52090A	0.026	0.494	0.539	0.393		
* SE07_01B	SE52090B	0.027	0.609	1.962			
* SE07_02	SE52262	0.020	0.694	0.910	0.227		
* SE07_03	SE52267	0.034	0.988	0.763	0.216		
* SE07_04	SE52273	0.030	0.638	0.934		0.174	-0.174
SE07_05Z	SE52015Z	0.039	0.847 (0.057)	-0.399 (0.057)			
* SE07_06	SE52051	0.078	1.005	0.815			
* SE07_07	SE52026	0.041	0.587	0.468	0.350		
* SE07_08	SE52130	0.026	0.909	1.202	0.215		
* SE07_09	SE52028	0.027	0.858	0.620	0.282		
* SE07_10	SE52189	0.033	1.041	0.450			
* SE07_11	SE52217	0.041	0.722	1.059	0.283		
* SE07_12	SE52038	0.033	0.994	0.977	0.290		
* SE07_13	SE52099	0.026	0.947	0.884			
* SE07_14	SE52118	0.031	0.766	1.293			
* SE08_01	SE72070	0.063	0.568	-0.193	0.207		
SE08_02	SE72400	0.035	0.878 (0.058)	-0.151 (0.049)			
* SE08_03	SE72024	0.048	0.891	-0.027	0.113		
* SE08_04	SE72462	0.036	0.490	0.792	0.198		
SE08_05	SE72443	0.026	0.969 (0.121)	0.320 (0.111)	0.249 (0.047)		
* SE08_06	SE72903	0.023	0.796	0.821		-0.090	0.090
* SE08_07	SE72145	0.041	0.949	1.441			
* SE08_08	SE72100	0.047	0.560	0.647	0.195		
* SE08_10	SE72137	0.082	0.836	0.435	0.194		
* SE08_11	SE72298	0.069	0.814	0.626			
* SE08_12	SE72215	0.057	0.515	1.031		-0.538	0.538

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SE08_13	SE72260	0.031	0.667 (0.051)	0.356 (0.054)			
SE08_14	SE72265	0.028	0.692 (0.051)	0.191 (0.053)			
* SE08_15	SE72347	0.026	1.061	1.184	0.186		
* SE08_16	SE72351	0.025	0.847	0.997			
* SE08_17	SE72367	0.020	1.114	0.705	0.156		
* SE09_01	SE62099	0.043	0.842	0.324	0.146		
* SE09_02	SE62095	0.028	0.501	0.750		-0.076	0.076
* SE09_03	SE62106	0.041	0.750	-0.654	0.116		
* SE09_04	SE62064	0.032	0.879	-0.289			
* SE09_05	SE62132	0.035	0.992	0.400	0.289		
* SE09_06	SE62163	0.035	1.196	1.375			
* SE09_07	SE62153	0.027	1.278	0.921	0.294		
SE09_08	SE62018	0.029	0.567 (0.033)	1.446 (0.067)		-0.582 (0.087)	0.582 (0.113)
* SE09_09	SE62143	0.047	0.850	1.772			
* SE09_10	SE62276	0.042	0.718	1.062			
* SE09_11	SE62050	0.032	0.920	1.074			
* SE09_12	SE62205	0.033	1.100	0.892	0.158		
* SE09_13	SE62190	0.034	0.883	0.091	0.140		
* SE09_14A	SE62024A	0.035	0.605	0.944	0.226		
* SE09_14B	SE62024B	0.025	0.801	1.514			
SE10_01	SE72033	0.029	0.789 (0.034)	0.298 (0.029)		-0.465 (0.065)	0.465 (0.065)
* SE10_02	SE72440	0.037	0.670	-0.280			
* SE10_03	SE72032	0.046	1.540	1.069	0.315		
* SE10_04	SE72031	0.037	0.655	1.009	0.137		
SE10_05	SE72086	0.028	0.637 (0.050)	-0.680 (0.082)			
* SE10_06	SE72005	0.057	1.030	0.797		0.248	-0.248
* SE10_08	SE72123	0.052	0.551	0.064	0.249		
* SE10_09	SE72116	0.042	0.574	1.240	0.198		
SE10_10	SE72920	0.083	0.544 (0.033)	0.954 (0.053)		0.985 (0.069)	-0.985 (0.095)
* SE10_11	SE72294	0.056	0.914	0.274			
* SE10_12	SE72231	0.043	1.257	0.990	0.265		
SE10_13	SE72261	0.031	0.868 (0.058)	-0.200 (0.049)			
* SE10_14	SE72220	0.081	1.761	1.800	0.210		
* SE10_15	SE72348	0.037	0.805	-0.777			
* SE10_16	SE72720	0.023	0.412	1.812	0.135		
* SE11_01	SE62279	0.042	1.185	0.075	0.187		
* SE11_02	SE62112	0.027	0.534	0.284			
* SE11_03	SE62119	0.033	1.214	0.226	0.249		
* SE11_04	SE62093	0.046	0.630	0.131		0.306	-0.306

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
* SE11_05	SE62089	0.039	1.347	1.002	0.153		
* SE11_06	SE62006	0.041	0.953	0.430			
* SE11_07	SE62067	0.029	0.823	0.433			
* SE11_08	SE62247	0.035	0.977	1.300	0.268		
* SE11_09	SE62177	0.037	0.711	1.076	0.207		
* SE11_10	SE62186	0.026	1.545	1.245	0.263		
* SE11_11A	SE62211A	0.024	0.814	0.413			
* SE11_11B	SE62211B	0.019	0.868	2.149			
* SE11_13	SE62033	0.039	1.106	0.762			
* SE11_14	SE62037	0.034	0.747	0.631	0.305		
SE11_15	SE62242	0.049	0.885 (0.064)	-1.001 (0.079)			
* SE12_01	SE72078	0.061	1.019	0.526			
* SE12_02	SE72460	0.058	0.962	0.778	0.254		
SE12_03	SE72000	0.029	0.639 (0.031)	0.329 (0.035)		-0.348 (0.074)	0.348 (0.073)
* SE12_05	SE72901	0.041	0.612	1.189	0.273		
* SE12_06	SE72038	0.057	0.487	0.364	0.103		
* SE12_07	SE72120	0.066	0.441	0.022	0.092		
SE12_08	SE72143	0.023	0.892 (0.062)	0.800 (0.048)			
* SE12_09	SE72523	0.042	0.663	0.387		0.309	-0.309
* SE12_10	SE72168	0.057	1.195	0.387	0.176		
* SE12_11	SE72205	0.088	1.159	0.948	0.244		
* SE12_12	SE72293	0.045	0.959	0.926			
* SE12_13A	SE72280A	0.025	1.309	0.823			
* SE12_13B	SE72280B	0.045	1.433	0.005	0.387		
* SE12_14	SE72370	0.072	1.461	0.487	0.289		
* SE13_01A	SE62091A	0.043	0.958	-0.639	0.304		
* SE13_01B	SE62091B	0.071	0.587	-1.118	0.167		
* SE13_02	SE62100	0.035	0.898	0.403			
* SE13_03	SE62097	0.044	0.909	0.334	0.147		
SE13_04	SE62101	0.025	0.549 (0.034)	0.014 (0.044)		0.311 (0.081)	-0.311 (0.074)
* SE13_06	SE62128	0.027	0.867	0.043			
SE13_07	SE62047	0.052	0.457 (0.043)	0.356 (0.076)			
SE13_08	SE62042	0.040	0.539 (0.048)	0.859 (0.078)			
* SE13_09	SE62250	0.033	0.580	1.268			
* SE13_10	SE62246	0.033	0.924	1.256	0.288		
* SE13_11	SE62056	0.031	1.147	0.495			
* SE13_12	SE62235	0.030	0.765	0.922	0.195		
* SE13_13	SE62180	0.035	1.210	0.326	0.211		

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

Item		RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (d_{i1})	Step 2 (d_{i2})
SE13_14	SE62022	0.030	0.577 (0.049)	0.764 (0.070)			
SE13_15	SE62243	0.042	0.631 (0.032)	-0.153 (0.040)		-0.233 (0.080)	0.233 (0.071)
SE14_01	SE72011	0.025	1.239 (0.128)	-0.106 (0.093)	0.253 (0.046)		
SE14_02	SE72905	0.042	0.481 (0.044)	-0.112 (0.078)			
* SE14_03	SE72049	0.067	0.805	0.684	0.270		
* SE14_04	SE72016	0.046	0.782	0.627		-0.167	0.167
* SE14_05	SE72451	0.041	1.084	-0.094			
* SE14_06	SE72074	0.054	0.785	0.412			
* SE14_07	SE72091	0.040	1.170	0.830	0.233		
SE14_08	SE72109	0.032	0.685 (0.051)	0.475 (0.054)			
* SE14_09	SE72140	0.038	0.906	1.049	0.279		
* SE14_10	SE72132	0.045	0.853	1.761			
* SE14_11	SE72209	0.036	1.207	0.708	0.268		
SE14_12	SE72210	0.079	0.475 (0.035)	1.120 (0.066)		0.609 (0.079)	-0.609 (0.108)
* SE14_13	SE72249	0.045	1.008	0.997	0.143		
* SE14_14	SE72323	0.032	0.697	0.791	0.295		
SE14_15	SE72368	0.033	0.748 (0.089)	-0.248 (0.170)	0.215 (0.064)		
* SE14_16	SE72303	0.038	1.205	1.133	0.210		

* Invariant item—item parameters for invariant items were fixed from the paperTIMSS concurrent calibration; location parameters are transformations of the fixed paperTIMSS value.

CHAPTER 13

Examining eTIMSS Country Differences Between eTIMSS Data and Bridge Data

A Look at Country-Level Mode of Administration Effects

Matthias von Davier
Pierre Foy
Michael O. Martin
Ina V.S. Mullis

The TIMSS & PIRLS International Study Center made every effort to ensure a seamless transition from the TIMSS paper-and-pencil format to the new, computer based eTIMSS. A major priority in developing the TIMSS 2019 assessment was ensuring that the eTIMSS and paperTIMSS assessments measured the same mathematics and science constructs, using the same items as much as possible.

The TIMSS 2019 Bridge Between eTIMSS and paperTIMSS

The purpose of this chapter is to help each eTIMSS country understand how a comparison between its eTIMSS data and its bridge data can be used to study how the transition to eTIMSS may have affected its TIMSS 2019 achievement results.

Based on an item equivalence pilot in 2017 (Fishbein, Martin, Mullis, & Foy, 2018) that indicated a modest mode of administration effect, countries transitioning to eTIMSS included a bridge to paperTIMSS to control for this mode effect. To provide bridging data, eTIMSS countries administered the complete computer-based eTIMSS 2019 assessment as well as a smaller, paper-based version of the trend items. That is, eTIMSS countries re-administered their eight blocks of trend items from 2015 in paperTIMSS format. The bridge booklets were administered to an additional sample of 1,500 students, sampled from about one-third of the schools selected for the full eTIMSS sample. As a random sample from the same student population, the bridge sample taking the trend items in paperTIMSS format is randomly equivalent to the full eTIMSS sample. As such, the bridge data form an intermediate link between eTIMSS countries' computer-based data in 2019 and their paper-based data in 2015, as well as to the paperTIMSS countries in 2019.

Using the bridge data as a link between paperTIMSS and eTIMSS is an example of equivalent groups linking. Because students were randomly assigned to the new (eTIMSS) and old (paperTIMSS) formats of an assessment, the students taking the two formats could be expected to have the same underlying skills and knowledge. They differ only in that they were randomly assigned to different formats and are otherwise equivalent. Underlying this approach is the principle of randomization, one of the central principles of experimental design (Box, Hunter, & Hunter, 2005), which aims to ensure that observed differences in results of groups exposed to experimental treatments are due to the treatments themselves and not to pre-existing differences between the groups.

Item Equivalence Between eTIMSS and paperTIMSS

TIMSS 2019 paid particular attention in converting its paper trend items (items used in TIMSS 2015 and retained in the 2019 assessment) to the eTIMSS computer-based version to ensure as much comparability as possible between response modes. This resulted in a large number of trend items that were very similar in presentation and response format between the paper and the eTIMSS assessments. In developing new items, there was an effort to capitalize on the digital environment and produce more engaging item types.

Having a substantial percentage of equivalent items between paper and eTIMSS strengthens the validity and interpretability of achievement results based on linking the two modes and enhances the randomly equivalent groups design. The more similarity between the paper and computer-based items, the more achievement differences between them are likely to be due to a mode effect. Therefore, TIMSS 2019 devoted considerable effort to identifying items that were equivalent or invariant with respect to paper and eTIMSS format, in content and psychometric properties (see Chapter 12; von Davier et al., 2019a,b).

Exhibit 13.1 shows the counts of equivalent and non-equivalent items in eTIMSS 2019 at fourth grade and eighth grade for mathematics and science. The percentage of equivalent trend items ranged from 80 to 91 percent across fourth and eighth grades for mathematics and science. Moreover, high percentages of all the eTIMSS items were equivalent—ranging from 72 to 87 percent. The equivalent items come from the following three categories defined by response types: multiple choice, keyboard, or number pad. As could be anticipated, somewhat higher percentages of the trend items were equivalent compared to the new items.

Exhibit 13.1: eTIMSS 2019 Achievement Items by Mode of Administration Equivalence

eTIMSS 2019 Fourth Grade Item Equivalence

Item Type		Mathematics			Science		
		Trend	New	Total	Trend	New	Total
Equivalent Items	Multiple Choice Items	41	24	65	47	39	86
	Keyboard Items	3	3	6	39	22	61
	Number Pad Items	30	22	52	—	—	—
All Equivalent Items		**74**	**49**	**123**	**86**	**61**	**147**
All Non–Equivalent Items		**18**	**30**	**48**	**9**	**13**	**22**
All Items		**92**	**79**	**171**	**95**	**74**	**169**
Percentage of Equivalent Items		**80%**	**62%**	**72%**	**91%**	**82%**	**87%**

eTIMSS 2019 Eighth Grade Item Equivalence

Item Type		Mathematics			Science		
		Trend	New	Total	Trend	New	Total
Equivalent Items	Multiple Choice Items	60	26	86	58	44	102
	Keyboard Items	9	10	19	47	26	73
	Number Pad Items	33	29	62	2	1	3
All Equivalent Items		**102**	**65**	**167**	**107**	**71**	**178**
All Non–Equivalent Items		**12**	**27**	**39**	**10**	**23**	**33**
All Items		**114**	**92**	**206**	**117**	**94**	**211**
Percentage of Equivalent Items		**89%**	**71%**	**81%**	**91%**	**76%**	**84%**

Country-Level Differences in Average Percent Correct on TIMSS 2019 Trend Items by Mode of Administration

To help users of the TIMSS 2019 data gain an understanding of the effect of changing from paperTIMSS to eTIMSS, the analyses in this section compare average performance between the paper bridge and eTIMSS on the trend items. This approach provides a model for investigating country mode effects for different types of items or student groups and a useful avenue for beginning to explore a country's transition to eTIMSS. The computations are described in Appendix 13A and are relatively straightforward.

Although the approach could be applied to any group of items, the analyses below were restricted to the more than 80 percent of the trend items that were found to be invariant between the two modes (see Exhbibit 13.1). These items are virtually identical except for administration mode, and the two samples were designed to be randomly equivalent. Apart from sampling differences and deviations from the sampling design that have caused some departure from this equivalence of comparison groups, the performance differences between the paper bridge and eTIMSS can be attributed to a mode effect.

Exhibits 13.2 through13.5 show for each eTIMSS country average performance on the invariant trend items for the paper bridge and eTIMSS samples as well as the average across the countries. Exhibits 13.2 and 13.3 show the results for fourth grade mathematics and science. Exhibits 13.4 and 13.5 show the same for eighth grade mathematics and science.

Exhibit 13.2: eTIMSS 2019 Average Percent Correct on Paper Bridge and eTIMSS Invariant Items – Fourth Grade Mathematics

Country	Paper_INV	eTIMSS_INV
Austria	54.84 (0.94)	52.87 (0.56)
Canada	47.89 (1.05)	46.24 (0.49)
Chile	31.98 (0.97)	29.93 (0.52)
Chinese Taipei	71.94 (0.66)	69.89 (0.42)
Croatia	47.24 (0.90)	44.25 (0.53)
Czech Republic	50.96 (1.53)	52.01 (0.69)
Denmark	52.23 (0.96)	49.42 (0.58)
England	59.45 (1.31)	57.32 (0.82)
Finland	53.93 (0.87)	51.51 (0.55)
France	40.04 (0.90)	39.08 (0.74)
Georgia	48.13 (1.91)	39.84 (0.84)
Germany	50.97 (1.06)	47.75 (0.61)
Hong Kong SAR	72.90 (1.78)	69.45 (0.83)
Hungary	52.80 (1.31)	48.72 (0.69)
Italy	48.14 (1.31)	46.38 (0.72)
Korea, Rep. of	69.26 (0.65)	67.56 (0.57)
Lithuania	60.02 (0.70)	54.65 (0.80)
Netherlands	53.13 (1.09)	51.39 (0.62)
Norway (5)	55.52 (1.07)	54.00 (0.56)
Portugal	54.75 (1.18)	49.37 (0.75)
Qatar	36.32 (1.15)	33.93 (0.76)
Russian Federation	61.82 (1.01)	60.39 (0.85)
Singapore	76.91 (1.18)	74.98 (0.90)
Slovak Republic	47.68 (1.32)	45.41 (0.75)
Spain	45.18 (1.12)	42.36 (0.49)
Sweden	49.55 (1.65)	47.90 (0.80)
United Arab Emirates	46.10 (1.92)	41.09 (0.34)
United States	56.22 (1.18)	53.94 (0.70)
International Average	**53.42 (0.23)**	**50.77 (0.13)**

( ) Standard errors appear in parentheses.

Exhibit 13.3: eTIMSS 2019 Average Percent Correct on Paper Bridge and eTIMSS Invariant Items – Fourth Grade Science

Country	Paper_INV	eTIMSS_INV
Austria	48.60 (0.80)	48.11 (0.44)
Canada	48.71 (1.05)	48.50 (0.34)
Chile	39.07 (0.83)	38.47 (0.48)
Chinese Taipei	58.19 (0.62)	55.46 (0.34)
Croatia	50.83 (0.75)	49.46 (0.46)
Czech Republic	50.55 (1.60)	51.08 (0.45)
Denmark	48.85 (0.89)	47.86 (0.46
England	55.33 (1.07)	53.51 (0.52)
Finland	56.64 (0.81)	56.32 (0.49)
France	40.93 (0.80)	41.04 (0.61)
Georgia	42.82 (1.45)	35.40 (0.62)
Germany	51.51 (1.05)	48.87 (0.56)
Hong Kong SAR	55.19 (1.55)	51.47 (0.74)
Hungary	53.49 (1.21)	51.37 (0.58)
Italy	47.87 (0.79)	46.16 (0.60)
Korea, Rep. of	66.45 (0.56)	65.10 (0.48)
Lithuania	54.52 (0.68)	52.85 (0.55)
Netherlands	48.30 (0.99)	47.66 (0.60)
Norway (5)	54.53 (0.70)	52.62 (0.49)
Portugal	47.03 (0.76)	43.52 (0.43)
Qatar	42.23 (1.50)	37.06 (0.71)
Russian Federation	61.53 (0.99)	59.61 (0.78)
Singapore	69.00 (1.12)	67.32 (0.76)
Slovak Republic	49.02 (0.95)	48.08 (0.64)
Spain	49.04 (0.97)	46.74 (0.40)
Sweden	51.52 (1.34)	53.06 (0.74)
United Arab Emirates	46.06 (1.63)	41.82 (0.34)
United States	54.33 (1.14)	52.76 (0.57)
International Average	**51.51 (0.20)**	**49.69 (0.11)**

( ) Standard errors appear in parentheses.

Exhibit 13.4: eTIMSS 2019 Average Percent Correct on Paper Bridge and eTIMSS Invariant Items – Eighth Grade Mathematics

Country	Paper_INV	eTIMSS_INV
Chile	26.94 (0.72)	26.58 (0.45)
Chinese Taipei	70.20 (1.15)	65.04 (0.56)
England	48.02 (1.63)	42.11 (1.16)
Georgia	32.78 (1.21)	30.43 (0.84)
Hong Kong SAR	62.45 (1.68)	57.16 (0.95)
Hungary	47.42 (1.84)	43.53 (0.71)
Israel	46.23 (1.65)	44.86 (1.02)
Italy	39.46 (0.73)	37.16 (0.59)
Korea, Rep. of	69.71 (0.75)	65.02 (0.55)
Lithuania	44.23 (1.26)	42.99 (0.67)
Malaysia	37.21 (2.04)	32.00 (0.55)
Norway (9)	43.85 (0.98)	40.83 (0.56)
Qatar	33.69 (1.19)	28.55 (0.80)
Russian Federation	52.34 (1.91)	48.21 (1.23)
Singapore	74.53 (1.57)	67.46 (1.06)
Sweden	44.71 (1.11)	40.48 (0.62)
Turkey	41.13 (1.51)	39.75 (0.84)
United Arab Emirates	39.88 (1.90)	34.59 (0.38)
United States	45.31 (1.70)	43.84 (1.01)
International Average	**47.37 (0.33)**	**43.72 (0.18)**

( ) Standard errors appear in parentheses.

Exhibit 13.5: eTIMSS 2019 Average Percent Correct on Paper Bridge and eTIMSS Invariant Items – Eighth Grade Science

Country	Paper_INV	eTIMSS_INV
Chile	36.21 (0.60)	35.03 (0.49)
Chinese Taipei	61.50 (0.96)	58.14 (0.44)
England	50.08 (1.03)	45.75 (0.88)
Georgia	33.50 (1.16)	32.75 (0.59)
Hong Kong SAR	50.41 (1.03)	43.10 (0.96)
Hungary	48.69 (1.35)	47.64 (0.56)
Israel	43.75 (1.27)	45.42 (0.83)
Italy	41.16 (0.82)	41.06 (0.53)
Korea, Rep. of	57.07 (0.77)	54.64 (0.49)
Lithuania	47.64 (0.88)	48.00 (0.64)
Malaysia	39.23 (1.86)	36.98 (0.52)
Norway (9)	43.57 (0.89)	41.15 (0.48)
Qatar	44.34 (1.02)	39.76 (0.82)
Russian Federation	52.27 (1.18)	50.38 (0.97)
Singapore	67.62 (1.26)	65.64 (0.78)
Sweden	49.21 (1.08)	47.56 (0.62)
Turkey	48.39 (1.23)	47.01 (0.76)
United Arab Emirates	44.84 (1.72)	40.80 (0.38)
United States	48.99 (1.16)	47.87 (0.82)
International Average	**47.81 (0.27)**	**45.72 (0.16)**

( ) Standard errors appear in parentheses.

Exhibit 13.6 shows the international average percent correct across countries for the invariant trend items for the paper bridge and eTIMSS as well as the difference between them, together with their standard errors. At both fourth and eighth grades, there was a small but significant average international difference favoring the paper bridge in each subject, with a smaller difference in science than mathematics. These international mode effects require an international adjustment for each subject and grade before country differences can be properly evaluated.

Exhibit 13.6: eTIMSS 2019 International Average Percent Correct on Paper Bridge and eTIMSS Invariant Items

Grade 4	Bridge	eTIMSS	Difference
Mathematics	53.42 (0.23)	50.77 (0.13)	2.65 (0.26) ▲
Science	51.51 (0.20)	49.69 (0.11)	1.82 (0.23) ▲
Grade 8	**Bridge**	**eTIMSS**	**Difference**
Mathematics	47.37 (0.33)	43.72 (0.18)	3.66 (0.38) ▲
Science	47.81 (0.27)	45.72 (0.16)	2.09 (0.31) ▲

( ) Standard errors appear in parentheses.
▲ indicates the bridge students performed significantly higher than the eTIMSS students (α= 0.05).

Exhibits 13.7 and 13.8 show for fourth grade mathematics and science the country mode differences between the paper bridge and eTIMSS data, having adjusted for the average international differences. The country deviations from the international percent correct (difference between country average percent correct and the international average percent correct) are shown for the paper bridge and eTIMSS, together with their standard errors. For example, Austria's deviation for the bridge was 1.41 (0.94) and for eTIMSS was 2.10 (0.56). The relative difference for the country is the difference between the two deviations, e.g., –0.69 for Austria, which is not significant. The relative difference represents the country mode difference adjusted for the average international difference between modes (see Appendix 13A).

Exhibit 13.7: eTIMSS 2019 Country Deviations from International Average Percent Correct for Paper Bridge and eTIMSS Invariant Items and their Differences – Fourth Grade Mathematics

Country	Bridge	eTIMSS	Difference	
Austria	1.41 (0.94)	2.10 (0.56)	–0.69 (1.09)	
Canada	–5.54 (1.04)	–4.53 (0.49)	–1.01 (1.15)	
Chile	–21.45 (0.96)	–20.85 (0.52)	–0.60 (1.09)	
Chinese Taipei	18.51 (0.68)	19.12 (0.42)	–0.61 (0.80)	
Croatia	–6.18 (0.90)	–6.52 (0.53)	0.34 (1.04)	
Czech Republic	–2.46 (1.49)	1.23 (0.67)	–3.70 (1.63)	▽
Denmark	–1.20 (0.95)	–1.36 (0.58)	0.16 (1.12)	
England	6.02 (1.29)	6.55 (0.80)	–0.52 (1.51)	
Finland	0.50 (0.87)	0.74 (0.54)	–0.23 (1.02)	
France	–13.38 (0.90)	–11.70 (0.72)	–1.69 (1.15)	
Georgia	–5.29 (1.85)	–10.93 (0.82)	5.64 (2.03)	▲
Germany	–2.46 (1.05)	–3.02 (0.60)	0.56 (1.21)	
Hong Kong SAR	19.47 (1.73)	18.68 (0.81)	0.80 (1.91)	
Hungary	–0.62 (1.29)	–2.05 (0.68)	1.42 (1.45)	
Italy	–5.28 (1.28)	–4.39 (0.71)	–0.89 (1.47)	
Korea, Rep. of	15.83 (0.66)	16.79 (0.56)	–0.95 (0.87)	
Lithuania	6.59 (0.71)	3.88 (0.78)	2.71 (1.06)	▲
Netherlands	–0.29 (1.08)	0.62 (0.61)	–0.91 (1.24)	
Norway (5)	2.10 (1.06)	3.22 (0.55)	–1.13 (1.19)	
Portugal	1.33 (1.16)	–1.41 (0.73)	2.73 (1.37)	▲
Qatar	–17.11 (1.13)	–16.84 (0.74)	–0.27 (1.35)	
Russian Federation	8.40 (1.00)	9.62 (0.83)	–1.22 (1.30)	
Singapore	23.48 (1.16)	24.21 (0.88)	–0.72 (1.46)	
Slovak Republic	–5.75 (1.29)	–5.37 (0.74)	–0.38 (1.49)	
Spain	–8.24 (1.10)	–8.41 (0.49)	0.17 (1.21)	
Sweden	–3.87 (1.61)	–2.87 (0.78)	–1.00 (1.79)	
United Arab Emirates	–7.32 (1.86)	–9.69 (0.36)	2.36 (1.90)	
United States	2.79 (1.16)	3.17 (0.68)	–0.38 (1.35)	

( ) Standard errors appear in parentheses.
▲ indicates the bridge students performed significantly higher than the eTIMSS students (α= 0.05).
▽ indicates the bridge students performed significantly lower than the eTIMSS students (α= 0.05).

Exhibit 13.8: TIMSS 2019 Country Deviations from International Average Percent Correct for Paper Bridge and eTIMSS Invariant Items and their Differences – Fourth Grade Science

Country	Bridge	eTIMSS	Difference	
Austria	–2.91 (0.80)	–1.58 (0.44)	–1.33 (0.91)	
Canada	–2.79 (1.03)	–1.19 (0.35)	–1.60 (1.09)	
Chile	–12.43 (0.83)	–11.21 (0.47)	–1.22 (0.95)	
Chinese Taipei	6.69 (0.63)	5.77 (0.35)	0.91 (0.72)	
Croatia	–0.67 (0.75)	–0.23 (0.46)	–0.44 (0.88)	
Czech Republic	–0.96 (1.55)	1.39 (0.45)	–2.35 (1.62)	
Denmark	–2.66 (0.88)	–1.83 (0.46)	–0.83 (0.99)	
England	3.83 (1.05)	3.82 (0.52)	0.01 (1.17)	
Finland	5.14 (0.81)	6.63 (0.48)	–1.49 (0.94)	
France	–10.57 (0.80)	–8.65 (0.60)	–1.92 (1.00)	
Georgia	–8.69 (1.41)	–14.29 (0.61)	5.60 (1.53)	▲
Germany	0.01 (1.03)	–0.82 (0.55)	0.83 (1.17)	
Hong Kong SAR	3.68 (1.51)	1.78 (0.72)	1.90 (1.67)	
Hungary	1.98 (1.18)	1.69 (0.57)	0.30 (1.31)	
Italy	–3.64 (0.79)	–3.53 (0.58)	–0.10 (0.98)	
Korea, Rep. of	14.94 (0.57)	15.41 (0.47)	–0.47 (0.74)	
Lithuania	3.02 (0.68)	3.16 (0.54)	–0.14 (0.87)	
Netherlands	–3.21 (0.98)	–2.03 (0.59)	–1.18 (1.14)	
Norway (5)	3.03 (0.70)	2.93 (0.48)	0.09 (0.85)	
Portugal	–4.47 (0.76)	–6.17 (0.43)	1.69 (0.87)	
Qatar	–9.28 (1.46)	–12.62 (0.70)	3.35 (1.61)	▲
Russian Federation	10.03 (0.97)	9.93 (0.76)	0.10 (1.23)	
Singapore	17.49 (1.10)	17.63 (0.74)	–0.14 (1.32)	
Slovak Republic	–2.49 (0.94)	–1.60 (0.63)	–0.89 (1.13)	
Spain	–2.47 (0.96)	–2.95 (0.40)	0.48 (1.04)	
Sweden	0.02 (1.30)	3.37 (0.72)	–3.35 (1.49)	▽
United Arab Emirates	–5.44 (1.58)	–7.87 (0.35)	2.43 (1.62)	
United States	2.82 (1.12)	3.07 (0.56)	–0.24 (1.25)	

( ) Standard errors appear in parentheses.
▲ indicates the bridge students performed significantly higher than the eTIMSS students (α= 0.05).
▽ indicates the bridge students performed significantly lower than the eTIMSS students (α= 0.05).

Exhibits 13.9 and 13.10 provide the relative differences in the percentage correct metric for eighth grade mathematics and science, respectively. Note that overall international differences were accounted for so that within country comparisons reflect the relative differences.

Exhibit 13.9: eTIMSS 2019 Country Deviations from International Average Percent Correct for Paper Bridge and eTIMSS Invariant Items and their Differences – Eighth Grade Mathematics

Country	Bridge	eTIMSS	Difference	
Chile	–20.44 (0.76)	–17.13 (0.47)	–3.30 (0.89)	▽
Chinese Taipei	22.83 (1.09)	21.32 (0.53)	1.51 (1.27)	
England	0.64 (1.54)	–1.61 (1.10)	2.25 (1.93)	
Georgia	–14.59 (1.15)	–13.29 (0.80)	–1.30 (1.45)	
Hong Kong SAR	15.07 (1.59)	13.44 (0.90)	1.63 (1.87)	
Hungary	0.05 (1.74)	–0.18 (0.67)	0.23 (1.90)	
Israel	–1.15 (1.56)	1.14 (0.97)	–2.29 (1.88)	
Italy	–7.91 (0.69)	–6.55 (0.56)	–1.36 (0.97)	
Korea, Rep. of	22.34 (0.70)	21.31 (0.52)	1.03 (0.96)	
Lithuania	–3.15 (1.19)	–0.73 (0.63)	–2.42 (1.40)	
Malaysia	–10.16 (1.93)	–11.71 (0.52)	1.55 (2.04)	
Norway (9)	–3.52 (0.93)	–2.89 (0.53)	–0.64 (1.14)	
Qatar	–13.68 (1.12)	–15.17 (0.75)	1.49 (1.41)	
Russian Federation	4.97 (1.81)	4.50 (1.17)	0.47 (2.19)	
Singapore	27.16 (1.49)	23.74 (1.01)	3.42 (1.83)	
Sweden	–2.66 (1.05)	–3.23 (0.59)	0.57 (1.26)	
Turkey	–6.24 (1.43)	–3.97 (0.80)	–2.28 (1.68)	
United Arab Emirates	–7.50 (1.80)	–9.12 (0.36)	1.63 (1.87)	
United States	–2.06 (1.61)	0.13 (0.95)	–2.19 (1.91)	

( ) Standard errors appear in parentheses.
▲ indicates the bridge students performed significantly higher than the eTIMSS students (α= 0.05).
▽ indicates the bridge students performed significantly lower than the eTIMSS students (α= 0.05).

Exhibit 13.10: eTIMSS 2019 Country Deviations from International Average Percent Correct for Paper Bridge and eTIMSS Invariant Items and their Differences – Eighth Grade Science

Country	Bridge	eTIMSS	Difference	
Chile	–11.60 (0.63)	–10.69 (0.49)	–0.91 (0.80)	
Chinese Taipei	13.69 (0.91)	12.42 (0.41)	1.26 (1.05)	
England	2.27 (0.97)	0.03 (0.84)	2.24 (1.32)	
Georgia	–14.32 (1.09)	–12.97 (0.56)	–1.35 (1.27)	
Hong Kong SAR	2.60 (0.98)	–2.62 (0.91)	5.22 (1.37)	▲
Hungary	0.88 (1.27)	1.92 (0.53)	–1.04 (1.41)	
Israel	–4.06 (1.20)	–0.30 (0.79)	–3.76 (1.47)	▽
Italy	–6.66 (0.77)	–4.66 (0.51)	–1.99 (0.97)	▽
Korea, Rep. of	9.26 (0.73)	8.92 (0.46)	0.33 (0.91)	
Lithuania	–0.17 (0.83)	2.28 (0.61)	–2.46 (1.07)	▽
Malaysia	–8.59 (1.76)	–8.74 (0.49)	0.15 (1.85)	
Norway (9)	–4.24 (0.84)	–4.57 (0.45)	0.33 (1.00)	
Qatar	–3.47 (0.96)	–5.96 (0.77)	2.49 (1.27)	
Russian Federation	4.45 (1.12)	4.66 (0.92)	–0.21 (1.48)	
Singapore	19.81 (1.19)	19.92 (0.74)	–0.12 (1.44)	
Sweden	1.40 (1.02)	1.84 (0.59)	–0.44 (1.22)	
Turkey	0.58 (1.16)	1.29 (0.72)	–0.72 (1.40)	
United Arab Emirates	–2.98 (1.63)	–4.92 (0.36)	1.95 (1.69)	
United States	1.17 (1.10)	2.15 (0.78)	–0.98 (1.38)	

( ) Standard errors appear in parentheses.
▲ indicates the bridge students performed significantly higher than the eTIMSS students (α= 0.05).
▽ indicates the bridge students performed significantly lower than the eTIMSS students (α= 0.05).

Exhibits 13.7 through 13.10 provide a way to evaluate whether countries had positive or negative mode effects over and above the international effect. Although most differences were not statistically significant given their standard errors, there were some differences, mostly small. Also, when computing a large number of significance tests, some number (5% or so) will appear statistically significant just by means of random sampling variability rather than underlying mode differences.

The **estimated** differences observed when looking at a relatively large number of country-mean differences follow a statistical distribution around the 'true' differences. Some are smaller and some are larger, and should the exercise be repeated, and another set of bridge and eTIMSS samples collected, a country's difference most likely would be slightly lower or slightly higher. This is the well-known 'regression to the mean' effect, whereby if one repeats a data collection the observations showing the most extreme estimates in the original study may not show as extreme estimates in the replication (e.g. Efron, 2011).

Country-Level Differences in TIMSS 2019 Average Scale Scores (Plausible Values) by Mode of Administration

As a consequence of the eTIMSS-paperTIMSS linking approach used in the achievement scaling, the eTIMSS scale scores and bridge scale scores are on the same TIMSS 2019 achievement scales and can be directly compared. However, the eTIMSS scale scores are based on all of the 2019 achievement items, while the bridge scale scores are based on only the trend items and were estimated from samples one-third the size of eTIMSS. For each grade and subject, the item parameters from the paperTIMSS trend scaling were applied to the eTIMSS data with a small constant adjustment to account for the average international difference (the international mode effect) between the paper and eTIMSS versions (see Chapter 12; von Davier et al., 2019a,b). Thus, country differences in scale scores between the eTIMSS and paper bridge data are the result of country mode effects and sampling differences.

Exhibits 13.11 and 13.12 show average country differences between average eTIMSS and paper bridge scale scores for fourth grade mathematics and science.

Exhibit 13.11: eTIMSS 2019 Average Scale Scores for eTIMSS and Paper Bridge and their Differences – Fourth Grade Mathematics

Country	Bridge Average Score	eTIMSS Average Score	Difference	
Austria	534 (3.3)	539 (2.0)	–5 (3.8)	
Canada	512 (3.8)	512 (1.9)	0 (4.3)	
Chile	436 (4.4)	441 (2.7)	–5 (5.1)	
Chinese Taipei	603 (2.6)	599 (1.9)	4 (3.2)	
Croatia	511 (3.4)	509 (2.2)	2 (4.0)	
Czech Republic	519 (7.3)	533 (2.5)	–14 (7.7)	
Denmark	528 (3.6)	525 (1.9)	3 (4.1)	
England	553 (5.0)	556 (3.0)	–3 (5.9)	
Finland	533 (3.4)	532 (2.3)	1 (4.1)	
France	481 (3.7)	485 (3.0)	–4 (4.8)	
Georgia	505 (8.0)	482 (3.7)	23 (8.8)	▲
Germany	519 (4.2)	521 (2.3)	–2 (4.8)	
Hong Kong SAR	607 (7.9)	602 (3.3)	6 (8.5)	
Hungary	530 (5.1)	523 (2.6)	7 (5.8)	
Italy	511 (4.9)	515 (2.4)	–4 (5.5)	
Korea, Rep. of	595 (2.5)	600 (2.2)	–5 (3.3)	
Lithuania	547 (2.8)	542 (2.8)	5 (3.9)	
Netherlands	528 (4.1)	538 (2.2)	–9 (4.6)	▽
Norway (5)	540 (3.9)	543 (2.2)	–2 (4.4)	
Portugal	536 (4.5)	525 (2.6)	11 (5.2)	▲
Qatar	450 (6.4)	449 (3.4)	0 (7.2)	
Russian Federation	559 (3.9)	567 (3.3)	–8 (5.1)	
Singapore	631 (5.6)	625 (3.9)	6 (6.8)	
Slovak Republic	505 (4.7)	510 (3.5)	–5 (5.9)	
Spain	502 (4.8)	502 (2.1)	–1 (5.2)	
Sweden	517 (5.8)	521 (2.8)	–5 (6.4)	
United Arab Emirates	496 (7.9)	481 (1.7)	14 (8.1)	
United States	537 (5.1)	535 (2.5)	2 (5.7)	
International Average	**529 (1.0)**	**528 (0.6)**	**1 (1.2)**	

( ) Standard errors appear in parentheses.
▲ indicates the bridge students performed significantly higher than the eTIMSS students (α= 0.05).
▽ indicates the bridge students performed significantly lower than the eTIMSS students (α= 0.05).

Exhibit 13.12: eTIMSS 2019 Average Scale Scores for eTIMSS and Paper Bridge and their Differences – Fourth Grade Science

Country	Bridge Average Score	eTIMSS Average Score	Difference	
Austria	511 (3.9)	522 (2.6)	–11 (4.7)	▽
Canada	512 (4.5)	523 (1.9)	–11 (4.9)	▽
Chile	461 (4.5)	469 (2.6)	–8 (5.1)	
Chinese Taipei	554 (2.9)	558 (1.8)	–4 (3.4)	
Croatia	524 (3.5)	524 (2.2)	0 (4.1)	
Czech Republic	517 (9.4)	534 (2.6)	–16 (9.8)	
Denmark	514 (4.3)	522 (2.4)	–8 (4.9)	
England	543 (4.7)	537 (2.7)	6 (5.4)	
Finland	547 (4.0)	555 (2.6)	–8 (4.7)	
France	478 (4.0)	488 (3.0)	–10 (4.9)	
Georgia	477 (8.1)	454 (3.9)	23 (9.0)	▲
Germany	522 (4.7)	518 (2.2)	4 (5.2)	
Hong Kong SAR	542 (7.3)	531 (3.3)	11 (8.0)	
Hungary	533 (6.3)	529 (2.7)	3 (6.8)	
Italy	507 (4.1)	510 (3.0)	–3 (5.1)	
Korea, Rep. of	588 (2.6)	588 (2.1)	0 (3.4)	
Lithuania	539 (3.1)	538 (2.5)	1 (4.0)	
Netherlands	511 (4.5)	518 (2.9)	–7 (5.3)	
Norway (5)	536 (3.5)	539 (2.2)	–3 (4.1)	
Portugal	509 (3.5)	504 (2.6)	5 (4.3)	
Qatar	463 (8.6)	449 (3.9)	14 (9.5)	
Russian Federation	567 (4.2)	567 (3.0)	0 (5.1)	
Singapore	599 (5.1)	595 (3.4)	5 (6.2)	
Slovak Republic	512 (4.9)	521 (3.7)	–9 (6.1)	
Spain	514 (4.3)	511 (2.0)	3 (4.8)	
Sweden	523 (6.4)	537 (3.3)	–15 (7.2)	▽
United Arab Emirates	485 (8.6)	473 (2.1)	12 (8.8)	
United States	535 (5.4)	539 (2.7)	–3 (6.1)	
International Average	**522 (1.2)**	**523 (1.0)**	**0 (1.6)**	

( ) Standard errors appear in parentheses.
▲ indicates the bridge students performed significantly higher than the eTIMSS students (α= 0.05).
▽ indicates the bridge students performed significantly lower than the eTIMSS students (α= 0.05).

Exhibits 13.11 and 13.12 show that the differences for grade 4 are mostly non–significant. The only country with significant results in both mathematics and science was Georgia, where students taking eTIMSS had lower achievement than those taking the paper bridge. In Portugal, the students taking the eTIMSS appear to perform lower in mathematics compared to the bridge but not in science. In the Netherlands, students taking eTIMSS performed better in mathematics, but there was no difference in science. In Sweden, Austria, and Canada, the students taking eTIMSS performed better in science than those who took the bridge, while there was no difference in mathematics.

Exhibits 13.13 and 13.14 show for eighth grade mathematics and science the comparisons of bridge and eTIMSS samples in terms of achievement estimates (plausible values) based on the randomly equivalent groups.

Exhibit 13.13: eTIMSS 2019 Average Scale Scores for eTIMSS and Paper Bridge and their Differences – Eighth Grade Mathematics

Country	Bridge Average Score	eTIMSS Average Score	Difference	
Chile	434 (3.3)	441 (2.8)	–6 (4.3)	
Chinese Taipei	618 (5.4)	612 (2.7)	5 (6.1)	
England	526 (6.0)	515 (5.3)	11 (8.0)	
Georgia	452 (7.1)	461 (4.3)	–9 (8.3)	
Hong Kong SAR	581 (6.9)	578 (4.1)	3 (8.0)	
Hungary	521 (7.1)	517 (2.9)	5 (7.7)	
Israel	511 (7.1)	519 (4.3)	–8 (8.3)	
Italy	495 (3.5)	497 (2.7)	–2 (4.4)	
Korea, Rep. of	613 (3.6)	607 (2.8)	7 (4.6)	
Lithuania	510 (5.2)	520 (2.9)	–11 (6.0)	
Malaysia	473 (9.9)	461 (3.2)	13 (10.4)	
Norway (9)	509 (3.9)	503 (2.4)	7 (4.6)	
Qatar	452 (6.3)	443 (4.0)	9 (7.5)	
Russian Federation	543 (7.5)	543 (4.5)	–1 (8.8)	
Singapore	630 (6.5)	616 (4.0)	15 (7.7)	
Sweden	513 (4.8)	503 (2.5)	11 (5.5)	▲
Turkey	487 (7.1)	496 (4.3)	–9 (8.3)	
United Arab Emirates	482 (8.6)	473 (1.9)	8 (8.8)	
United States	512 (6.4)	515 (4.8)	–4 (8.0)	
International Average	**519 (1.5)**	517 **(0.8)**	**2 (1.7)**	

( ) Standard errors appear in parentheses.
▲ indicates the bridge students performed significantly higher than the eTIMSS students (α= 0.05).
▽ indicates the bridge students performed significantly lower than the eTIMSS students (α= 0.05).

Exhibit 13.14: eTIMSS 2019 Average Scale Scores for eTIMSS and Paper Bridge and their Differences – Eighth Grade Science

Country	Bridge Average Score	eTIMSS Average Score	Difference	
Chile	458 (3.7)	462 (2.9)	–4 (4.7)	
Chinese Taipei	584 (5.0)	574 (1.9)	10 (5.3)	
England	529 (5.0)	517 (4.8)	13 (7.0)	
Georgia	435 (7.5)	447 (3.9)	–12 (8.4)	
Hong Kong SAR	531 (5.3)	504 (5.2)	27 (7.4)	▲
Hungary	523 (6.5)	530 (2.6)	–6 (7.0)	
Israel	498 (7.0)	513 (4.2)	–16 (8.2)	
Italy	487 (4.3)	500 (2.6)	–13 (5.0)	▽
Korea, Rep. of	563 (3.6)	561 (2.1)	2 (4.2)	
Lithuania	522 (4.5)	534 (3.0)	–12 (5.4)	▽
Malaysia	469 (10.9)	460 (3.5)	9 (11.4)	
Norway (9)	500 (4.7)	495 (3.1)	5 (5.6)	
Qatar	495 (5.1)	475 (4.4)	20 (6.7)	▲
Russian Federation	544 (6.1)	543 (4.2)	1 (7.4)	
Singapore	611 (6.1)	608 (3.9)	3 (7.3)	
Sweden	521 (6.1)	521 (3.2)	0 (6.8)	
Turkey	518 (6.4)	515 (3.7)	2 (7.4)	
United Arab Emirates	490 (9.9)	473 (2.2)	17 (10.2)	
United States	524 (6.0)	522 (4.7)	1 (7.6)	
International Average	**516 (1.6)**	513 **(0.8)**	**3 (1.8)**	

( ) Standard errors appear in parentheses.
▲ indicates the bridge students performed significantly higher than the eTIMSS students (α= 0.05).
▽ indicates the bridge students performed significantly lower than the eTIMSS students (α= 0.05).

As shown in Exhibits 13.13 and 13.14, the differences for the eighth grade are mostly non–significant. In Sweden, students taking eTIMSS performed lower in mathematics, but there was no difference in science. In Hong Kong SAR and Qatar, students taking eTIMSS performed lower in science than those who took the bridge, while there was no difference in mathematics. In Italy and Lithuania, the eTIMSS students performed higher in science than the bridge students.

Exhibits 13.11 through 13.14 show a small number of significant differences between eTIMSS and the paper bridge. Of the few countries with mode differences, they were either in mathematics or in science and occurred in both directions, with the exception of a paper bridge advantage for Georgia at the fourth grade. Also, these exhibits contain a total of 94 mode comparisons, 56 for fourth grade and 38 for eighth grade, and we have not made adjustments for multiple comparisons (e.g. Shaffer, 1995; Benjamini & Hochberg, 1995). About 5 significant differences would be expected to occur purely at random among 94 comparisons at the α=0.05 level.

Summary

The present chapter provides an overview of how countries can use their bridge data together with their eTIMSS data to evaluate the extent of mode differences in their TIMSS 2019 data. First, to provide an accessible approach to the study of country mode effects, country differences in the average percent correct between the paper bridge and eTIMSS were examined for those trend items found to be equivalent. Then, after subtracting the average international difference from the country average for both the bridge and eTIMSS, the difference between the bridge and eTIMSS country deviations provides an estimate of the country mode effect. However, only a few countries had significant mode effects, and these were for most countries isolated instances of one subject in one grade.

Second, country differences between average eTIMSS scale scores and scale scores estimated for the paper bridge were examined. Similar to the results from the percent correct analyses, country differences in average scale scores between eTIMSS and the paper bridge were small, and few were flagged as statistically significant. Because the bridge scale scores were based on smaller samples and fewer items than the eTIMSS scale scores, these differences may be due to sample differences in addition to residual differences that were present in the linked scales.

This chapter is intended to encourage researchers interested in examining how mode effects can differ among countries, types of items, or student groups. More in depth studies by country experts may be worthwhile to explore to what extent differences in performance between paper bridge and eTIMSS can be attributed to residual mode effects versus sample differences.

References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57(1), 289–300.

Box, G. E. P., Hunter, J. S., & Hunter, W.G . (2005). *Statistics for experimenters: Design, innovation, and discovery, Second Edition*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7*, 1–26. doi:10.1214/aos/1176344552

Efron, B. (2011) Tweedie's formula and selection bias. *Journal of the American Statistical Association, 106*, 1602–1614.

Fishbein, B., Martin, M. O., Mullis, I. V. S., & Foy, P. (2018). The TIMSS 2019 item equivalence study: Examining mode effects for computer-based assessment and implications for measuring trends. *Large-scale Assessments in Education*, 6(1), 11. https://doi.org/10.1186/s40536-018-0064-z

Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology, 46*, 561–584. doi:10.1146/annurev.ps.46.020195.003021

von Davier, M., Yamamoto, K., Shin, H-J., Chen, H., Khorramdel, L., Weeks, J., Davis, S., Kong, N., & Kandathil, M. (2019a). Evaluating item response theory linking and model fit for data from PISA 2000–2012. *Assessment in Education: Principles, Policy & Practice, 26*(4), 466–488. https://doi.org/10.1080/0969594X.2019.1586642

von Davier, M., Khorramdel, L., He, Q., Shin, H. J., & Chen, H. (2019b). Developments in psychometric population models for technology-based large-scale assessments: An overview of challenges and opportunities. *Journal of Educational and Behavioral Statistics, 44*(6), 671–705. https://doi.org/10.3102/1076998619881789

Appendix 13A

Comparing Country Level Proportion Correct to International Averages

Consider the international average of a statistic, for example an average proportion correct over a number of item responses. In our case, these are the items that were designed for a paper based assessment, TIMSS 2015, and that were still used in TIMSS 2019 as trend items for computer based countries, and were re-implemented for computer delivery for countries that chose to use the eTIMSS assessment. These trend items were carefully designed for computer delivery so that a majority of 80% or more per grade and subject domain was considered equivalent in terms of how they relate to achievement on the TIMSS scale.

The international average of the average percent-correct typically based on equal contribution of all participating countries, that is, they are defined as an unweighted average. Formally, we have

$$\mu_I = \frac{1}{C}\sum_{c=1}^{C}\mu_c.$$

Obviously, we do not have the true population values at the country level, as we only collect a sample of schools, and 1 or 2 classrooms per school. The best estimate of the country average percentages are the weighted estimates of the proportion correct, i.e., the weighted sum of correct responses, divided by the sum of weights, over the items that are considered comparable.

The international estimate $\widehat{M}_I$ of this proportion correct has estimation error as well, as it is also based on sampling, albeit over multiple countries. We denote the standard error associated with this average by $\widehat{S}_I$. Assuming unbiased sample-based estimates, we have

$$E(\widehat{M}_I) = E\left(\sum_{c=1}^{C}\widehat{M}_c\right) = \mu_I$$

with estimates of country means $\widehat{M}_c$ that are based on the country sample, we also assume these are unbiased, i.e.,

$$E(\widehat{M}_c) = \mu_c$$

and denote the associated standard errors by $\widehat{S}_c$. For an estimate of the difference, $\widehat{d}_k = \widehat{M}_k - \widehat{M}_I$ of a country k's mean and overall mean $\Delta_k = \mu_k - \mu_I$ we observe the following complication. The estimate of the international mean $\widehat{M}_I$ contains the country mean $\widehat{M}_k$ as one component. This implies

$$\hat{S}_{d(k)} = \sqrt{\hat{S}_I^2 + \hat{S}_k^2 - 2cov(\widehat{M}_I, \widehat{M}_k)}$$

with

$$cov(\widehat{M}_I, \widehat{M}_k) = cov\left(\frac{1}{C}\sum_{c=1}^{C}\widehat{M}_c, \widehat{M}_k\right) = \frac{1}{C}cov(\widehat{M}_k, \widehat{M}_k) = \frac{1}{C}\hat{S}_k^2.$$

Plugging this result into the estimate provides

$$\hat{S}_{d(k)} = \sqrt{\hat{S}_I^2 + \hat{S}_k^2 - \frac{2}{C}\hat{S}_k^2} = \sqrt{\hat{S}_I^2 + \left[\frac{C-2}{C}\right]\hat{S}_k^2}$$

which is well defined whenever there are at least two countries, i.e., whenever $C \geq 2$.

Country Mode Differences, Corrected for International Mode Differences

The international estimate and the expected values of proportion correct of paper items ('P'-samples) will be denoted by

$$E(\widehat{M}_{IP}) = \mu_{IP}$$

and the mean of proportion correct across computer based ('E'-samples) is

$$E(\widehat{M}_{IC}) = \mu_{IE}.$$

Similarly, we have associated standard errors for the estimate of the international proportion correct for paper, S_{IP}, and computer, S_{IE}, respectively, as we have for the country level estimates S_{kP} and S_{kE}. These can be calculated separately using the jackknife procedures and defined as given above. The bridge and the eTIMSS samples do provide an estimate $\hat{d}_{P-E}$ of the mode difference

$$\Delta_{P-E} = \mu_{IP} - \mu_{IE}$$

at the international level. This mode difference is being controlled for in the linking design that uses the bridge and eTIMSS samples in a customary equivalent groups approach. That means this difference is no longer relevant and can be taken out of country level comparisons of the effect of mode on achievement results. Only any remaining differences that are based on differences at the country levels are relevant, as the overall difference is no longer affecting the plausible values that are provided in the international database.

That means, in order to examine whether there is a difference between the paper- and the eTIMSS proportion correct at the country-level that goes beyond what would be expected internationally. Only differences that go beyond this are relevant, as the international average of correct response differences is already taken care of by the mode effect adjustment. Consequently, the difference

$$\hat{d}_{Pk-Ek} = \hat{d}_{Pk} - \hat{d}_{Ek} \approx \widehat{M}_{Pk} - \widehat{M_{Ek}} - \Delta_{P-E}$$

quantifies the relative paper versus eTIMSS difference of proportions correct that not accounted for by the international linking in the bridge study. For this estimated difference, we can use the standard error

$$\hat{S}_{Pk-Ek} = \sqrt{\hat{S}_{Pd(k)} + \hat{S}_{Ed(k)}}$$

With the estimates defined as above

$$\hat{S}_{Pd(k)} = \sqrt{\hat{S}_{IP}^2 + \left[\frac{C-2}{C}\right]\hat{S}_{kP}^2}$$

and

$$\hat{S}_{Ed(k)} = \sqrt{\hat{S}_{IE}^2 + \left[\frac{C-2}{C}\right]\hat{S}_{kE}^2}.$$

Note that these are almost the same as the s.e. for the country mean proportions correct for paper versus eTIMSS, calculated separately. This statistic is adjusted by the s.e. for the international proportion correct (separately calculated by mode) but adjusted for the number of countries included in the international mean proportions.

Achievement data comparisons based on Bridge and eTIMSS samples

The comparison, once the linking is accomplished, is rather straightforward. The standard error estimates for the bridge sample averages and the eTIMSS averages can be used to calculate the standard error of the difference for countries where schools were selected to test either using the paperTIMSS or the eTIMSS assessment. These can, within countries, be assumed to be independent samples, and if the schools were randomly assigned to the mode of assessment, these independent samples can be assumed to be identically distributed. In practice, this may not be completely true, as schools that were able to test on computer may have been somewhat different from schools that were assigned to test with the bridge/paperTIMSS instruments. For example, hardware availability in schools may be associated with average socio-economic status of students' parents.

Assuming independent samples from the same population, the mean difference

$$\mu_{cB} - \mu_{cE} = \Delta$$

Between bridge sample (B) and eTIMSS sample (E) in country c can be evaluated using the standard error of the difference for independent samples,

$$s.e.(\Delta) = \sqrt{s.e.(\mu_{cB})^2 + s.e.(\mu_{cE})^2}.$$

However, this is no longer appropriate and may overestimate the s.e. if students were assigned to paper or eTIMSS within schools. In this case, samples are dependent, and the difference of the achievement per school needs to be calculated and the variance of this difference needs to be estimated using an appropriate resampling method (Efron, 1979). The bridge and the eTIMSS samples would in some countries be drawn in the same schools, but different classes, while in other countries the two samples would come from schools without overlap, while a third set of countries would have some schools that assign one class to paperTIMSS and another to eTIMSS, and other schools would only assign one class to one of the modes. The assumption of independent samples is applicable in the case that the different classes perform independently of being sampled in the same or in different schools. If schools are very different compared to between class differences within schools (i.e., there is tracking between schools, but little tracking within schools) this will lead to overestimation of standard errors.

In the exhibits, we assume for simplicity of exposition, we assume independent samples of students taking the eTIMSS and the paperTIMSS assessment.

IEA TIMSS 2019

CHAPTER 14

Estimating Standard Errors in the TIMSS 2019 Results

Pierre Foy
Sylvie LaRoche

To obtain estimates of students' proficiency in mathematics and science that are both accurate and cost-effective, TIMSS 2019 made extensive use of probability sampling techniques to sample students from the national fourth and eighth grade student populations, and applied matrix-sampling assessment designs to target individual students with a subset of the complete pool of assessment items. This approach made efficient use of resources, in particular keeping student response burden to a minimum, but at a cost of some variance or uncertainty in the reported statistics, such as the means and percentages computed to estimate population parameters.

To quantify this uncertainty, each statistic in the TIMSS 2019 international report is accompanied by an estimate of its standard error. Statistics based on differences arising from comparing two estimated results also have standard errors, which serve to calculate confidence intervals or to perform statistical tests of significance. For statistics reporting student achievement, which are based on plausible values, standard errors are calculated based on two components. The first reflects the uncertainty due to generalizing from a student sample to the entire student population from which it was drawn, referred to as sampling variance. The second is known as imputation variance and reflects uncertainty due to inferring students' achievement estimates from their observed performance on a subset of achievement items and other achievement related information. This variance component reflects the posterior variance of the achievement variables given all available information used in the achievement imputation model described in Chapter 11 of this volume. For parameter estimates of variables that are not plausible values, the estimates of standard errors are based entirely on sampling variance.

Estimating Sampling Variance

TIMSS makes extensive use of probability sampling to derive achievement results from national samples of students. Because many such samples are possible but only one sample is drawn, some uncertainty about how well the sample represents the population is to be expected. The uncertainty caused by sampling students from a target population, known as sampling variance, can be estimated from the data of the one sample drawn.

Whereas estimating the sampling variance from simple random samples is a relatively simple task, estimating the sampling variance from the complex sample design of TIMSS is a more challenging endeavor. A common way to estimate the sampling variance in multistage cluster sampling designs is through resampling schemes (Efron, 1982) such as the balanced repeated replication and Jackknife techniques (Johnson & Rust, 1992; Quenouille, 1949; Tukey, 1958; Wolter, 1985). TIMSS uses a variation of the Jackknife, Jackknife Repeated Replication (JRR), to estimate sampling variances. JRR was chosen because it is computationally straightforward and provides approximately unbiased estimates of the sampling variance of means, totals, and percentages.

At the core of the JRR technique is the repeated resampling from the one sample drawn, under identical sample design conditions. In the context of TIMSS, this entails the grouping of primary sampling units into sampling zones based on the TIMSS sample design and repeated draws of subsamples from these zones. The main features of the TIMSS sample design that JRR incorporates in its repeated replication are the stratification of schools and the clustering of students within schools. This was done by defining Jackknife sampling zones as pairs of successive schools[1] to model the stratification and clustering from the national samples (see Chapter 3 for information on the TIMSS Sample Design). The repeated subsampling required by JRR was applied within each sampling zone.

Sampling zones were constructed within explicit strata. When an explicit stratum had an odd number of schools, either by design or because of non-responding schools, the students in the lone school of the last sampling zone were divided randomly to make up two "quasi" schools for the purposes of calculating jackknife standard errors.[2] Each sampling zone then consisted of a pair of schools or "quasi" schools. Since most national samples consisted of a minimum of 150 schools, a total of 75 zones were created. If more than 150 schools were sampled, the additional zones were collapsed into the first 75 zones.[3]

Exhibit 14.1 shows the number of constructed Jackknife sampling zones, prior to any collapsing, for the participating countries and benchmarking participants in TIMSS 2019.[4]

1 When schools were sampled, they were ordered within explicit strata by implicit stratification variables and their measure of size. Based on this sorting, successively sampled schools had similar stratification attributes. More information can be found in Appendix 3A of Chapter 3.

2 If a remaining school consisted of 2 sampled classrooms, each classroom became a "quasi" school.

3 The randomization used in the resampling within sampling zones preserves the sampling variance measured in the original sampling zones after collapsing.

4 Note that jackknife sampling zones may be constructed in a different manner under specific national conditions. Country-specific information on the construction of Jackknife sampling zones is available in Appendix 9A of Chapter 9.

Exhibit 14.1: Number of Jackknife Sampling Zones in the TIMSS 2019 National Samples

Country	Fourth Grade		Eighth Grade	
	TIMSS Sample	Bridge Sample	TIMSS Sample	Bridge Sample
Albania	84	—	—	—
Armenia	76	—	—	—
Australia	145	—	145	—
Austria	98	50	—	—
Azerbaijan	97	—	—	—
Bahrain	118	—	131	—
Belgium (Flemish)	77	—	—	—
Bosnia and Herzegovina	99	—	—	—
Bulgaria	76	—	—	—
Canada	358	44	—	—
Chile	85	29	83	28
Chinese Taipei	81	34	105	29
Croatia	79	38	—	—
Cyprus	76	—	109	—
Czech Republic	77	30	—	—
Denmark	84	31	—	—
Egypt	—	—	86	—
England	71	26	69	25
Finland	80	36	79	—
France	78	30	75	—
Georgia	78	29	74	27
Germany	102	39	—	—
Hong Kong SAR	70	28	69	27
Hungary	76	26	78	27
Iran, Islamic Rep. of	112	—	110	—
Ireland	75	—	77	—
Israel	—	—	80	37
Italy	81	30	79	29
Japan	75	—	72	—
Jordan	—	—	122	—
Kazakhstan	85	—	85	—
Korea, Rep. of	76	34	85	33
Kosovo	73	—	—	—
Kuwait	82	—	87	—
Latvia	84	—	—	—

Exhibit 14.1: Number of Jackknife Sampling Zones in the TIMSS 2019 National Samples (continued)

Country	Fourth Grade		Eighth Grade	
	TIMSS Sample	Bridge Sample	TIMSS Sample	Bridge Sample
Lebanon	—	—	109	—
Lithuania	105	37	98	36
Malaysia	—	—	91	22
Malta	226	—	—	—
Montenegro	238	—	—	—
Morocco	149	—	132	—
Netherlands	57	21	—	—
New Zealand	80	—	70	—
North Macedonia	75	—	—	—
Northern Ireland	71	—	—	—
Norway	77	28	79	27
Oman	126	—	119	—
Pakistan	71	—	—	—
Philippines	91	—	—	—
Poland	75	—	—	—
Portugal	91	45	78	—
Qatar	138	33	109	32
Romania	—	—	99	—
Russian Federation	59	32	60	32
Saudi Arabia	111	—	106	—
Serbia	84	—	—	—
Singapore	187	28	153	28
Slovak Republic	80	35	—	—
South Africa	149	—	261	—
Spain	257	35	—	—
Sweden	73	26	77	27
Turkey	91	—	91	36
United Arab Emirates	722	52	671	47
United States	145	42	138	35
Benchmarking Participants				
Ontario, Canada	83	—	80	—
Quebec, Canada	75	—	63	—
Moscow City, Russian Fed.	76	—	76	—

Exhibit 14.1: Number of Jackknife Sampling Zones in the TIMSS 2019 National Samples (continued)

Country	Fourth Grade		Eighth Grade	
	TIMSS Sample	Bridge Sample	TIMSS Sample	Bridge Sample
Gauteng, RSA	—	—	75	—
Western Cape, RSA	—	—	75	—
Madrid, Spain	85	—	—	—
Abu Dhabi, UAE	263	—	254	—
Dubai, UAE	226	—	185	—

The JRR procedure draws two subsamples from each sampling zone: one where the first school in the pair is included and the second school is removed, and the other where the second school is included and the first school is removed. When a school is removed from a sampling zone, the sampling weights of the students in the remaining school are doubled to make up for the omitted school. In both subsamples, all students in the other sampling zones are included with their sampling weights unchanged. With this process applied in each of the 75 sampling zones, the JRR procedure yields a total of 150 replicate subsamples, each one with its own set of replicate sampling weights to account for the successive removal of each school from the pair of schools in any given sampling zone.

The process of creating replicate sampling weights for the replicate subsamples defines replicate factors k_{hi} as follows:

$$k_{hi} = \begin{cases} 2 \text{ for students in school } i \text{ of sampling zone } h \\ 0 \text{ for students in the other school of sampling zone } h \\ 1 \text{ for students in any other sampling zone} \end{cases} \tag{14.1}$$

These replicate factors are used to compute the 150 sets of replicate sampling weights as follows:

$$W_{hij} = k_{hi} \cdot W_{0j} \tag{14.2}$$

where W_{0j} is the overall sampling weight of student j and W_{hij} is the resulting replicate sampling weight of student j when school i from sampling zone h is included and the other school in the pair is removed.

Exhibit 14.2 illustrates the calculation of the replicate factors necessary to produce the replicate sampling weights. Within each sampling zone, each school is assigned randomly an indicator u_{hi}, coded either 0 or 1, such that one school has a value of 0 and the other a value of 1. This indicator serves to determine how schools within each zone will be successively included and removed. When a school is removed from a zone, the replicate factor is set to 0 and the sampling weights of all students in that school are set to 0. When a school is included, the replicate factor is set to 2 and the sampling weights of all students in that school are doubled. The sampling weights of students in all the other sampling zones remain unchanged.

Exhibit 14.2: Construction of Replicate Factors Across Sampling Zones

<table>
<tr><th rowspan="3">Sampling Zone</th><th rowspan="3">School Replicate Indicator (u_{hi})</th><th colspan="12">Replicate Factors for Computing JRR Replicate Sampling Weights (k_{hi})</th></tr>
<tr><th colspan="2">Zone 1</th><th colspan="2">Zone 2</th><th colspan="2">Zone 3</th><th rowspan="2">. . .</th><th colspan="2">Zone h</th><th rowspan="2">. . .</th><th colspan="2">Zone 75</th></tr>
<tr><th>(1)</th><th>(2)</th><th>(3)</th><th>(4)</th><th>(5)</th><th>(6)</th><th>(2h-1)</th><th>(2h)</th><th>(149)</th><th>(150)</th></tr>
<tr><td rowspan="2">1</td><td>0</td><td>2</td><td>0</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">. . .</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">. . .</td><td rowspan="2">1</td><td rowspan="2">1</td></tr>
<tr><td>1</td><td>0</td><td>2</td></tr>
<tr><td rowspan="2">2</td><td>0</td><td rowspan="2">1</td><td rowspan="2">1</td><td>2</td><td>0</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">. . .</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">. . .</td><td rowspan="2">1</td><td rowspan="2">1</td></tr>
<tr><td>1</td><td>0</td><td>2</td></tr>
<tr><td rowspan="2">3</td><td>0</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">1</td><td>2</td><td>0</td><td rowspan="2">. . .</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">. . .</td><td rowspan="2">1</td><td rowspan="2">1</td></tr>
<tr><td>1</td><td>0</td><td>2</td></tr>
<tr><td>⋮</td><td>⋮</td><td>⋮</td><td>⋮</td><td>⋮</td><td>⋮</td><td>⋮</td><td>⋮</td><td>⋱</td><td>⋮</td><td>⋮</td><td>⋮</td><td>⋮</td><td>⋮</td></tr>
<tr><td rowspan="2">h</td><td>0</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">. . .</td><td>2</td><td>0</td><td rowspan="2">. . .</td><td rowspan="2">1</td><td rowspan="2">1</td></tr>
<tr><td>1</td><td>0</td><td>2</td></tr>
<tr><td>⋮</td><td>⋮</td><td>⋮</td><td>⋮</td><td>⋮</td><td>⋮</td><td>⋮</td><td>⋮</td><td>⋮</td><td>⋮</td><td>⋮</td><td>⋱</td><td>⋮</td><td>⋮</td></tr>
<tr><td rowspan="2">75</td><td>0</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">. . .</td><td rowspan="2">1</td><td rowspan="2">1</td><td rowspan="2">. . .</td><td>2</td><td>0</td></tr>
<tr><td>1</td><td>0</td><td>2</td></tr>
</table>

For example, sampling Zone 1 yields two sets of replicate sampling weights, hence the two columns for Zone 1. The first set has doubled sampling weights ($k_{11} = 2$) for the students in the first school ($u_{11} = 0$) of Zone 1, zeroed sampling weights ($k_{12} = 0$) for the students in the second school ($u_{12} = 1$) of Zone 1, and unchanged sampling weights ($k_{hi} = 1$) for all students in the other sampling zones, e.g., Zones 2 through 75. This is shown in the first Zone 1 column. The second set of replicate sampling weights (shown in the second Zone 1 column) has zeroed sampling weights ($k_{11} = 0$) for the students in the first school ($u_{11} = 0$) of Zone 1, doubled sampling weights ($k_{12} = 2$) for the students in the second school ($u_{12} = 1$) of Zone 1, and unchanged sampling weights ($k_{hi} = 1$) for all students in the other sampling zones.

The process is repeated across all 75 possible sampling zones, generating 150 sets of replicate sampling weights. The replicate sampling weights are then used to estimate any statistic of interest 150 times. The variation across these 150 jackknife estimates determines the sampling variance.

Given a statistic to be computed from a national sample, the formula used to estimate the sampling variance of that statistic, based on the TIMSS JRR algorithm, is given by the following equation:

$$Var_{jrr}(t_0) = \frac{1}{2}\sum_{h=1}^{75}\sum_{i=1}^{2}(t_{hi} - t_0)^2 \qquad (14.3)$$

where the term t_0 denotes the statistic of interest estimated with the overall student sampling weights W_{0j} and the term t_{hi} denotes the same statistic computed using the set of replicate sampling weights W_{hij} obtained from sampling zone h (h=1,...,75), where the i^{th} school (1st or 2nd) in the zone is included and the other removed. , where the school (1st or 2nd) in the zone is included and the other removed. Efron (1982) provides a proof of why the variance can be calculated based on these squared deviations of the t_{hi} from the total sample statistics in jackknife based resampling schemes.

The sampling variance estimated with the TIMSS JRR method properly accounts for the variation arising from having sampled students using the TIMSS 2019 multi-stage stratified cluster sample design. Its square root is used as the standard error for any statistic derived from variables other than plausible values. Examples of such statistics include the mean age of students, the percentage of students with at least one parent with a university degree, and other variables that can be assessed objectively and likely only minimally affected by response variability.

Estimating Imputation Variance

For variables other than plausible values, standard errors were the result solely of sampling variation, and were computed using the JRR technique. However, the situation with achievement estimates is more complex. Achievement estimates are based on observations of how students perform on a subset of the TIMSS 2019 items. As described in the *TIMSS 2019 Assessment Frameworks*, the TIMSS 2019 item pool was far too extensive to be administered in its entirety to any one student, and a matrix-sampling assessment design was adopted whereby each student was given a single test booklet containing only a part of the entire assessment. The results from all students and booklets were then analyzed using item response theory to provide estimates of achievement on the TIMSS 2019 scale. Any estimate of achievement based on a set of observed variables is affected by measurement error. In order to generalize to the full assessment, an imputation (Rubin, 1987) model that incorporates performance on TIMSS 2019 of each student as well as information about similarities between students was applied. This imputation model is a latent regression model described in Chapter 11 of this volume and was used to derive estimates of student performance (plausible values). Student proficiency estimates incorporate uncertainty that can be quantified through measurement error and variability due to the latent regression. TIMSS 2019 followed the customary procedure of generating five imputations, or plausible values, for each student and using the variability among them as a measure of that uncertainty, known as the imputation variance.

The general procedure for estimating the imputation variance when analyzing student achievement data follows the basic principle of performing any statistical analysis five times—once for each set of plausible values—and aggregating the five sets of results (Mislevy, Beaton, Kaplan, & Sheehan, 1992). Thus, for any given achievement-based statistic t, estimating that statistic from each plausible value yields five estimates t_m, $m = 1, \ldots, 5$, all of them computed using the overall student sampling weights W_{0j}. The final estimate of that statistic, t_0, is the average of these five estimates:

$$t_0 = \frac{1}{5}\sum_{m=1}^{5} t_m \tag{14.4}$$

The imputation variance of the statistic t_0 is simply the variance of the five results from the plausible values, computed as follows:

$$Var_{imp}(t_0) = \frac{6}{5}\sum_{m=1}^{5}\frac{(t_m - t_0)^2}{4} \tag{14.5}$$

where the factor $\frac{6}{5}$ is a correction factor required by the multiple imputation methodology (Rubin, 1987). This imputation variance is then added to the sampling variance to produce the total variance estimate of the statistic t_0, as follows:

$$Var_{tot}(t_0) = Var_{jrr}(t_0) + Var_{imp}(t_0) \tag{14.6}$$

The sampling variance $Var_{jrr}(t_0)$ in this context is the average of the sampling variances from the five plausible values $Var_{jrr}(t_m)$ $m = 1, \ldots, 5$, as follows:

$$Var_{jrr}(t_0) = \frac{1}{5}\sum_{m=1}^{5} Var_{jrr}(t_m) \tag{14.7}$$

where

$$Var_{jrr}(t_m) = \frac{1}{2}\sum_{h=1}^{75}\sum_{i=1}^{2}(t_{mhi} - t_m)^2 \tag{14.8}$$

and t_{mhi} is the appropriate JRR estimate for plausible value and computed using the set of replicate sampling weights of sampling zone h where school i is included. The square root of the total variance is then the proper standard error for any statistic based on plausible values, such as the average TIMSS mathematics achievement for girls, or the percentage of students at or above the TIMSS Advanced International Benchmark of mathematics achievement.

Appendices 14A through 14D provide details on the jackknife sampling variance, the imputation variance, the total variance, and the overall standard error for each country's mean proficiency estimates in mathematics at the fourth grade, science at the fourth grade, mathematics at the eighth grade, and science at the eighth grade, respectively, and including the content and cognitive domains. Appendices 14E and 14F provide the same details for the bridge samples, limited to overall mathematics and science.[5]

5 Information on the bridge samples is available in Chapter 3.

Estimating Standard Errors for International Averages

Some exhibits in the TIMSS 2019 international report include international averages and their standard errors. For example, Exhibit 1.5 of *TIMSS 2019 International Results in Mathematics and Science* reports the international average for the percentages of girls and boys and their fourth grade mathematics achievement. International averages are computed using the data from the participating countries included in the main table of an exhibit. Results from the benchmarking participants are not included in the estimation of international averages.

For any given statistic t_0, its international average is given by

$$t_{intl} = \frac{1}{N}\sum_{k=1}^{N} t_{0k} \tag{14.9}$$

where N is the number of countries contributing to the international average and t_{0k} is the estimate of our statistic of interest for country k.

The total variance of the international average t_{intl} is given by

$$Var_{tot}(t_{intl}) = \frac{1}{N^2}\sum_{k=1}^{N} Var_{tot}(t_{0k}) \tag{14.10}$$

where $Var_{tot}(t_{0k})$ is the total variance of our statistic of interest for country k. For statistics based on plausible values, the total variance includes the sampling variance and the imputation variance, as given in equation (14.6) above. For statistics not based on plausible values, such as percentages, the total variance is based entirely on the sampling variance, as shown in equation (14.3) above. The standard error of the international average is the square root of the total variance.

Estimating Standard Errors for Comparing Independent Results

Standard errors, along with providing a measure of uncertainty for TIMSS results, also serve to perform statistical test of significance when comparing two or more results. A basic objective of TIMSS is to provide fair and accurate comparisons of student achievement across TIMSS assessment cycles. Exhibit 1.4 in the *TIMSS 2019 International Results in Mathematics and Science* report is one such example, showing fourth grade mathematics trend comparisons for the TIMSS 2019 countries across the TIMSS assessment cycles. The reports also include comparisons of results across the participating countries. Exhibit 1.2 shows all pairwise country comparisons for fourth grade mathematics achievement. All of these comparisons and their statistical significance tests require the computation of a standard error for a difference between two comparable results.

TIMSS 2019 results were reported by way of a statistic such as a mean or percentage, and each statistic is accompanied by its standard error, computed using either equation (14.3) or equation (14.6), as appropriate. When comparing results, either between assessment cycles or between countries, it is necessary to compute the standard error of the difference between two results. Because national samples are drawn independently of each other within an assessment cycle, as well as between assessment cycles, computing the standard error of a difference is straightforward.

When computing the difference between two TIMSS results t_A and t_B on the same TIMSS scale, which could be comparing the science achievement of countries *A* and *B*, or the science achievement of a country between assessment cycles *A* and *B*, the standard error of that difference is given by

$$SE(t_A - t_B) = \sqrt{Var_{tot}(t_A) + Var_{tot}(t_B)} \tag{14.11}$$

or, more simply

$$SE(t_A - t_B) = \sqrt{SE(t_A)^2 + SE(t_B)^2} \tag{14.12}$$

which can be stated as follows: the standard error of the difference between two independent results is the square root of the sum of their respective squared standard errors.

Estimating Standard Errors for Comparing Dependent Results

In the context of TIMSS, dependent results are statistics derived from the same national, or benchmarking, sample. The achievement difference between girls and boys, as shown in Exhibit 1.5 in the *TIMSS 2019 International Results in Mathematics and Science* report, is an example of two dependent results and their difference. This dependence occurs because girls and boys are selected from the same sample of classrooms and schools. Girls and boys from the same school tend to perform more similar as compared to subgroups selected from different schools, thus inducing a correlation that needs to be accounted in the computation of the standard error of their performance difference.

The difference between two statistics is itself a statistic. With this in mind, the standard error of any difference between two dependent results is computed in the same way as any other statistic, as was described earlier. The 150 sets of replicate weights produce 150 replicate estimates of the difference of interest and equations (14.3) and (14.6) apply.

Estimating Standard Errors for Comparing Against International Average

In TIMSS 2019 international report exhibits showing international averages, it can be of interest to compare a country's results to its corresponding international average. Exhibit 1.10.1 is one such example, showing percent correct statistics for a fourth grade mathematics item anchored at the TIMSS Low International Benchmark, along with a statistical significance test for comparisons of national results with the international average.

When comparing a country's result with the international average, TIMSS accounts for the fact that the country contributed to the international standard error. To correct for this contribution, the standard error of the difference needs to be adjusted. The total variance of the difference $t_k - t_{intl}$, comparing country k to the international average for a statistic t, is given by

$$Var_{tot}(t_k - t_{intl}) = Var_{tot}(t_{intl}) + \frac{(N-1)^2 - 1}{N^2} Var_{tot}(t_k) \tag{14.13}$$

where N is the number of countries contributing to the international average, $Var_{tot}(t_{intl})$ is the total variance of the international average as computed by equation (14.10), and $Var_{tot}(t_k)$ is the total variance for country k as computed by equation (14.6).

Equation (14.13) can be simplified and expressed in terms of standard errors as follows

$$SE(t_k - t_{intl}) = \sqrt{SE(t_{intl})^2 + \frac{N-2}{N} SE(t_k)^2} \tag{14.14}$$

where $SE(t_{intl})$ is the standard error of the international average and $SE(t_k)$ is the standard error for country k.

References

Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans*. Philadelphia, PA.: Society for Industrial and Applied Mathematics (SIAM).

Johnson, E. G., & Rust, K. F. (1992). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics, 17*(2), 175-190.

Mislevy, R. J., Beaton, A., Kaplan, B. A., & Sheehan, K. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, *29*(2), 133-161.

Quenouille, M. H. (1949). Problems in plane sampling. *The Annals of Mathematical Statistics, 20*(3), 355–375. doi:10.1214/aoms/1177729989.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons.

Tukey, J. W. (1958). Bias and confidence in not quite large samples. *The Annals of Mathematical Statistics, 29*(2), 614. doi:10.1214/aoms/1177706647

Wolter, K. M. (1985). *Introduction to variance estimation*. New York: Springer-Verlag.

Appendix 14A: Summary Statistics and Standard Errors for Proficiency in Grade 4 Mathematics

Summary Statistics and Standard Errors for Proficiency in Overall Mathematics—Grade 4

Country	Sample Size	Overall Mathematics				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Albania	4426	494.017	10.964	0.440	11.404	3.377
Armenia	5399	498.185	6.056	0.384	6.440	2.538
Australia	5890	515.880	7.397	0.309	7.706	2.776
Austria	5097	539.219	3.807	0.261	4.069	2.017
Azerbaijan	5245	515.455	6.950	0.492	7.443	2.728
Bahrain	5762	479.853	5.609	1.131	6.740	2.596
Belgium (Flemish)	4655	532.443	3.072	0.567	3.638	1.907
Bosnia and Herzegovina	5617	451.682	5.626	0.180	5.806	2.410
Bulgaria	4268	515.019	18.121	0.268	18.389	4.288
Canada	15572	511.564	3.135	0.313	3.448	1.857
Chile	4773	440.972	6.526	0.860	7.386	2.718
Chinese Taipei	4295	599.240	2.117	1.580	3.697	1.923
Croatia	4335	509.498	4.194	0.463	4.657	2.158
Cyprus	4062	532.094	7.948	0.221	8.169	2.858
Czech Republic	5357	532.975	5.863	0.486	6.348	2.520
Denmark	3692	524.542	3.278	0.387	3.666	1.915
England	3871	555.843	8.661	0.207	8.868	2.978
Finland	5394	532.071	4.911	0.532	5.443	2.333
France	4792	484.811	8.480	0.594	9.074	3.012
Georgia	4312	481.819	12.713	0.710	13.423	3.664
Germany	3932	520.981	4.777	0.403	5.180	2.276
Hong Kong SAR	3386	601.622	9.827	1.128	10.955	3.310
Hungary	5227	523.431	6.526	0.468	6.994	2.645
Iran, Islamic Rep. of	6010	443.035	13.526	1.343	14.869	3.856
Ireland	4582	548.465	5.998	0.173	6.172	2.484
Italy	4268	514.923	5.811	0.130	5.941	2.437
Japan	4196	592.957	2.914	0.151	3.065	1.751
Kazakhstan	4791	512.089	5.605	0.646	6.251	2.500
Korea, Rep. of	4448	599.608	4.762	0.212	4.974	2.230
Kosovo	4496	444.354	8.370	0.563	8.933	2.989
Kuwait	4437	383.318	21.475	1.009	22.483	4.742
Latvia	4481	546.133	6.301	0.391	6.693	2.587
Lithuania	4265	542.131	6.962	0.687	7.650	2.766
Malta	4152	509.125	1.205	0.858	2.063	1.436
Montenegro	5076	452.780	3.486	0.329	3.815	1.953
Morocco	7723	383.388	17.893	0.578	18.471	4.298
Netherlands	3829	537.509	4.265	0.424	4.689	2.165
New Zealand	5019	487.192	5.330	1.462	6.793	2.606
North Macedonia	3270	471.720	27.385	0.675	28.060	5.297
Northern Ireland	3497	565.800	6.632	0.917	7.550	2.748
Norway (5)	4526	542.669	4.344	0.427	4.771	2.184
Oman	6814	430.874	13.191	0.724	13.915	3.730
Pakistan	3980	327.691	140.907	2.041	142.948	11.956
Philippines	5515	296.675	37.419	3.099	40.517	6.365

Summary Statistics and Standard Errors for Proficiency in Overall Mathematics—Grade 4 (continued)

Country	Sample Size	Overall Mathematics				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Poland	4882	520.140	6.809	0.357	7.167	2.677
Portugal	4914	525.091	6.299	0.472	6.771	2.602
Qatar	5646	449.412	10.287	1.121	11.408	3.378
Russian Federation	4596	566.947	10.762	0.391	11.153	3.340
Saudi Arabia	5453	398.093	12.000	0.620	12.620	3.552
Serbia	4380	507.881	9.790	0.382	10.172	3.189
Singapore	6839	625.429	14.165	0.823	14.989	3.872
Slovak Republic	4861	509.841	10.957	0.959	11.917	3.452
South Africa (5)	11891	373.564	11.769	0.942	12.711	3.565
Spain	10945	502.472	3.404	1.152	4.556	2.134
Sweden	4532	521.229	7.505	0.272	7.778	2.789
Turkey (5)	4599	522.856	19.333	0.446	19.779	4.447
United Arab Emirates	29511	481.387	2.147	0.761	2.908	1.705
United States	10028	534.732	6.192	0.258	6.449	2.540
Benchmarking Participants						
Ontario, Canada	4358	511.715	10.510	0.447	10.957	3.310
Quebec, Canada	4383	532.133	4.595	0.636	5.231	2.287
Moscow City, Russian Fed.	4392	592.772	4.721	0.183	4.903	2.214
Madrid, Spain	3878	518.255	4.406	0.216	4.623	2.150
Abu Dhabi, UAE	10324	440.536	3.334	1.672	5.006	2.237
Dubai, UAE	8299	544.050	2.252	0.455	2.707	1.645

Summary Statistics and Standard Errors for Proficiency in Number—Grade 4

Country	Sample Size	Number				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Albania	4426	495.037	12.316	0.774	13.089	3.618
Armenia	5399	517.920	4.452	0.760	5.212	2.283
Australia	5890	506.179	8.708	0.702	9.410	3.068
Austria	5097	541.815	3.008	0.627	3.635	1.907
Azerbaijan	5245	525.634	6.492	0.697	7.189	2.681
Bahrain	5762	478.348	6.162	0.349	6.511	2.552
Belgium (Flemish)	4655	526.283	3.066	0.870	3.935	1.984
Bosnia and Herzegovina	5617	458.811	5.057	0.243	5.301	2.302
Bulgaria	4268	521.387	14.965	1.096	16.060	4.008
Canada	15582	505.263	3.266	1.209	4.475	2.115
Chile	4775	438.254	6.533	2.477	9.010	3.002
Chinese Taipei	4295	599.412	2.072	0.692	2.764	1.663
Croatia	4335	511.817	3.286	0.145	3.431	1.852
Cyprus	4062	537.968	7.600	0.447	8.047	2.837
Czech Republic	5358	535.705	5.121	0.873	5.995	2.448
Denmark	3693	517.765	4.080	0.216	4.296	2.073
England	3872	558.990	10.487	0.510	10.997	3.316
Finland	5397	527.881	4.710	0.646	5.355	2.314
France	4792	480.262	9.222	1.209	10.431	3.230
Georgia	4316	500.924	10.726	1.968	12.694	3.563
Germany	3933	517.367	4.227	0.390	4.617	2.149
Hong Kong SAR	3386	597.913	11.105	1.715	12.821	3.581
Hungary	5227	530.588	6.008	0.617	6.625	2.574
Iran, Islamic Rep. of	6010	446.090	14.802	0.944	15.746	3.968
Ireland	4582	554.615	6.340	0.906	7.246	2.692
Italy	4269	521.999	5.393	0.730	6.123	2.475
Japan	4196	585.911	2.711	0.633	3.344	1.829
Kazakhstan	4791	522.658	4.882	0.761	5.643	2.375
Korea, Rep. of	4448	593.447	5.589	0.152	5.742	2.396
Kosovo	4496	447.397	7.405	0.388	7.793	2.792
Kuwait	4437	–	–	–	–	–
Latvia	4481	546.956	6.553	0.052	6.604	2.570
Lithuania	4265	538.090	7.405	0.317	7.722	2.779
Malta	4154	511.977	1.220	1.107	2.326	1.525
Montenegro	5076	453.534	3.182	1.709	4.891	2.211
Morocco	7723	383.032	17.238	2.026	19.263	4.389
Netherlands	3831	532.919	3.937	0.801	4.738	2.177
New Zealand	5019	478.489	6.662	1.526	8.188	2.861
North Macedonia	3270	472.331	25.779	1.016	26.794	5.176
Northern Ireland	3497	572.448	8.218	1.439	9.657	3.107
Norway (5)	4527	539.844	4.074	0.085	4.159	2.039
Oman	6814	423.562	14.899	0.914	15.814	3.977
Pakistan	3980	351.471	115.358	3.999	119.356	10.925
Philippines	5515	308.074	33.043	3.841	36.884	6.073
Poland	4882	512.658	7.242	0.742	7.984	2.826
Portugal	4915	524.243	7.023	1.600	8.623	2.936
Qatar	5646	454.894	10.939	0.465	11.404	3.377
Russian Federation	4596	567.443	10.191	1.229	11.420	3.379

Summary Statistics and Standard Errors for Proficiency in Number—Grade 4 (continued)

Country	Sample Size	Number: Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Saudi Arabia	5453	–	–	–	–	–
Serbia	4380	517.752	7.751	0.536	8.287	2.879
Singapore	6839	635.167	15.775	0.284	16.059	4.007
Slovak Republic	4862	512.314	10.238	2.434	12.672	3.560
South Africa (5)	11891	370.249	11.994	1.250	13.245	3.639
Spain	10946	506.240	3.120	0.621	3.741	1.934
Sweden	4535	516.995	7.230	1.056	8.285	2.878
Turkey (5)	4599	525.396	22.198	0.345	22.543	4.748
United Arab Emirates	29515	485.159	2.250	0.481	2.731	1.652
United States	10029	542.260	6.521	0.292	6.812	2.610
Benchmarking Participants						
Ontario, Canada	4360	501.494	10.897	1.903	12.800	3.578
Quebec, Canada	4384	529.549	4.826	0.868	5.694	2.386
Moscow City, Russian Fed.	4392	590.825	4.572	0.396	4.968	2.229
Madrid, Spain	3879	524.494	4.725	0.206	4.931	2.221
Abu Dhabi, UAE	10328	442.787	3.240	0.832	4.072	2.018
Dubai, UAE	8299	547.844	2.542	0.485	3.028	1.740

Summary Statistics and Standard Errors for Proficiency in Measurement and Geometry—Grade 4

Country	Sample Size	Measurement and Geometry				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Albania	4426	495.776	10.239	1.329	11.568	3.401
Armenia	5399	490.284	8.306	0.952	9.259	3.043
Australia	5890	515.924	8.113	2.516	10.629	3.260
Austria	5097	541.628	4.420	1.303	5.723	2.392
Azerbaijan	5245	502.585	8.528	1.623	10.151	3.186
Bahrain	5762	473.914	6.582	0.166	6.748	2.598
Belgium (Flemish)	4655	550.815	3.208	0.899	4.107	2.027
Bosnia and Herzegovina	5617	457.795	6.477	1.983	8.460	2.909
Bulgaria	4268	521.531	20.206	4.090	24.295	4.929
Canada	15582	510.605	3.081	0.047	3.128	1.769
Chile	4775	438.502	5.106	1.247	6.353	2.521
Chinese Taipei	4295	607.480	2.502	0.677	3.179	1.783
Croatia	4335	517.771	5.301	1.745	7.046	2.654
Cyprus	4062	525.855	7.670	1.893	9.563	3.092
Czech Republic	5358	539.826	6.182	2.194	8.376	2.894
Denmark	3693	536.224	3.116	2.527	5.642	2.375
England	3872	545.075	9.475	1.373	10.849	3.294
Finland	5397	538.434	5.356	3.534	8.890	2.982
France	4792	497.646	8.568	2.319	10.887	3.300
Georgia	4316	469.507	15.030	1.642	16.672	4.083
Germany	3933	531.170	5.153	1.459	6.612	2.571
Hong Kong SAR	3386	607.990	8.797	0.751	9.549	3.090
Hungary	5227	519.326	8.306	2.793	11.099	3.332
Iran, Islamic Rep. of	6010	445.206	12.254	0.373	12.627	3.553
Ireland	4582	540.368	6.293	0.753	7.046	2.654
Italy	4269	510.261	7.611	2.340	9.951	3.155
Japan	4196	601.117	3.823	3.414	7.238	2.690
Kazakhstan	4791	513.122	6.556	1.030	7.586	2.754
Korea, Rep. of	4448	607.585	4.979	1.603	6.582	2.566
Kosovo	4496	450.249	9.239	1.621	10.861	3.296
Kuwait	4437	–	–	–	–	–
Latvia	4481	547.744	7.137	0.483	7.621	2.761
Lithuania	4265	543.118	7.024	1.766	8.790	2.965
Malta	4154	496.988	1.180	2.084	3.264	1.807
Montenegro	5076	459.380	3.790	0.649	4.438	2.107
Morocco	7723	385.647	19.198	0.911	20.109	4.484
Netherlands	3831	537.123	3.983	0.667	4.649	2.156
New Zealand	5019	481.268	4.792	2.465	7.256	2.694
North Macedonia	3270	475.047	29.272	4.244	33.516	5.789
Northern Ireland	3497	555.761	7.610	1.459	9.068	3.011
Norway (5)	4527	546.419	5.130	2.512	7.642	2.764
Oman	6814	428.651	16.539	0.862	17.401	4.171
Pakistan	3980	286.038	191.857	7.959	199.815	14.136
Philippines	5515	259.291	45.122	4.782	49.904	7.064
Poland	4882	529.183	6.871	0.443	7.313	2.704
Portugal	4915	520.073	7.092	1.138	8.230	2.869
Qatar	5646	434.405	8.803	2.455	11.259	3.355
Russian Federation	4596	570.651	12.611	0.900	13.511	3.676

Summary Statistics and Standard Errors for Proficiency in Measurement and Geometry—Grade 4 (continued)

Country	Sample Size	Measurement and Geometry				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Saudi Arabia	5453	–	–	–	–	–
Serbia	4380	499.171	12.382	1.408	13.790	3.713
Singapore	6839	620.252	14.310	0.567	14.877	3.857
Slovak Republic	4862	505.503	10.877	2.544	13.421	3.663
South Africa (5)	11891	362.254	12.841	1.042	13.884	3.726
Spain	10946	493.645	4.008	0.899	4.906	2.215
Sweden	4535	521.350	9.811	1.565	11.376	3.373
Turkey (5)	4599	526.938	16.622	2.777	19.399	4.404
United Arab Emirates	29515	471.672	2.400	0.881	3.280	1.811
United States	10029	519.502	6.697	0.155	6.852	2.618
Benchmarking Participants						
Ontario, Canada	4360	516.387	9.539	0.805	10.344	3.216
Quebec, Canada	4384	531.905	4.690	2.077	6.767	2.601
Moscow City, Russian Fed.	4392	590.479	5.506	0.394	5.899	2.429
Madrid, Spain	3879	508.131	5.073	6.155	11.227	3.351
Abu Dhabi, UAE	10328	429.207	3.481	0.882	4.363	2.089
Dubai, UAE	8299	535.274	2.822	1.722	4.545	2.132

Summary Statistics and Standard Errors for Proficiency in Data—Grade 4

Country	Sample Size	Data				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Albania	4426	489.603	12.580	3.084	15.664	3.958
Armenia	5399	446.495	14.881	2.513	17.394	4.171
Australia	5890	534.158	8.039	3.826	11.865	3.445
Austria	5097	528.303	6.040	1.338	7.377	2.716
Azerbaijan	5245	503.973	8.201	0.995	9.196	3.032
Bahrain	5762	483.168	6.779	3.825	10.603	3.256
Belgium (Flemish)	4655	526.539	4.283	0.561	4.844	2.201
Bosnia and Herzegovina	5617	412.975	8.287	6.037	14.324	3.785
Bulgaria	4268	489.750	28.516	3.143	31.659	5.627
Canada	15582	522.589	3.937	1.613	5.551	2.356
Chile	4775	438.175	7.576	1.537	9.113	3.019
Chinese Taipei	4295	590.000	2.463	3.069	5.533	2.352
Croatia	4335	494.031	5.375	1.832	7.207	2.685
Cyprus	4062	523.541	11.570	0.078	11.648	3.413
Czech Republic	5358	517.888	7.235	1.145	8.379	2.895
Denmark	3693	525.274	4.783	0.652	5.435	2.331
England	3872	564.772	7.085	2.409	9.494	3.081
Finland	5397	533.949	6.282	1.324	7.606	2.758
France	4792	475.921	9.368	1.968	11.336	3.367
Georgia	4316	444.167	19.010	2.027	21.037	4.587
Germany	3933	514.969	7.740	1.565	9.305	3.050
Hong Kong SAR	3386	606.787	8.553	4.393	12.946	3.598
Hungary	5227	507.997	9.727	0.547	10.274	3.205
Iran, Islamic Rep. of	6010	424.447	12.946	1.439	14.385	3.793
Ireland	4582	542.599	8.254	0.938	9.192	3.032
Italy	4269	498.162	8.479	0.728	9.208	3.034
Japan	4196	605.568	3.826	0.433	4.259	2.064
Kazakhstan	4791	481.279	8.150	0.943	9.092	3.015
Korea, Rep. of	4448	602.348	4.966	1.453	6.418	2.533
Kosovo	4496	423.399	12.797	1.132	13.929	3.732
Kuwait	4437	–	–	–	–	–
Latvia	4481	542.028	7.480	3.073	10.553	3.249
Lithuania	4265	544.701	7.244	1.906	9.150	3.025
Malta	4154	511.745	1.769	1.415	3.184	1.784
Montenegro	5076	438.941	5.236	2.093	7.328	2.707
Morocco	7723	374.307	26.721	0.967	27.689	5.262
Netherlands	3831	549.356	6.651	2.214	8.864	2.977
New Zealand	5019	503.795	6.152	3.394	9.547	3.090
North Macedonia	3270	464.375	34.069	2.781	36.850	6.070
Northern Ireland	3497	564.088	5.388	0.743	6.131	2.476
Norway (5)	4527	546.747	7.074	3.198	10.272	3.205
Oman	6814	432.765	12.206	2.497	14.703	3.834
Pakistan	3980	277.861	206.922	2.972	209.894	14.488
Philippines	5515	290.984	45.650	5.163	50.813	7.128
Poland	4882	524.057	7.417	1.258	8.675	2.945
Portugal	4915	528.085	5.533	0.977	6.510	2.551
Qatar	5646	445.438	13.336	0.950	14.286	3.780
Russian Federation	4596	559.843	12.285	2.540	14.825	3.850

Summary Statistics and Standard Errors for Proficiency in Data—Grade 4 (continued)

Country	Sample Size	Data				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Saudi Arabia	5453	–	–	–	–	–
Serbia	4380	489.187	15.143	2.259	17.402	4.172
Singapore	6839	613.499	13.784	0.750	14.534	3.812
Slovak Republic	4862	505.622	15.045	2.092	17.137	4.140
South Africa (5)	11891	390.013	12.252	2.380	14.632	3.825
Spain	10946	499.410	4.318	2.348	6.665	2.582
Sweden	4535	527.046	10.149	1.890	12.039	3.470
Turkey (5)	4599	510.321	19.449	0.787	20.236	4.498
United Arab Emirates	29515	476.228	1.997	1.132	3.130	1.769
United States	10029	533.224	6.384	2.877	9.261	3.043
Benchmarking Participants						
Ontario, Canada	4360	527.205	13.685	1.917	15.603	3.950
Quebec, Canada	4384	534.805	5.964	3.392	9.356	3.059
Moscow City, Russian Fed.	4392	603.254	5.642	0.443	6.085	2.467
Madrid, Spain	3879	512.794	5.275	2.916	8.191	2.862
Abu Dhabi, UAE	10328	434.920	3.302	1.866	5.168	2.273
Dubai, UAE	8299	546.189	2.653	1.307	3.960	1.990

Summary Statistics and Standard Errors for Proficiency in Knowing in Mathematics—Grade 4

Country	Sample Size	Mathematics Knowing				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Albania	4426	492.008	12.230	1.733	13.963	3.737
Armenia	5399	496.887	6.012	1.058	7.070	2.659
Australia	5890	509.225	8.101	2.569	10.670	3.266
Austria	5097	540.214	3.265	0.612	3.877	1.969
Azerbaijan	5245	513.116	5.116	0.280	5.396	2.323
Bahrain	5762	477.568	6.100	1.156	7.256	2.694
Belgium (Flemish)	4655	546.158	3.938	2.051	5.989	2.447
Bosnia and Herzegovina	5617	444.171	5.257	1.908	7.166	2.677
Bulgaria	4268	510.965	15.592	1.144	16.736	4.091
Canada	15582	506.483	3.526	0.940	4.466	2.113
Chile	4775	427.279	6.096	0.831	6.927	2.632
Chinese Taipei	4295	621.564	2.272	1.380	3.651	1.911
Croatia	4335	507.643	3.934	0.868	4.802	2.191
Cyprus	4062	529.883	8.817	1.931	10.748	3.278
Czech Republic	5358	528.194	5.622	3.193	8.815	2.969
Denmark	3693	523.892	4.057	0.803	4.860	2.205
England	3872	562.728	10.076	0.545	10.621	3.259
Finland	5397	530.888	5.218	0.530	5.749	2.398
France	4792	487.518	9.102	2.051	11.152	3.340
Georgia	4316	473.424	13.333	1.550	14.883	3.858
Germany	3933	523.431	4.736	0.625	5.361	2.315
Hong Kong SAR	3386	599.917	8.586	0.562	9.148	3.025
Hungary	5227	524.599	6.495	0.454	6.948	2.636
Iran, Islamic Rep. of	6010	436.394	14.250	1.341	15.591	3.949
Ireland	4582	550.224	7.395	1.652	9.046	3.008
Italy	4269	514.972	6.325	2.730	9.055	3.009
Japan	4196	597.367	3.325	0.540	3.865	1.966
Kazakhstan	4791	509.915	4.589	0.745	5.334	2.310
Korea, Rep. of	4448	612.367	6.050	6.579	12.629	3.554
Kosovo	4496	444.770	9.651	0.494	10.145	3.185
Kuwait	4437	–	–	–	–	–
Latvia	4481	537.095	6.113	0.637	6.750	2.598
Lithuania	4265	535.443	6.091	1.802	7.893	2.809
Malta	4154	509.508	1.238	0.630	1.868	1.367
Montenegro	5076	444.654	3.261	1.249	4.510	2.124
Morocco	7723	379.300	19.089	0.364	19.454	4.411
Netherlands	3831	534.084	3.751	0.467	4.218	2.054
New Zealand	5019	475.997	6.090	1.065	7.155	2.675
North Macedonia	3270	469.801	28.765	3.154	31.919	5.650
Northern Ireland	3497	574.459	8.240	2.448	10.688	3.269
Norway (5)	4527	540.539	4.141	1.028	5.168	2.273
Oman	6814	423.923	16.661	2.574	19.235	4.386
Pakistan	3980	326.675	153.107	4.480	157.587	12.553
Philippines	5515	301.905	38.734	1.333	40.067	6.330
Poland	4882	509.333	6.648	0.665	7.314	2.704
Portugal	4915	523.035	6.398	1.684	8.081	2.843
Qatar	5646	447.336	12.633	0.248	12.881	3.589
Russian Federation	4596	554.537	8.452	0.304	8.757	2.959

Summary Statistics and Standard Errors for Proficiency in Knowing in Mathematics—Grade 4 (continued)

Country	Sample Size	Mathematics Knowing				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Saudi Arabia	5453	–	–	–	–	–
Serbia	4380	504.015	8.524	2.490	11.014	3.319
Singapore	6839	640.096	14.607	0.572	15.179	3.896
Slovak Republic	4862	501.590	10.324	0.583	10.906	3.302
South Africa (5)	11891	372.353	12.734	1.028	13.763	3.710
Spain	10946	499.496	3.767	1.974	5.741	2.396
Sweden	4535	515.376	7.412	2.289	9.701	3.115
Turkey (5)	4599	514.413	19.051	0.479	19.530	4.419
United Arab Emirates	29515	479.344	2.365	0.305	2.671	1.634
United States	10029	536.492	6.608	0.276	6.884	2.624
Benchmarking Participants						
Ontario, Canada	4360	503.914	11.510	2.407	13.917	3.731
Quebec, Canada	4384	535.016	5.116	2.128	7.243	2.691
Moscow City, Russian Fed.	4392	576.540	3.440	1.004	4.443	2.108
Madrid, Spain	3879	514.609	5.230	10.531	15.761	3.970
Abu Dhabi, UAE	10328	439.150	3.707	0.289	3.996	1.999
Dubai, UAE	8299	542.494	2.552	0.626	3.177	1.783

Summary Statistics and Standard Errors for Proficiency in Applying in Mathematics—Grade 4

Country	Sample Size	Mathematics Applying				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Albania	4426	497.712	10.503	0.649	11.152	3.339
Armenia	5399	501.345	6.687	1.964	8.651	2.941
Australia	5890	515.948	7.378	0.905	8.282	2.878
Austria	5097	537.632	3.843	0.365	4.208	2.051
Azerbaijan	5245	519.348	8.536	0.794	9.330	3.055
Bahrain	5762	479.466	6.657	0.205	6.861	2.619
Belgium (Flemish)	4655	526.453	3.080	0.748	3.828	1.957
Bosnia and Herzegovina	5617	451.822	6.086	2.439	8.525	2.920
Bulgaria	4268	518.288	19.395	1.120	20.515	4.529
Canada	15582	512.930	3.097	0.621	3.719	1.928
Chile	4775	445.626	6.786	2.321	9.107	3.018
Chinese Taipei	4295	600.223	2.124	0.173	2.297	1.516
Croatia	4335	509.240	3.923	1.250	5.173	2.274
Cyprus	4062	536.422	8.417	0.525	8.942	2.990
Czech Republic	5358	531.499	6.009	0.850	6.860	2.619
Denmark	3693	519.860	3.732	1.761	5.493	2.344
England	3872	553.177	8.442	2.201	10.643	3.262
Finland	5397	531.330	5.275	0.386	5.661	2.379
France	4792	481.687	8.915	0.645	9.560	3.092
Georgia	4316	489.562	11.944	1.356	13.300	3.647
Germany	3933	514.027	4.884	1.196	6.079	2.466
Hong Kong SAR	3386	606.477	9.781	1.329	11.110	3.333
Hungary	5227	521.226	7.109	0.976	8.084	2.843
Iran, Islamic Rep. of	6010	449.546	14.199	1.726	15.925	3.991
Ireland	4582	551.333	6.180	0.917	7.098	2.664
Italy	4269	517.381	6.215	0.788	7.003	2.646
Japan	4196	592.841	2.799	1.386	4.185	2.046
Kazakhstan	4791	514.005	6.603	0.619	7.222	2.687
Korea, Rep. of	4448	594.160	4.985	1.356	6.341	2.518
Kosovo	4496	445.293	8.054	0.894	8.948	2.991
Kuwait	4437	–	–	–	–	–
Latvia	4481	546.604	6.509	1.046	7.554	2.749
Lithuania	4265	546.946	7.435	0.022	7.457	2.731
Malta	4154	507.516	1.068	0.368	1.436	1.198
Montenegro	5076	453.817	3.806	0.691	4.497	2.121
Morocco	7723	387.008	17.215	3.429	20.643	4.544
Netherlands	3831	535.971	4.467	0.513	4.980	2.232
New Zealand	5019	487.375	5.498	0.503	6.001	2.450
North Macedonia	3270	476.657	25.566	1.332	26.898	5.186
Northern Ireland	3497	564.756	7.353	0.444	7.798	2.792
Norway (5)	4527	539.724	4.871	0.615	5.485	2.342
Oman	6814	434.099	12.421	0.173	12.594	3.549
Pakistan	3980	306.379	167.860	3.563	171.423	13.093
Philippines	5515	286.298	41.963	5.853	47.816	6.915
Poland	4882	521.408	7.172	0.542	7.713	2.777
Portugal	4915	528.083	6.178	0.559	6.737	2.596
Qatar	5646	453.181	9.907	1.337	11.244	3.353
Russian Federation	4596	570.548	11.325	1.314	12.639	3.555

Summary Statistics and Standard Errors for Proficiency in Applying in Mathematics—Grade 4 (continued)

Country	Sample Size	Mathematics Applying				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Saudi Arabia	5453	–	–	–	–	–
Serbia	4380	508.983	11.027	1.116	12.143	3.485
Singapore	6839	625.586	14.408	0.559	14.967	3.869
Slovak Republic	4862	508.027	10.737	1.074	11.811	3.437
South Africa (5)	11891	375.471	11.031	1.722	12.752	3.571
Spain	10946	505.530	3.171	0.465	3.637	1.907
Sweden	4535	517.812	7.540	0.313	7.852	2.802
Turkey (5)	4599	530.746	19.559	0.235	19.794	4.449
United Arab Emirates	29515	484.287	2.153	0.602	2.755	1.660
United States	10029	537.435	6.417	0.589	7.006	2.647
Benchmarking Participants						
Ontario, Canada	4360	513.734	10.374	1.062	11.436	3.382
Quebec, Canada	4384	533.351	4.554	0.867	5.422	2.328
Moscow City, Russian Fed.	4392	598.612	4.973	1.334	6.308	2.511
Madrid, Spain	3879	520.490	4.207	1.539	5.746	2.397
Abu Dhabi, UAE	10328	441.848	3.349	0.577	3.926	1.981
Dubai, UAE	8299	547.365	2.365	0.328	2.693	1.641

Summary Statistics and Standard Errors for Proficiency in Reasoning in Mathematics—Grade 4

Country	Sample Size	Mathematics Reasoning: Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Albania	4426	489.866	12.706	1.225	13.931	3.732
Armenia	5399	482.947	6.571	1.814	8.385	2.896
Australia	5890	522.334	7.181	1.955	9.136	3.023
Austria	5097	537.327	4.399	1.271	5.670	2.381
Azerbaijan	5245	506.384	8.801	1.102	9.904	3.147
Bahrain	5762	479.032	4.998	1.080	6.078	2.465
Belgium (Flemish)	4655	530.363	3.215	0.627	3.843	1.960
Bosnia and Herzegovina	5617	461.310	6.931	2.009	8.940	2.990
Bulgaria	4268	509.400	23.474	1.308	24.782	4.978
Canada	15582	513.184	2.821	1.323	4.144	2.036
Chile	4775	447.788	6.567	9.457	16.025	4.003
Chinese Taipei	4295	576.130	3.178	0.087	3.265	1.807
Croatia	4335	509.579	6.524	1.198	7.722	2.779
Cyprus	4062	526.492	8.086	0.102	8.188	2.861
Czech Republic	5358	541.354	6.771	0.911	7.683	2.772
Denmark	3693	534.908	3.519	1.364	4.882	2.210
England	3872	554.243	8.213	3.475	11.688	3.419
Finland	5397	535.274	4.889	1.304	6.193	2.489
France	4792	480.293	10.053	0.635	10.688	3.269
Georgia	4316	468.961	17.836	2.064	19.900	4.461
Germany	3933	531.209	5.146	2.480	7.626	2.761
Hong Kong SAR	3386	595.785	13.208	4.373	17.580	4.193
Hungary	5227	522.152	8.319	0.952	9.271	3.045
Iran, Islamic Rep. of	6010	426.403	13.031	5.822	18.854	4.342
Ireland	4582	541.952	5.564	0.606	6.170	2.484
Italy	4269	503.988	5.031	3.154	8.186	2.861
Japan	4196	588.749	4.091	0.845	4.937	2.222
Kazakhstan	4791	507.328	6.371	0.665	7.035	2.652
Korea, Rep. of	4448	596.499	5.798	2.513	8.311	2.883
Kosovo	4496	441.253	8.773	1.752	10.525	3.244
Kuwait	4437	–	–	–	–	–
Latvia	4481	554.327	8.639	0.558	9.198	3.033
Lithuania	4265	533.508	8.818	1.957	10.775	3.282
Malta	4154	507.999	1.284	0.631	1.916	1.384
Montenegro	5076	462.869	4.294	2.896	7.189	2.681
Morocco	7723	379.809	21.977	5.363	27.340	5.229
Netherlands	3831	545.763	5.618	3.026	8.644	2.940
New Zealand	5019	501.111	5.137	2.072	7.210	2.685
North Macedonia	3270	470.146	25.049	7.096	32.145	5.670
Northern Ireland	3497	558.367	5.297	3.076	8.373	2.894
Norway (5)	4527	550.732	4.681	3.590	8.271	2.876
Oman	6814	424.489	13.180	0.679	13.859	3.723
Pakistan	3980	354.332	84.595	2.266	86.861	9.320
Philippines	5515	271.924	41.199	2.558	43.758	6.615
Poland	4882	527.177	6.716	0.927	7.643	2.765
Portugal	4915	519.489	6.618	1.702	8.320	2.884
Qatar	5646	439.848	10.867	1.109	11.976	3.461
Russian Federation	4596	572.825	12.976	0.341	13.317	3.649

Summary Statistics and Standard Errors for Proficiency in Reasoning in Mathematics—Grade 4 (continued)

Country	Sample Size	Mathematics Reasoning				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Saudi Arabia	5453	–	–	–	–	–
Serbia	4380	503.200	13.045	0.865	13.910	3.730
Singapore	6839	613.975	15.570	0.650	16.220	4.027
Slovak Republic	4862	521.638	11.487	0.675	12.162	3.487
South Africa (5)	11891	370.204	11.837	2.304	14.141	3.760
Spain	10946	496.836	3.676	0.342	4.018	2.005
Sweden	4535	535.770	7.667	0.830	8.497	2.915
Turkey (5)	4599	508.961	22.865	3.009	25.874	5.087
United Arab Emirates	29515	474.278	2.188	0.784	2.972	1.724
United States	10029	523.744	5.847	0.534	6.381	2.526
Benchmarking Participants						
Ontario, Canada	4360	515.653	9.052	3.423	12.475	3.532
Quebec, Canada	4384	523.872	4.827	2.926	7.752	2.784
Moscow City, Russian Fed.	4392	601.547	6.724	1.067	7.791	2.791
Madrid, Spain	3879	514.454	4.514	2.627	7.141	2.672
Abu Dhabi, UAE	10328	434.535	3.313	2.337	5.650	2.377
Dubai, UAE	8299	537.919	2.417	0.880	3.297	1.816

Appendix 14B: Summary Statistics and Standard Errors for Proficiency in Grade 4 Science

Summary Statistics and Standard Errors for Proficiency in Overall Science—Grade 4

Country	Sample Size	Overall Science				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Albania	4426	489.480	11.533	0.908	12.441	3.527
Armenia	5399	466.432	8.525	3.268	11.793	3.434
Australia	5890	532.575	5.420	0.460	5.881	2.425
Austria	5097	522.061	4.281	2.351	6.632	2.575
Azerbaijan	5245	426.735	9.989	0.817	10.807	3.287
Bahrain	5762	492.542	10.410	1.214	11.624	3.409
Belgium (Flemish)	4655	501.017	3.383	1.052	4.435	2.106
Bosnia and Herzegovina	5617	458.594	6.755	1.928	8.682	2.947
Bulgaria	4268	521.302	23.747	0.593	24.340	4.934
Canada	15577	523.033	2.235	1.304	3.539	1.881
Chile	4773	469.116	5.797	0.820	6.618	2.572
Chinese Taipei	4295	558.050	2.031	1.067	3.098	1.760
Croatia	4335	523.864	2.843	1.810	4.653	2.157
Cyprus	4062	511.419	6.960	2.263	9.223	3.037
Czech Republic	5358	533.720	3.991	2.707	6.699	2.588
Denmark	3692	522.163	3.415	2.197	5.612	2.369
England	3871	537.026	6.153	1.023	7.176	2.679
Finland	5395	554.561	4.877	1.746	6.623	2.573
France	4791	487.728	8.011	0.783	8.794	2.966
Georgia	4313	454.213	12.496	2.990	15.486	3.935
Germany	3933	518.346	4.696	0.190	4.886	2.210
Hong Kong SAR	3386	531.250	9.782	1.427	11.210	3.348
Hungary	5227	529.438	6.151	0.978	7.129	2.670
Iran, Islamic Rep. of	6010	440.828	16.546	0.419	16.964	4.119
Ireland	4582	527.970	9.038	0.968	10.006	3.163
Italy	4269	509.728	6.219	2.829	9.048	3.008
Japan	4196	561.659	2.705	0.432	3.137	1.771
Kazakhstan	4791	494.155	8.139	1.260	9.399	3.066
Korea, Rep. of	4448	587.607	3.775	0.745	4.520	2.126
Kosovo	4496	413.059	11.384	2.272	13.656	3.695
Kuwait	4437	392.295	32.017	5.232	37.250	6.103
Latvia	4481	541.858	5.508	0.179	5.687	2.385
Lithuania	4265	538.079	5.350	0.987	6.337	2.517
Malta	4153	495.791	1.379	0.229	1.608	1.268
Montenegro	5076	453.321	4.138	2.071	6.208	2.492
Morocco	7723	374.066	29.024	4.927	33.951	5.827
Netherlands	3829	518.471	6.588	1.821	8.410	2.900
New Zealand	5019	502.551	4.061	1.245	5.305	2.303
North Macedonia	3270	426.043	35.158	3.731	38.889	6.236
Northern Ireland	3497	518.491	3.486	1.809	5.295	2.301
Norway (5)	4526	539.402	4.314	0.494	4.808	2.193
Oman	6814	434.944	14.574	1.977	16.551	4.068
Pakistan	3980	290.097	171.577	9.136	180.713	13.443
Philippines	5515	249.018	48.177	8.390	56.567	7.521

Summary Statistics and Standard Errors for Proficiency in Overall Science—Grade 4 (continued)

Country	Sample Size	Overall Science				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Poland	4882	530.833	5.574	1.119	6.693	2.587
Portugal	4915	503.815	3.840	2.666	6.507	2.551
Qatar	5645	449.472	14.335	1.183	15.518	3.939
Russian Federation	4596	567.255	8.334	0.769	9.103	3.017
Saudi Arabia	5453	402.237	12.027	4.535	16.562	4.070
Serbia	4380	516.907	10.649	1.604	12.252	3.500
Singapore	6837	594.529	11.309	0.250	11.558	3.400
Slovak Republic	4862	520.732	12.738	0.955	13.693	3.700
South Africa (5)	11891	324.234	21.945	2.378	24.323	4.932
Spain	10945	511.282	2.891	1.152	4.043	2.011
Sweden	4534	537.233	8.897	1.999	10.897	3.301
Turkey (5)	4599	526.355	16.508	1.317	17.825	4.222
United Arab Emirates	29508	472.544	2.748	1.464	4.212	2.052
United States	10028	538.643	5.876	1.640	7.516	2.742
Benchmarking Participants						
Ontario, Canada	4359	523.918	7.784	2.160	9.944	3.153
Quebec, Canada	4384	521.962	4.190	2.215	6.405	2.531
Moscow City, Russian Fed.	4392	594.862	4.178	0.812	4.989	2.234
Madrid, Spain	3879	522.822	3.446	0.519	3.965	1.991
Abu Dhabi, UAE	10324	417.825	4.409	3.156	7.565	2.751
Dubai, UAE	8299	544.504	2.065	0.849	2.914	1.707

Summary Statistics and Standard Errors for Proficiency in Life Science—Grade 4

Country	Sample Size	Life Science				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Albania	4426	488.235	10.851	2.967	13.818	3.717
Armenia	5399	475.812	8.178	2.171	10.349	3.217
Australia	5890	539.141	5.857	1.831	7.688	2.773
Austria	5097	523.171	4.066	1.321	5.386	2.321
Azerbaijan	5245	423.083	9.538	1.789	11.327	3.366
Bahrain	5762	491.680	11.564	1.315	12.880	3.589
Belgium (Flemish)	4655	499.731	3.645	2.358	6.003	2.450
Bosnia and Herzegovina	5617	471.113	6.833	4.032	10.865	3.296
Bulgaria	4268	525.291	24.469	2.436	26.904	5.187
Canada	15582	531.891	2.155	1.316	3.471	1.863
Chile	4775	477.897	4.975	1.304	6.278	2.506
Chinese Taipei	4295	540.497	2.185	1.991	4.175	2.043
Croatia	4335	520.242	3.043	2.025	5.067	2.251
Cyprus	4062	514.903	7.284	3.828	11.112	3.333
Czech Republic	5358	535.472	4.427	0.597	5.024	2.241
Denmark	3693	526.397	3.644	1.013	4.657	2.158
England	3872	537.469	6.008	0.738	6.746	2.597
Finland	5397	558.312	4.761	3.490	8.251	2.873
France	4792	493.601	8.348	1.117	9.465	3.077
Georgia	4316	456.951	12.997	3.346	16.342	4.043
Germany	3933	521.427	4.763	0.617	5.380	2.319
Hong Kong SAR	3386	523.155	10.597	2.250	12.847	3.584
Hungary	5227	533.350	6.240	5.540	11.780	3.432
Iran, Islamic Rep. of	6010	429.949	16.037	4.422	20.460	4.523
Ireland	4582	527.948	9.977	2.013	11.990	3.463
Italy	4269	513.748	6.311	4.565	10.875	3.298
Japan	4196	550.278	2.536	1.567	4.103	2.026
Kazakhstan	4791	486.196	8.623	3.341	11.963	3.459
Korea, Rep. of	4448	574.267	4.081	2.415	6.497	2.549
Kosovo	4496	407.760	12.803	5.756	18.559	4.308
Kuwait	4437	–	–	–	–	–
Latvia	4481	534.568	5.420	1.804	7.224	2.688
Lithuania	4265	536.658	5.187	2.532	7.719	2.778
Malta	4154	499.335	1.531	4.543	6.074	2.464
Montenegro	5076	464.311	4.530	0.521	5.050	2.247
Morocco	7723	363.768	31.147	4.138	35.285	5.940
Netherlands	3831	517.876	7.104	3.596	10.700	3.271
New Zealand	5019	510.077	3.999	1.405	5.404	2.325
North Macedonia	3270	421.842	32.352	2.157	34.510	5.874
Northern Ireland	3497	520.134	3.438	4.331	7.769	2.787
Norway (5)	4527	547.003	4.338	4.802	9.140	3.023
Oman	6814	434.459	14.580	6.970	21.550	4.642
Pakistan	3980	–	–	–	–	–
Philippines	5515	–	–	–	–	–
Poland	4882	533.749	5.181	4.297	9.478	3.079
Portugal	4915	508.825	3.420	0.370	3.791	1.947
Qatar	5646	448.118	15.116	6.112	21.228	4.607
Russian Federation	4596	570.364	7.330	2.565	9.895	3.146

Summary Statistics and Standard Errors for Proficiency in Life Science—Grade 4 (continued)

Country	Sample Size	Life Science: Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Saudi Arabia	5453	–	–	–	–	–
Serbia	4380	520.783	10.714	3.916	14.630	3.825
Singapore	6839	603.016	12.558	0.743	13.300	3.647
Slovak Republic	4862	520.034	12.170	2.918	15.088	3.884
South Africa (5)	11891	–	–	–	–	–
Spain	10946	513.882	2.646	2.096	4.742	2.178
Sweden	4535	541.434	9.250	1.340	10.590	3.254
Turkey (5)	4599	518.678	17.496	3.325	20.821	4.563
United Arab Emirates	29515	466.709	2.923	1.071	3.993	1.998
United States	10029	546.431	5.566	0.821	6.386	2.527
Benchmarking Participants						
Ontario, Canada	4360	534.859	7.100	1.391	8.491	2.914
Quebec, Canada	4384	529.545	4.069	1.928	5.998	2.449
Moscow City, Russian Fed.	4392	595.338	4.339	3.029	7.368	2.714
Madrid, Spain	3879	524.595	3.650	7.943	11.592	3.405
Abu Dhabi, UAE	10328	413.051	4.445	1.714	6.159	2.482
Dubai, UAE	8299	537.447	2.062	1.429	3.490	1.868

Summary Statistics and Standard Errors for Proficiency in Physical Science—Grade 4

Country	Sample Size	Physical Science				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Albania	4426	493.368	13.752	2.873	16.626	4.077
Armenia	5399	453.897	9.309	2.272	11.580	3.403
Australia	5890	525.828	6.274	0.928	7.202	2.684
Austria	5097	519.431	4.820	1.897	6.716	2.592
Azerbaijan	5245	426.872	9.955	0.881	10.836	3.292
Bahrain	5762	496.246	12.388	2.424	14.812	3.849
Belgium (Flemish)	4655	502.074	3.390	1.844	5.235	2.288
Bosnia and Herzegovina	5617	450.345	7.573	3.596	11.169	3.342
Bulgaria	4268	518.080	32.938	7.467	40.405	6.357
Canada	15582	512.830	2.423	0.788	3.211	1.792
Chile	4775	457.694	7.631	6.984	14.614	3.823
Chinese Taipei	4295	573.203	2.577	1.126	3.703	1.924
Croatia	4335	527.709	2.895	2.684	5.579	2.362
Cyprus	4062	511.033	9.208	0.867	10.074	3.174
Czech Republic	5358	527.994	4.430	2.010	6.441	2.538
Denmark	3693	507.047	3.254	2.124	5.378	2.319
England	3872	536.978	7.463	2.839	10.302	3.210
Finland	5397	544.061	5.400	5.076	10.476	3.237
France	4792	477.467	7.913	1.827	9.740	3.121
Georgia	4316	452.496	15.278	5.678	20.956	4.578
Germany	3933	518.375	5.375	3.824	9.199	3.033
Hong Kong SAR	3386	528.796	11.179	1.271	12.450	3.528
Hungary	5227	523.884	6.019	1.607	7.626	2.762
Iran, Islamic Rep. of	6010	452.770	20.607	1.766	22.373	4.730
Ireland	4582	522.775	8.264	1.919	10.183	3.191
Italy	4269	501.845	6.379	5.495	11.874	3.446
Japan	4196	578.571	3.285	0.283	3.569	1.889
Kazakhstan	4791	506.307	10.574	0.513	11.088	3.330
Korea, Rep. of	4448	606.851	4.132	3.185	7.316	2.705
Kosovo	4496	415.201	12.106	5.180	17.286	4.158
Kuwait	4437	–	–	–	–	–
Latvia	4481	553.378	6.309	6.663	12.972	3.602
Lithuania	4265	547.325	6.077	3.199	9.275	3.046
Malta	4154	491.791	1.676	6.505	8.182	2.860
Montenegro	5076	445.837	4.458	3.470	7.927	2.816
Morocco	7723	378.516	33.324	5.369	38.692	6.220
Netherlands	3831	515.520	6.715	1.147	7.861	2.804
New Zealand	5019	492.277	3.897	0.367	4.264	2.065
North Macedonia	3270	431.900	47.856	3.482	51.338	7.165
Northern Ireland	3497	510.827	4.432	0.335	4.767	2.183
Norway (5)	4527	525.225	4.148	5.020	9.168	3.028
Oman	6814	436.716	18.617	3.024	21.641	4.652
Pakistan	3980	–	–	–	–	–
Philippines	5515	–	–	–	–	–
Poland	4882	525.764	6.554	1.710	8.263	2.875
Portugal	4915	496.429	3.688	2.167	5.855	2.420
Qatar	5646	450.999	14.751	1.397	16.149	4.019
Russian Federation	4596	572.092	7.604	1.000	8.605	2.933

Summary Statistics and Standard Errors for Proficiency in Physical Science—Grade 4 (continued)

Country	Sample Size	Physical Science				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Saudi Arabia	5453	–	–	–	–	–
Serbia	4380	524.273	12.399	5.165	17.564	4.191
Singapore	6839	613.119	13.060	0.586	13.646	3.694
Slovak Republic	4862	525.382	13.915	1.168	15.084	3.884
South Africa (5)	11891	–	–	–	–	–
Spain	10946	503.461	3.344	2.153	5.497	2.344
Sweden	4535	525.075	8.860	1.969	10.829	3.291
Turkey (5)	4599	537.992	17.660	3.341	21.001	4.583
United Arab Emirates	29515	477.422	3.591	1.215	4.807	2.192
United States	10029	527.099	6.262	1.650	7.912	2.813
Benchmarking Participants						
Ontario, Canada	4360	512.245	8.292	0.225	8.517	2.918
Quebec, Canada	4384	513.586	4.803	3.171	7.974	2.824
Moscow City, Russian Fed.	4392	598.457	4.564	2.501	7.065	2.658
Madrid, Spain	3879	514.091	3.962	2.538	6.500	2.550
Abu Dhabi, UAE	10328	417.734	5.420	1.315	6.735	2.595
Dubai, UAE	8299	555.628	2.803	1.682	4.485	2.118

Summary Statistics and Standard Errors for Proficiency in Earth Science—Grade 4

Country	Sample Size	Earth Science				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Albania	4426	474.909	16.389	1.234	17.623	4.198
Armenia	5399	451.275	12.441	2.054	14.494	3.807
Australia	5890	526.772	6.078	1.595	7.673	2.770
Austria	5097	523.718	5.431	6.593	12.024	3.468
Azerbaijan	5245	423.738	10.102	11.926	22.028	4.693
Bahrain	5762	477.747	11.341	4.901	16.242	4.030
Belgium (Flemish)	4655	496.356	3.955	0.746	4.701	2.168
Bosnia and Herzegovina	5617	436.773	8.279	1.755	10.034	3.168
Bulgaria	4268	514.125	21.546	1.481	23.027	4.799
Canada	15582	518.659	2.720	2.024	4.745	2.178
Chile	4775	460.043	7.619	11.055	18.674	4.321
Chinese Taipei	4295	567.902	2.169	1.047	3.216	1.793
Croatia	4335	523.109	4.302	4.827	9.129	3.021
Cyprus	4062	499.825	6.945	0.397	7.342	2.710
Czech Republic	5358	535.596	4.791	4.138	8.928	2.988
Denmark	3693	534.842	3.508	3.744	7.253	2.693
England	3872	532.805	6.119	2.576	8.695	2.949
Finland	5397	563.107	6.158	5.935	12.093	3.478
France	4792	488.298	9.236	0.776	10.012	3.164
Georgia	4316	434.646	15.236	2.483	17.719	4.209
Germany	3933	508.928	5.505	10.538	16.043	4.005
Hong Kong SAR	3386	549.322	9.454	10.439	19.893	4.460
Hungary	5227	531.264	7.825	2.409	10.233	3.199
Iran, Islamic Rep. of	6010	438.022	15.497	1.835	17.331	4.163
Ireland	4582	536.223	9.872	4.856	14.729	3.838
Italy	4269	506.898	7.312	6.521	13.834	3.719
Japan	4196	559.222	3.602	0.175	3.776	1.943
Kazakhstan	4791	487.558	9.533	0.975	10.507	3.242
Korea, Rep. of	4448	586.934	4.573	3.629	8.202	2.864
Kosovo	4496	410.241	12.118	3.370	15.488	3.935
Kuwait	4437	–	–	–	–	–
Latvia	4481	535.307	7.600	5.912	13.512	3.676
Lithuania	4265	524.756	7.533	1.515	9.049	3.008
Malta	4154	491.459	1.839	2.410	4.248	2.061
Montenegro	5076	433.585	4.608	5.297	9.905	3.147
Morocco	7723	349.645	39.183	4.815	43.997	6.633
Netherlands	3831	520.962	9.917	2.496	12.413	3.523
New Zealand	5019	503.404	5.182	4.598	9.780	3.127
North Macedonia	3270	409.198	45.408	6.945	52.353	7.236
Northern Ireland	3497	524.625	3.572	3.375	6.947	2.636
Norway (5)	4527	546.773	4.305	3.843	8.148	2.854
Oman	6814	415.929	15.816	4.771	20.587	4.537
Pakistan	3980	–	–	–	–	–
Philippines	5515	–	–	–	–	–
Poland	4882	529.304	5.651	5.177	10.828	3.291
Portugal	4915	500.943	5.364	3.674	9.038	3.006
Qatar	5646	442.276	17.920	14.323	32.242	5.678
Russian Federation	4596	554.389	10.384	9.394	19.778	4.447

Summary Statistics and Standard Errors for Proficiency in Earth Science—Grade 4 (continued)

Country	Sample Size	Earth Science				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Saudi Arabia	5453	–	–	–	–	–
Serbia	4380	494.207	11.376	8.843	20.219	4.497
Singapore	6839	556.853	12.002	3.567	15.569	3.946
Slovak Republic	4862	513.026	13.831	5.525	19.356	4.400
South Africa (5)	11891	–	–	–	–	–
Spain	10946	518.068	3.132	2.789	5.922	2.433
Sweden	4535	546.607	10.093	4.151	14.243	3.774
Turkey (5)	4599	524.451	15.115	1.208	16.323	4.040
United Arab Emirates	29515	473.901	2.477	0.175	2.652	1.629
United States	10029	538.555	6.639	3.672	10.311	3.211
Benchmarking Participants						
Ontario, Canada	4360	518.094	9.690	1.805	11.495	3.390
Quebec, Canada	4384	518.638	5.573	4.402	9.975	3.158
Moscow City, Russian Fed.	4392	589.320	4.707	4.317	9.024	3.004
Madrid, Spain	3879	533.059	3.524	0.481	4.006	2.001
Abu Dhabi, UAE	10328	421.662	3.897	0.369	4.267	2.066
Dubai, UAE	8299	541.727	2.000	3.326	5.326	2.308

Summary Statistics and Standard Errors for Proficiency in Knowing in Science—Grade 4

Country	Sample Size	Science Knowing				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Albania	4426	493.520	15.190	0.095	15.285	3.910
Armenia	5399	463.452	8.847	2.897	11.744	3.427
Australia	5890	537.851	6.396	2.826	9.222	3.037
Austria	5097	522.948	4.913	4.963	9.876	3.143
Azerbaijan	5245	425.221	9.629	6.675	16.304	4.038
Bahrain	5762	496.453	12.074	1.267	13.341	3.653
Belgium (Flemish)	4655	493.254	3.748	3.447	7.196	2.683
Bosnia and Herzegovina	5617	451.232	6.924	3.607	10.531	3.245
Bulgaria	4268	526.484	27.340	1.409	28.749	5.362
Canada	15582	524.359	2.410	1.125	3.535	1.880
Chile	4775	472.679	6.354	7.642	13.996	3.741
Chinese Taipei	4295	560.459	2.572	1.055	3.628	1.905
Croatia	4335	526.477	3.679	1.919	5.598	2.366
Cyprus	4062	502.802	9.196	1.493	10.690	3.269
Czech Republic	5358	538.284	4.201	4.468	8.669	2.944
Denmark	3693	520.882	3.179	0.915	4.094	2.023
England	3872	543.531	7.029	3.920	10.949	3.309
Finland	5397	553.164	5.371	0.656	6.027	2.455
France	4792	485.417	8.070	5.122	13.192	3.632
Georgia	4316	451.581	14.805	0.468	15.274	3.908
Germany	3933	519.653	5.203	0.318	5.520	2.349
Hong Kong SAR	3386	537.079	9.815	0.556	10.371	3.220
Hungary	5227	533.444	6.781	0.329	7.110	2.666
Iran, Islamic Rep. of	6010	444.077	19.224	1.604	20.828	4.564
Ireland	4582	531.669	10.309	1.525	11.833	3.440
Italy	4269	514.558	6.983	2.098	9.081	3.013
Japan	4196	534.805	3.218	3.628	6.846	2.616
Kazakhstan	4791	488.507	7.698	0.685	8.383	2.895
Korea, Rep. of	4448	584.422	4.261	1.772	6.033	2.456
Kosovo	4496	419.297	13.257	6.842	20.099	4.483
Kuwait	4437	–	–	–	–	–
Latvia	4481	539.164	5.333	4.695	10.029	3.167
Lithuania	4265	539.375	5.624	4.055	9.678	3.111
Malta	4154	496.408	2.099	0.493	2.592	1.610
Montenegro	5076	451.058	4.275	5.767	10.042	3.169
Morocco	7723	362.055	32.135	5.560	37.695	6.140
Netherlands	3831	514.683	6.580	1.041	7.621	2.761
New Zealand	5019	504.910	4.333	2.690	7.024	2.650
North Macedonia	3270	422.905	43.112	4.163	47.275	6.876
Northern Ireland	3497	522.837	4.269	4.121	8.390	2.897
Norway (5)	4527	540.324	4.503	1.881	6.384	2.527
Oman	6814	432.373	18.347	2.677	21.024	4.585
Pakistan	3980	–	–	–	–	–
Philippines	5515	–	–	–	–	–
Poland	4882	524.361	5.964	1.037	7.001	2.646
Portugal	4915	502.347	3.756	3.891	7.647	2.765
Qatar	5646	454.610	16.662	2.396	19.058	4.366
Russian Federation	4596	562.018	10.081	0.652	10.733	3.276

Summary Statistics and Standard Errors for Proficiency in Knowing in Science—Grade 4 (continued)

Country	Sample Size	Science Knowing				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Saudi Arabia	5453	–	–	–	–	–
Serbia	4380	506.225	10.924	0.262	11.185	3.344
Singapore	6839	587.656	12.735	0.708	13.444	3.667
Slovak Republic	4862	527.135	12.786	2.312	15.099	3.886
South Africa (5)	11891	–	–	–	–	–
Spain	10946	514.045	3.435	1.409	4.844	2.201
Sweden	4535	540.293	9.282	2.498	11.780	3.432
Turkey (5)	4599	530.733	19.372	1.290	20.662	4.546
United Arab Emirates	29515	481.735	3.413	1.392	4.805	2.192
United States	10029	542.058	6.545	0.985	7.530	2.744
Benchmarking Participants						
Ontario, Canada	4360	524.869	8.615	0.841	9.455	3.075
Quebec, Canada	4384	522.978	4.614	3.195	7.808	2.794
Moscow City, Russian Fed.	4392	591.652	4.335	0.072	4.406	2.099
Madrid, Spain	3879	523.214	4.168	9.646	13.814	3.717
Abu Dhabi, UAE	10328	421.851	5.606	2.871	8.478	2.912
Dubai, UAE	8299	559.673	2.925	1.311	4.236	2.058

Summary Statistics and Standard Errors for Proficiency in Applying in Science—Grade 4

Country	Sample Size	Science Applying				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Albania	4426	485.101	10.642	3.652	14.294	3.781
Armenia	5399	453.249	8.895	2.159	11.054	3.325
Australia	5890	523.645	6.218	3.930	10.148	3.186
Austria	5097	523.330	4.473	1.417	5.890	2.427
Azerbaijan	5245	418.971	10.559	9.524	20.083	4.481
Bahrain	5762	494.376	10.727	1.002	11.729	3.425
Belgium (Flemish)	4655	501.090	3.187	1.454	4.641	2.154
Bosnia and Herzegovina	5617	458.511	7.321	1.790	9.111	3.018
Bulgaria	4268	522.599	26.521	2.668	29.189	5.403
Canada	15582	519.564	2.235	1.718	3.953	1.988
Chile	4775	460.625	6.635	4.845	11.480	3.388
Chinese Taipei	4295	560.549	2.337	1.641	3.978	1.994
Croatia	4335	521.202	2.339	3.010	5.349	2.313
Cyprus	4062	519.222	7.550	1.288	8.838	2.973
Czech Republic	5358	526.297	4.752	1.261	6.013	2.452
Denmark	3693	519.113	3.383	2.785	6.168	2.483
England	3872	525.617	6.096	2.655	8.751	2.958
Finland	5397	550.927	5.432	1.047	6.479	2.545
France	4792	494.742	8.121	0.737	8.858	2.976
Georgia	4316	445.263	12.904	1.046	13.950	3.735
Germany	3933	515.952	4.637	1.862	6.499	2.549
Hong Kong SAR	3386	526.193	9.048	0.599	9.646	3.106
Hungary	5227	525.553	6.436	3.168	9.604	3.099
Iran, Islamic Rep. of	6010	440.479	15.754	3.147	18.901	4.347
Ireland	4582	525.392	8.565	0.354	8.919	2.986
Italy	4269	503.695	6.128	1.418	7.545	2.747
Japan	4196	576.340	2.626	2.417	5.044	2.246
Kazakhstan	4791	493.660	10.344	1.252	11.595	3.405
Korea, Rep. of	4448	596.002	4.948	1.582	6.530	2.555
Kosovo	4496	406.283	10.870	2.857	13.726	3.705
Kuwait	4437	–	–	–	–	–
Latvia	4481	540.032	6.380	0.465	6.846	2.616
Lithuania	4265	531.074	5.044	0.128	5.172	2.274
Malta	4154	496.039	1.494	6.059	7.553	2.748
Montenegro	5076	453.692	4.740	2.644	7.384	2.717
Morocco	7723	378.025	32.747	6.026	38.773	6.227
Netherlands	3831	517.129	6.855	2.687	9.542	3.089
New Zealand	5019	497.463	3.942	2.595	6.536	2.557
North Macedonia	3270	422.930	37.216	1.699	38.915	6.238
Northern Ireland	3497	514.108	2.963	2.367	5.330	2.309
Norway (5)	4527	536.815	4.354	1.509	5.863	2.421
Oman	6814	433.588	15.144	1.250	16.393	4.049
Pakistan	3980	–	–	–	–	–
Philippines	5515	–	–	–	–	–
Poland	4882	537.985	5.012	1.397	6.409	2.532
Portugal	4915	502.194	4.233	5.254	9.487	3.080
Qatar	5646	450.600	13.408	4.174	17.583	4.193
Russian Federation	4596	571.833	9.492	1.749	11.241	3.353

Summary Statistics and Standard Errors for Proficiency in Applying in Science—Grade 4 (continued)

Country	Sample Size	Science Applying				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Saudi Arabia	5453	–	–	–	–	–
Serbia	4380	525.707	11.839	3.545	15.384	3.922
Singapore	6839	595.084	11.510	2.194	13.704	3.702
Slovak Republic	4862	515.345	14.272	3.814	18.086	4.253
South Africa (5)	11891	–	–	–	–	–
Spain	10946	510.634	2.986	1.061	4.047	2.012
Sweden	4535	532.461	8.868	0.878	9.746	3.122
Turkey (5)	4599	528.137	16.898	1.302	18.200	4.266
United Arab Emirates	29515	469.974	2.669	1.699	4.368	2.090
United States	10029	535.051	6.543	3.291	9.835	3.136
Benchmarking Participants						
Ontario, Canada	4360	519.720	8.028	1.378	9.406	3.067
Quebec, Canada	4384	519.627	4.303	8.808	13.111	3.621
Moscow City, Russian Fed.	4392	602.505	4.702	0.916	5.618	2.370
Madrid, Spain	3879	521.469	3.949	10.752	14.701	3.834
Abu Dhabi, UAE	10328	415.135	4.373	4.343	8.716	2.952
Dubai, UAE	8299	540.986	2.081	3.002	5.083	2.255

Summary Statistics and Standard Errors for Proficiency in Reasoning in Science—Grade 4

Country	Sample Size	Science Reasoning				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Albania	4426	487.010	12.216	0.726	12.943	3.598
Armenia	5399	485.656	8.880	3.830	12.710	3.565
Australia	5890	537.743	5.627	3.096	8.723	2.954
Austria	5097	518.283	4.539	6.429	10.969	3.312
Azerbaijan	5245	429.553	10.824	1.269	12.092	3.477
Bahrain	5762	481.455	10.543	2.588	13.131	3.624
Belgium (Flemish)	4655	510.883	3.549	2.099	5.648	2.377
Bosnia and Herzegovina	5617	469.002	7.838	0.988	8.827	2.971
Bulgaria	4268	507.498	26.974	3.316	30.290	5.504
Canada	15582	525.462	2.235	0.962	3.196	1.788
Chile	4775	472.026	5.440	1.946	7.385	2.718
Chinese Taipei	4295	552.213	2.581	4.738	7.319	2.705
Croatia	4335	521.849	4.220	1.797	6.017	2.453
Cyprus	4062	510.740	6.372	3.831	10.203	3.194
Czech Republic	5358	538.648	4.274	5.729	10.003	3.163
Denmark	3693	527.333	3.486	4.070	7.556	2.749
England	3872	543.506	7.158	6.188	13.345	3.653
Finland	5397	562.640	4.779	0.841	5.620	2.371
France	4792	474.755	9.975	12.162	22.137	4.705
Georgia	4316	465.311	12.706	6.971	19.677	4.436
Germany	3933	518.469	5.974	2.386	8.360	2.891
Hong Kong SAR	3386	530.616	11.279	1.473	12.752	3.571
Hungary	5227	531.802	5.589	1.406	6.996	2.645
Iran, Islamic Rep. of	6010	432.459	14.476	9.066	23.542	4.852
Ireland	4582	525.444	8.616	6.097	14.713	3.836
Italy	4269	507.940	6.022	1.131	7.153	2.674
Japan	4196	579.490	3.107	2.523	5.630	2.373
Kazakhstan	4791	501.699	8.597	3.152	11.749	3.428
Korea, Rep. of	4448	581.415	3.681	1.866	5.547	2.355
Kosovo	4496	402.304	14.905	2.960	17.865	4.227
Kuwait	4437	–	–	–	–	–
Latvia	4481	546.550	6.258	0.223	6.481	2.546
Lithuania	4265	547.895	6.038	2.563	8.601	2.933
Malta	4154	490.053	1.445	12.695	14.140	3.760
Montenegro	5076	451.059	4.784	6.035	10.818	3.289
Morocco	7723	365.474	26.434	3.542	29.976	5.475
Netherlands	3831	523.318	7.593	2.688	10.280	3.206
New Zealand	5019	504.804	4.022	2.593	6.616	2.572
North Macedonia	3270	424.594	39.456	4.672	44.128	6.643
Northern Ireland	3497	519.104	4.677	5.286	9.963	3.156
Norway (5)	4527	539.855	4.768	1.498	6.266	2.503
Oman	6814	433.209	13.810	2.971	16.781	4.096
Pakistan	3980	–	–	–	–	–
Philippines	5515	–	–	–	–	–
Poland	4882	525.421	5.493	1.443	6.936	2.634
Portugal	4915	503.602	3.390	0.470	3.860	1.965
Qatar	5646	433.558	13.636	4.638	18.274	4.275
Russian Federation	4596	569.144	6.894	0.913	7.807	2.794

Summary Statistics and Standard Errors for Proficiency in Reasoning in Science—Grade 4 (continued)

Country	Sample Size	Science Reasoning				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Saudi Arabia	5453	–	–	–	–	–
Serbia	4380	517.777	9.896	5.498	15.394	3.924
Singapore	6839	603.828	10.360	1.765	12.125	3.482
Slovak Republic	4862	516.101	13.445	4.293	17.738	4.212
South Africa (5)	11891	–	–	–	–	–
Spain	10946	506.571	2.890	0.426	3.316	1.821
Sweden	4535	540.965	9.032	1.155	10.187	3.192
Turkey (5)	4599	520.490	13.996	3.153	17.149	4.141
United Arab Emirates	29515	461.474	2.626	1.021	3.647	1.910
United States	10029	538.423	5.025	2.413	7.437	2.727
Benchmarking Participants						
Ontario, Canada	4360	527.619	7.398	1.409	8.807	2.968
Quebec, Canada	4384	524.675	4.240	5.002	9.242	3.040
Moscow City, Russian Fed.	4392	591.683	4.531	3.720	8.251	2.872
Madrid, Spain	3879	519.623	3.179	10.664	13.843	3.721
Abu Dhabi, UAE	10328	411.117	4.174	3.338	7.511	2.741
Dubai, UAE	8299	531.291	1.875	2.410	4.285	2.070

Appendix 14C: Summary Statistics and Standard Errors for Proficiency in Grade 8 Mathematics

Summary Statistics and Standard Errors for Proficiency in Overall Mathematics—Grade 8

Country	Sample Size	Overall Mathematics				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	9060	517.277	14.044	0.177	14.221	3.771
Bahrain	5725	481.087	2.777	0.120	2.897	1.702
Chile	4694	440.607	7.581	0.478	8.059	2.839
Chinese Taipei	5610	612.495	5.826	1.531	7.357	2.712
Cyprus	3521	501.082	2.377	0.242	2.619	1.618
Egypt	7210	412.880	26.447	0.721	27.168	5.212
England	3858	514.927	27.251	0.384	27.634	5.257
Finland	5565	508.916	6.152	0.601	6.753	2.599
France	4426	482.608	5.397	0.624	6.020	2.454
Georgia	3788	461.303	17.386	1.425	18.811	4.337
Hong Kong SAR	3730	578.312	15.364	1.134	16.497	4.062
Hungary	5219	516.541	8.018	0.456	8.474	2.911
Iran, Islamic Rep. of	5980	446.169	13.413	0.473	13.885	3.726
Ireland	4118	523.731	6.474	0.530	7.004	2.646
Israel	4269	519.113	17.907	0.297	18.204	4.267
Italy	4138	497.483	5.975	1.464	7.440	2.728
Japan	4446	594.229	6.945	0.437	7.382	2.717
Jordan	7176	420.268	17.166	1.002	18.169	4.262
Kazakhstan	4453	487.562	10.735	0.141	10.876	3.298
Korea, Rep. of	4409	606.822	6.574	1.253	7.827	2.798
Kuwait	4574	402.747	23.747	1.289	25.036	5.004
Lebanon	4730	429.308	7.823	0.659	8.482	2.912
Lithuania	4366	520.432	8.063	0.582	8.645	2.940
Malaysia	8077	460.567	9.418	0.607	10.026	3.166
Morocco	8458	388.187	4.828	0.382	5.210	2.282
New Zealand	6051	481.592	10.452	0.870	11.322	3.365
Norway (9)	5215	502.871	5.576	0.292	5.868	2.422
Oman	6751	410.657	6.371	1.256	7.627	2.762
Portugal	3867	500.318	9.614	0.496	10.110	3.180
Qatar	4437	443.414	15.834	0.246	16.081	4.010
Romania	4494	478.985	17.801	0.425	18.227	4.269
Russian Federation	4456	543.492	20.307	0.378	20.686	4.548
Saudi Arabia	5680	393.770	4.959	1.512	6.471	2.544
Singapore	5546	615.766	15.176	0.636	15.812	3.976
South Africa (9)	20829	389.477	3.967	1.168	5.135	2.266
Sweden	4564	502.516	6.253	0.243	6.496	2.549
Turkey	4662	495.630	17.061	1.380	18.441	4.294
United Arab Emirates	25538	473.427	3.156	0.299	3.454	1.859
United States	9941	515.441	22.595	0.268	22.862	4.781
Benchmarking Participants						
Ontario, Canada	4329	529.753	17.543	0.599	18.143	4.259
Quebec, Canada	3636	543.210	13.241	0.258	13.499	3.674
Moscow City, Russian Fed.	4324	575.346	16.433	0.969	17.401	4.172
Gauteng, RSA (9)	5633	420.703	8.041	0.774	8.815	2.969
Western Cape, RSA (9)	5351	441.185	17.367	1.927	19.294	4.393
Abu Dhabi, UAE	9380	435.773	7.821	0.733	8.554	2.925
Dubai, UAE	6544	536.581	3.956	0.153	4.109	2.027

Summary Statistics and Standard Errors for Proficiency in Number—Grade 8

Country	Sample Size	Number				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	9060	521.707	15.336	0.127	15.463	3.932
Bahrain	5725	472.615	3.559	1.353	4.913	2.216
Chile	4694	441.937	6.977	3.364	10.340	3.216
Chinese Taipei	5610	613.141	6.559	0.978	7.537	2.745
Cyprus	3521	499.381	2.658	2.219	4.878	2.209
Egypt	7210	413.502	27.138	2.483	29.621	5.443
England	3858	519.081	26.841	1.954	28.794	5.366
Finland	5565	514.909	5.684	1.001	6.685	2.585
France	4426	476.842	5.169	1.470	6.640	2.577
Georgia	3788	466.287	19.842	2.494	22.336	4.726
Hong Kong SAR	3730	569.550	13.882	3.703	17.584	4.193
Hungary	5219	515.395	8.684	1.211	9.894	3.146
Iran, Islamic Rep. of	5980	441.773	15.892	1.911	17.803	4.219
Ireland	4118	540.940	6.383	2.546	8.929	2.988
Israel	4269	518.925	16.694	0.811	17.505	4.184
Italy	4138	494.874	5.749	0.218	5.967	2.443
Japan	4446	578.198	9.576	2.336	11.912	3.451
Jordan	7176	408.396	17.692	2.538	20.231	4.498
Kazakhstan	4453	482.079	10.829	0.904	11.733	3.425
Korea, Rep. of	4409	605.157	6.148	0.871	7.019	2.649
Kuwait	4574	–	–	–	–	–
Lebanon	4730	431.698	7.194	0.148	7.343	2.710
Lithuania	4366	514.059	8.634	0.217	8.852	2.975
Malaysia	8077	457.697	9.027	0.741	9.768	3.125
Morocco	8458	377.425	5.811	1.476	7.287	2.699
New Zealand	6051	483.246	12.022	0.816	12.837	3.583
Norway (9)	5215	507.398	4.965	0.301	5.266	2.295
Oman	6751	392.131	6.695	2.474	9.169	3.028
Portugal	3867	492.416	10.103	1.084	11.187	3.345
Qatar	4437	441.431	16.001	0.296	16.298	4.037
Romania	4494	477.886	17.501	2.325	19.826	4.453
Russian Federation	4456	541.156	20.577	0.908	21.485	4.635
Saudi Arabia	5680	–	–	–	–	–
Singapore	5546	611.096	14.347	2.252	16.598	4.074
South Africa (9)	20829	–	–	–	–	–
Sweden	4564	501.828	4.814	0.773	5.586	2.364
Turkey	4662	493.175	18.001	0.129	18.130	4.258
United Arab Emirates	25538	474.301	2.984	0.455	3.439	1.855
United States	9941	519.901	19.597	0.214	19.811	4.451
Benchmarking Participants						
Ontario, Canada	4329	530.292	17.155	1.497	18.651	4.319
Quebec, Canada	3636	543.711	14.190	1.180	15.370	3.920
Moscow City, Russian Fed.	4324	574.077	18.389	1.504	19.893	4.460
Gauteng, RSA (9)	5633	420.761	8.103	2.396	10.499	3.240
Western Cape, RSA (9)	5351	444.598	17.850	9.306	27.155	5.211
Abu Dhabi, UAE	9380	439.203	7.014	2.110	9.124	3.021
Dubai, UAE	6544	536.637	4.104	0.433	4.537	2.130

Summary Statistics and Standard Errors for Proficiency in Algebra—Grade 8

Country	Sample Size	Algebra: Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	9060	501.398	16.060	1.049	17.109	4.136
Bahrain	5725	485.039	2.928	1.373	4.302	2.074
Chile	4694	438.542	8.257	1.402	9.659	3.108
Chinese Taipei	5610	618.254	5.960	0.997	6.956	2.638
Cyprus	3521	514.814	2.635	3.930	6.565	2.562
Egypt	7210	412.860	33.009	2.550	35.560	5.963
England	3858	503.735	32.566	1.273	33.839	5.817
Finland	5565	488.887	7.326	1.096	8.422	2.902
France	4426	467.980	6.001	2.019	8.020	2.832
Georgia	3788	473.258	17.780	0.949	18.729	4.328
Hong Kong SAR	3730	583.623	13.252	2.024	15.276	3.908
Hungary	5219	508.619	8.639	0.133	8.772	2.962
Iran, Islamic Rep. of	5980	450.423	13.567	0.707	14.274	3.778
Ireland	4118	505.358	7.269	0.652	7.921	2.814
Israel	4269	527.828	24.003	0.854	24.857	4.986
Italy	4138	490.877	5.727	1.573	7.299	2.702
Japan	4446	602.110	8.091	1.948	10.039	3.169
Jordan	7176	441.794	22.402	0.524	22.926	4.788
Kazakhstan	4453	503.546	13.311	0.297	13.608	3.689
Korea, Rep. of	4409	609.097	9.620	2.393	12.013	3.466
Kuwait	4574	–	–	–	–	–
Lebanon	4730	451.905	8.212	0.746	8.957	2.993
Lithuania	4366	518.057	7.700	0.448	8.147	2.854
Malaysia	8077	456.274	9.312	1.586	10.897	3.301
Morocco	8458	370.444	7.942	1.644	9.586	3.096
New Zealand	6051	464.141	11.246	1.109	12.355	3.515
Norway (9)	5215	476.997	7.895	0.966	8.861	2.977
Oman	6751	426.669	7.525	1.769	9.294	3.049
Portugal	3867	498.813	11.156	0.021	11.177	3.343
Qatar	4437	453.550	14.632	1.756	16.389	4.048
Romania	4494	489.858	20.548	0.783	21.331	4.619
Russian Federation	4456	559.931	22.756	1.942	24.698	4.970
Saudi Arabia	5680	–	–	–	–	–
Singapore	5546	618.972	18.670	2.250	20.920	4.574
South Africa (9)	20829	–	–	–	–	–
Sweden	4564	495.883	6.490	2.209	8.699	2.949
Turkey	4662	492.814	20.209	1.121	21.330	4.618
United Arab Emirates	25538	485.853	3.383	1.020	4.403	2.098
United States	9941	519.854	28.422	0.210	28.632	5.351
Benchmarking Participants						
Ontario, Canada	4329	514.768	18.620	0.780	19.400	4.405
Quebec, Canada	3636	530.778	14.055	2.148	16.202	4.025
Moscow City, Russian Fed.	4324	592.353	17.460	0.142	17.602	4.196
Gauteng, RSA (9)	5633	431.329	9.242	4.323	13.564	3.683
Western Cape, RSA (9)	5351	451.381	20.608	3.293	23.901	4.889
Abu Dhabi, UAE	9380	448.054	8.502	1.742	10.244	3.201
Dubai, UAE	6544	547.066	3.731	1.860	5.590	2.364

Summary Statistics and Standard Errors for Proficiency in Geometry—Grade 8

Country	Sample Size	Geometry				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	9060	513.265	15.234	0.959	16.193	4.024
Bahrain	5725	493.257	3.073	2.239	5.312	2.305
Chile	4694	434.459	7.426	10.850	18.276	4.275
Chinese Taipei	5610	623.173	6.427	0.950	7.377	2.716
Cyprus	3521	489.872	2.559	2.728	5.286	2.299
Egypt	7210	417.003	26.982	0.752	27.733	5.266
England	3858	508.636	27.562	0.849	28.412	5.330
Finland	5565	510.625	7.875	2.487	10.362	3.219
France	4426	493.173	6.033	1.397	7.430	2.726
Georgia	3788	448.845	17.643	1.836	19.479	4.413
Hong Kong SAR	3730	596.009	18.643	2.163	20.806	4.561
Hungary	5219	521.359	8.773	2.292	11.065	3.326
Iran, Islamic Rep. of	5980	441.513	16.060	3.684	19.744	4.443
Ireland	4118	506.181	6.978	0.779	7.757	2.785
Israel	4269	506.333	21.788	1.433	23.221	4.819
Italy	4138	509.750	7.832	5.724	13.556	3.682
Japan	4446	610.049	7.767	4.037	11.804	3.436
Jordan	7176	413.044	17.277	4.323	21.599	4.648
Kazakhstan	4453	486.047	13.153	1.426	14.579	3.818
Korea, Rep. of	4409	617.207	6.659	1.866	8.525	2.920
Kuwait	4574	–	–	–	–	–
Lebanon	4730	422.228	7.921	2.125	10.046	3.170
Lithuania	4366	529.409	8.990	0.212	9.202	3.034
Malaysia	8077	466.142	9.570	3.782	13.351	3.654
Morocco	8458	413.406	3.802	1.216	5.018	2.240
New Zealand	6051	476.769	10.827	0.704	11.531	3.396
Norway (9)	5215	501.938	4.891	0.265	5.157	2.271
Oman	6751	418.101	7.080	3.070	10.151	3.186
Portugal	3867	509.307	9.795	1.138	10.933	3.307
Qatar	4437	435.025	15.763	0.424	16.187	4.023
Romania	4494	472.008	21.499	0.414	21.913	4.681
Russian Federation	4456	540.349	26.078	1.142	27.219	5.217
Saudi Arabia	5680	–	–	–	–	–
Singapore	5546	618.882	14.866	0.509	15.375	3.921
South Africa (9)	20829	–	–	–	–	–
Sweden	4564	495.300	8.495	1.218	9.712	3.116
Turkey	4662	489.518	16.797	0.784	17.581	4.193
United Arab Emirates	25538	461.829	3.781	0.746	4.527	2.128
United States	9941	499.144	21.973	0.593	22.566	4.750
Benchmarking Participants						
Ontario, Canada	4329	535.681	19.355	3.935	23.290	4.826
Quebec, Canada	3636	549.050	15.176	4.456	19.632	4.431
Moscow City, Russian Fed.	4324	565.106	18.316	0.894	19.210	4.383
Gauteng, RSA (9)	5633	406.524	8.323	4.865	13.188	3.632
Western Cape, RSA (9)	5351	426.732	16.965	11.587	28.552	5.343
Abu Dhabi, UAE	9380	419.501	9.641	1.817	11.458	3.385
Dubai, UAE	6544	527.076	4.507	2.271	6.778	2.603

Summary Statistics and Standard Errors for Proficiency in Data and Probability—Grade 8

Country	Sample Size	Data and Probability				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	9060	532.746	13.573	1.314	14.888	3.858
Bahrain	5725	465.179	3.277	0.789	4.066	2.016
Chile	4694	434.424	8.324	2.062	10.386	3.223
Chinese Taipei	5610	593.234	5.644	0.844	6.488	2.547
Cyprus	3521	493.229	3.237	3.797	7.034	2.652
Egypt	7210	380.148	27.237	2.209	29.446	5.426
England	3858	523.446	36.522	1.678	38.200	6.181
Finland	5565	513.853	8.336	4.503	12.839	3.583
France	4426	495.500	5.916	1.067	6.982	2.642
Georgia	3788	429.485	18.013	8.267	26.279	5.126
Hong Kong SAR	3730	562.626	24.524	6.534	31.058	5.573
Hungary	5219	520.619	8.154	2.394	10.548	3.248
Iran, Islamic Rep. of	5980	435.406	13.982	1.910	15.892	3.986
Ireland	4118	540.628	9.156	2.516	11.672	3.416
Israel	4269	511.180	18.271	5.739	24.010	4.900
Italy	4138	493.873	6.943	3.831	10.774	3.282
Japan	4446	594.204	5.817	0.443	6.260	2.502
Jordan	7176	396.059	14.280	3.353	17.634	4.199
Kazakhstan	4453	462.561	9.716	1.307	11.023	3.320
Korea, Rep. of	4409	597.614	5.611	1.042	6.652	2.579
Kuwait	4574	–	–	–	–	–
Lebanon	4730	383.185	9.115	2.927	12.042	3.470
Lithuania	4366	522.383	8.474	0.871	9.345	3.057
Malaysia	8077	456.593	12.079	0.519	12.598	3.549
Morocco	8458	372.355	4.055	1.816	5.871	2.423
New Zealand	6051	495.868	11.422	2.248	13.671	3.697
Norway (9)	5215	518.124	7.704	1.158	8.863	2.977
Oman	6751	393.309	7.910	0.325	8.235	2.870
Portugal	3867	497.688	8.992	1.025	10.017	3.165
Qatar	4437	423.360	20.945	1.455	22.400	4.733
Romania	4494	457.526	18.614	1.395	20.009	4.473
Russian Federation	4456	517.118	18.875	3.294	22.170	4.708
Saudi Arabia	5680	–	–	–	–	–
Singapore	5546	620.001	16.790	6.767	23.557	4.854
South Africa (9)	20829	–	–	–	–	–
Sweden	4564	513.023	9.957	3.872	13.829	3.719
Turkey	4662	502.174	16.676	2.188	18.863	4.343
United Arab Emirates	25538	451.107	3.557	0.781	4.338	2.083
United States	9941	509.327	27.128	1.714	28.842	5.370
Benchmarking Participants						
Ontario, Canada	4329	541.614	22.310	4.911	27.220	5.217
Quebec, Canada	3636	554.380	14.887	5.644	20.530	4.531
Moscow City, Russian Fed.	4324	564.343	14.952	3.104	18.056	4.249
Gauteng, RSA (9)	5633	405.563	8.558	3.589	12.148	3.485
Western Cape, RSA (9)	5351	425.641	17.216	8.395	25.611	5.061
Abu Dhabi, UAE	9380	410.520	9.589	0.056	9.645	3.106
Dubai, UAE	6544	525.187	5.071	2.410	7.481	2.735

Summary Statistics and Standard Errors for Proficiency in Knowing in Mathematics—Grade 8

Country	Sample Size	Mathematics Knowing				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	9060	510.615	14.671	0.990	15.660	3.957
Bahrain	5725	471.190	2.837	0.173	3.010	1.735
Chile	4694	433.956	8.561	0.238	8.799	2.966
Chinese Taipei	5610	615.928	6.952	2.024	8.976	2.996
Cyprus	3521	508.644	2.720	1.265	3.986	1.996
Egypt	7210	416.001	32.315	1.061	33.375	5.777
England	3858	510.217	26.475	3.759	30.234	5.499
Finland	5565	504.945	5.502	0.791	6.293	2.509
France	4426	473.161	6.168	1.457	7.625	2.761
Georgia	3788	458.037	20.764	4.010	24.774	4.977
Hong Kong SAR	3730	580.084	14.053	2.160	16.212	4.026
Hungary	5219	515.904	8.767	0.764	9.530	3.087
Iran, Islamic Rep. of	5980	440.556	17.321	0.552	17.873	4.228
Ireland	4118	530.358	6.078	1.837	7.915	2.813
Israel	4269	515.757	20.888	1.962	22.850	4.780
Italy	4138	492.290	6.640	1.355	7.995	2.828
Japan	4446	588.824	8.248	1.326	9.574	3.094
Jordan	7176	413.661	24.478	0.215	24.693	4.969
Kazakhstan	4453	488.320	12.849	1.179	14.028	3.745
Korea, Rep. of	4409	613.927	7.847	2.484	10.331	3.214
Kuwait	4574	–	–	–	–	–
Lebanon	4730	455.591	7.310	1.220	8.530	2.921
Lithuania	4366	518.403	7.458	0.649	8.108	2.847
Malaysia	8077	451.478	11.224	3.150	14.374	3.791
Morocco	8458	382.182	6.697	1.553	8.250	2.872
New Zealand	6051	467.666	9.762	2.376	12.138	3.484
Norway (9)	5215	498.962	4.689	0.769	5.459	2.336
Oman	6751	406.422	7.165	0.424	7.589	2.755
Portugal	3867	498.281	10.647	1.409	12.056	3.472
Qatar	4437	442.695	17.381	3.384	20.765	4.557
Romania	4494	482.304	22.864	2.439	25.303	5.030
Russian Federation	4456	549.698	25.475	1.443	26.918	5.188
Saudi Arabia	5680	–	–	–	–	–
Singapore	5546	614.343	17.230	1.208	18.439	4.294
South Africa (9)	20829	–	–	–	–	–
Sweden	4564	495.989	5.309	1.678	6.987	2.643
Turkey	4662	494.133	22.677	1.846	24.523	4.952
United Arab Emirates	25538	478.196	3.330	0.423	3.753	1.937
United States	9941	521.881	26.312	0.762	27.074	5.203
Benchmarking Participants						
Ontario, Canada	4329	517.976	15.002	2.500	17.502	4.184
Quebec, Canada	3636	545.692	11.497	2.613	14.111	3.756
Moscow City, Russian Fed.	4324	589.071	17.299	0.617	17.916	4.233
Gauteng, RSA (9)	5633	411.383	10.303	2.779	13.082	3.617
Western Cape, RSA (9)	5351	432.474	23.570	11.365	34.935	5.911
Abu Dhabi, UAE	9380	440.428	8.702	1.401	10.102	3.178
Dubai, UAE	6544	539.911	4.468	0.327	4.795	2.190

Summary Statistics and Standard Errors for Proficiency in Applying in Mathematics—Grade 8

Country	Sample Size	Mathematics Applying				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	9060	521.308	14.010	0.689	14.699	3.834
Bahrain	5725	478.923	2.651	0.314	2.965	1.722
Chile	4694	437.717	7.632	0.770	8.402	2.899
Chinese Taipei	5610	609.694	5.754	1.232	6.986	2.643
Cyprus	3521	495.993	2.150	0.823	2.972	1.724
Egypt	7210	405.486	24.557	3.752	28.309	5.321
England	3858	518.150	27.723	0.763	28.486	5.337
Finland	5565	510.446	6.808	0.386	7.194	2.682
France	4426	485.091	5.204	1.768	6.973	2.641
Georgia	3788	460.071	15.774	1.570	17.344	4.165
Hong Kong SAR	3730	575.390	15.337	0.930	16.267	4.033
Hungary	5219	516.847	8.390	0.581	8.971	2.995
Iran, Islamic Rep. of	5980	442.614	11.568	0.779	12.347	3.514
Ireland	4118	526.484	6.995	0.430	7.425	2.725
Israel	4269	518.697	17.501	0.351	17.851	4.225
Italy	4138	496.799	5.774	0.185	5.959	2.441
Japan	4446	596.085	7.018	1.007	8.025	2.833
Jordan	7176	415.300	15.724	0.094	15.819	3.977
Kazakhstan	4453	486.405	9.952	0.417	10.369	3.220
Korea, Rep. of	4409	604.305	6.551	0.703	7.254	2.693
Kuwait	4574	–	–	–	–	–
Lebanon	4730	411.745	8.469	3.837	12.306	3.508
Lithuania	4366	523.888	8.062	1.293	9.355	3.059
Malaysia	8077	463.518	9.704	0.129	9.833	3.136
Morocco	8458	388.654	4.268	1.699	5.966	2.443
New Zealand	6051	486.140	9.632	0.240	9.871	3.142
Norway (9)	5215	503.581	6.496	0.540	7.035	2.652
Oman	6751	409.018	5.953	0.202	6.154	2.481
Portugal	3867	496.611	9.806	1.250	11.056	3.325
Qatar	4437	437.580	16.918	0.302	17.220	4.150
Romania	4494	475.174	16.825	0.329	17.154	4.142
Russian Federation	4456	542.934	19.889	0.268	20.157	4.490
Saudi Arabia	5680	–	–	–	–	–
Singapore	5546	614.102	14.009	0.464	14.473	3.804
South Africa (9)	20829	–	–	–	–	–
Sweden	4564	501.081	6.230	0.739	6.968	2.640
Turkey	4662	491.310	14.805	1.108	15.913	3.989
United Arab Emirates	25538	465.758	3.279	0.109	3.388	1.841
United States	9941	515.001	23.455	0.525	23.979	4.897
Benchmarking Participants						
Ontario, Canada	4329	530.562	19.156	0.765	19.922	4.463
Quebec, Canada	3636	544.493	15.334	1.326	16.660	4.082
Moscow City, Russian Fed.	4324	573.951	17.944	0.523	18.467	4.297
Gauteng, RSA (9)	5633	422.600	7.152	4.062	11.214	3.349
Western Cape, RSA (9)	5351	442.022	15.948	0.785	16.733	4.091
Abu Dhabi, UAE	9380	427.967	8.102	0.344	8.445	2.906
Dubai, UAE	6544	532.163	4.493	0.520	5.013	2.239

Summary Statistics and Standard Errors for Proficiency in Reasoning in Mathematics—Grade 8

Country	Sample Size	Mathematics Reasoning				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	9060	514.746	14.676	0.612	15.288	3.910
Bahrain	5725	489.441	3.510	0.738	4.248	2.061
Chile	4694	450.617	6.750	3.547	10.296	3.209
Chinese Taipei	5610	616.077	5.589	1.506	7.094	2.664
Cyprus	3521	504.638	2.640	1.948	4.587	2.142
Egypt	7210	411.012	31.147	0.278	31.425	5.606
England	3858	512.032	30.506	1.549	32.055	5.662
Finland	5565	506.284	6.748	1.655	8.404	2.899
France	4426	489.002	5.176	1.862	7.038	2.653
Georgia	3788	459.674	17.350	1.675	19.025	4.362
Hong Kong SAR	3730	581.964	17.189	1.880	19.069	4.367
Hungary	5219	512.398	8.000	1.013	9.013	3.002
Iran, Islamic Rep. of	5980	457.434	14.840	1.142	15.982	3.998
Ireland	4118	508.119	7.062	4.249	11.311	3.363
Israel	4269	524.958	19.719	2.643	22.363	4.729
Italy	4138	504.658	5.860	6.938	12.799	3.578
Japan	4446	598.864	7.213	2.812	10.024	3.166
Jordan	7176	431.144	17.653	1.930	19.583	4.425
Kazakhstan	4453	487.171	11.205	0.020	11.225	3.350
Korea, Rep. of	4409	608.986	5.765	3.126	8.892	2.982
Kuwait	4574	–	–	–	–	–
Lebanon	4730	406.844	9.989	3.840	13.829	3.719
Lithuania	4366	513.915	8.940	3.701	12.641	3.555
Malaysia	8077	461.703	7.904	1.925	9.829	3.135
Morocco	8458	381.036	5.081	3.290	8.371	2.893
New Zealand	6051	486.100	9.685	1.596	11.280	3.359
Norway (9)	5215	496.367	6.042	1.545	7.587	2.754
Oman	6751	412.153	6.895	0.755	7.649	2.766
Portugal	3867	507.758	8.659	2.378	11.036	3.322
Qatar	4437	447.563	14.233	0.466	14.699	3.834
Romania	4494	480.567	18.812	1.157	19.969	4.469
Russian Federation	4456	536.366	19.940	3.342	23.282	4.825
Saudi Arabia	5680	–	–	–	–	–
Singapore	5546	620.155	17.880	1.957	19.836	4.454
South Africa (9)	20829	–	–	–	–	–
Sweden	4564	513.563	7.000	1.635	8.636	2.939
Turkey	4662	503.856	16.548	0.298	16.846	4.104
United Arab Emirates	25538	478.954	2.972	0.786	3.758	1.939
United States	9941	507.342	19.328	1.449	20.777	4.558
Benchmarking Participants						
Ontario, Canada	4329	540.382	15.840	4.935	20.775	4.558
Quebec, Canada	3636	537.774	14.506	0.036	14.541	3.813
Moscow City, Russian Fed.	4324	567.658	16.698	1.225	17.923	4.234
Gauteng, RSA (9)	5633	426.782	7.104	4.470	11.574	3.402
Western Cape, RSA (9)	5351	444.480	15.998	6.991	22.989	4.795
Abu Dhabi, UAE	9380	441.348	6.935	1.131	8.065	2.840
Dubai, UAE	6544	541.314	3.978	0.486	4.464	2.113

Appendix 14D: Summary Statistics and Standard Errors for Proficiency in Grade 8 Science

Summary Statistics and Standard Errors for Proficiency in Overall Science—Grade 8

Country	Sample Size	Overall Science				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	9060	528.337	9.822	0.339	10.161	3.188
Bahrain	5725	486.036	2.538	1.142	3.679	1.918
Chile	4694	462.347	7.554	0.804	8.359	2.891
Chinese Taipei	5610	574.334	3.509	0.198	3.707	1.925
Cyprus	3521	483.555	3.225	0.428	3.653	1.911
Egypt	7210	389.328	28.050	1.595	29.645	5.445
England	3856	516.676	23.017	0.491	23.508	4.848
Finland	5565	542.585	9.162	0.716	9.877	3.143
France	4426	488.563	6.598	0.595	7.194	2.682
Georgia	3788	446.739	12.285	2.859	15.144	3.892
Hong Kong SAR	3730	503.515	26.307	0.834	27.141	5.210
Hungary	5217	529.755	6.278	0.539	6.817	2.611
Iran, Islamic Rep. of	5980	449.415	12.278	0.436	12.713	3.566
Ireland	4118	523.095	8.200	0.399	8.599	2.932
Israel	4268	513.305	17.175	0.449	17.624	4.198
Italy	4138	500.476	6.504	0.121	6.625	2.574
Japan	4446	569.506	4.363	0.208	4.571	2.138
Jordan	7176	452.010	20.906	0.885	21.790	4.668
Kazakhstan	4453	478.071	9.107	0.279	9.386	3.064
Korea, Rep. of	4409	560.680	4.232	0.270	4.502	2.122
Kuwait	4574	444.213	30.874	1.069	31.943	5.652
Lebanon	4730	376.883	16.615	4.994	21.609	4.649
Lithuania	4366	533.824	7.486	1.323	8.809	2.968
Malaysia	8077	460.237	11.635	0.506	12.141	3.484
Morocco	8458	394.100	5.580	1.478	7.057	2.657
New Zealand	6051	498.876	11.500	0.672	12.172	3.489
Norway (9)	5205	495.449	7.779	1.810	9.589	3.097
Oman	6751	457.184	7.284	1.032	8.316	2.884
Portugal	3867	518.738	8.000	0.355	8.355	2.891
Qatar	4436	474.528	18.268	0.752	19.020	4.361
Romania	4494	469.789	15.889	1.617	17.506	4.184
Russian Federation	4456	542.859	17.270	0.559	17.829	4.222
Saudi Arabia	5680	431.473	6.186	0.595	6.781	2.604
Singapore	5545	607.554	14.751	0.514	15.265	3.907
South Africa (9)	20829	369.972	6.717	2.887	9.604	3.099
Sweden	4556	521.393	9.877	0.345	10.223	3.197
Turkey	4662	515.488	13.137	0.474	13.611	3.689
United Arab Emirates	25539	472.983	4.170	0.816	4.986	2.233
United States	9942	522.341	21.733	0.042	21.775	4.666
Benchmarking Participants						
Ontario, Canada	4329	521.586	7.943	0.943	8.886	2.981
Quebec, Canada	3637	536.604	12.801	0.092	12.893	3.591
Moscow City, Russian Fed.	4323	566.525	8.303	0.304	8.608	2.934
Gauteng, RSA (9)	5633	422.150	12.150	2.836	14.986	3.871
Western Cape, RSA (9)	5351	439.280	23.053	2.970	26.023	5.101
Abu Dhabi, UAE	9380	420.055	11.268	1.834	13.102	3.620
Dubai, UAE	6544	547.816	3.599	0.464	4.063	2.016

Summary Statistics and Standard Errors for Proficiency in Biology—Grade 8

Country	Sample Size	Biology				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	9060	530.890	9.911	1.004	10.915	3.304
Bahrain	5725	492.387	2.900	0.609	3.509	1.873
Chile	4694	471.236	8.463	0.770	9.233	3.039
Chinese Taipei	5610	576.165	3.372	1.531	4.903	2.214
Cyprus	3521	489.032	4.061	1.465	5.526	2.351
Egypt	7210	381.063	30.835	0.857	31.692	5.630
England	3856	515.790	22.998	3.779	26.777	5.175
Finland	5565	534.044	8.903	1.958	10.861	3.296
France	4426	488.240	7.024	1.244	8.268	2.875
Georgia	3788	447.049	11.122	1.005	12.127	3.482
Hong Kong SAR	3730	500.523	31.026	1.292	32.318	5.685
Hungary	5217	530.057	5.918	1.114	7.032	2.652
Iran, Islamic Rep. of	5980	447.731	11.603	2.317	13.921	3.731
Ireland	4118	521.489	7.220	2.884	10.104	3.179
Israel	4268	512.142	16.610	1.213	17.823	4.222
Italy	4138	508.176	6.805	0.470	7.275	2.697
Japan	4446	573.735	4.373	1.099	5.472	2.339
Jordan	7176	457.275	24.673	2.174	26.847	5.181
Kazakhstan	4453	476.447	8.856	1.422	10.278	3.206
Korea, Rep. of	4409	559.513	3.853	1.083	4.936	2.222
Kuwait	4574	–	–	–	–	–
Lebanon	4730	354.843	18.294	8.097	26.391	5.137
Lithuania	4366	534.595	8.140	1.011	9.151	3.025
Malaysia	8077	462.552	12.143	1.393	13.536	3.679
Morocco	8458	386.625	5.590	3.372	8.961	2.994
New Zealand	6051	498.024	12.134	1.248	13.382	3.658
Norway (9)	5205	485.525	7.388	0.195	7.583	2.754
Oman	6751	466.130	7.465	3.194	10.658	3.265
Portugal	3867	526.842	8.360	0.453	8.814	2.969
Qatar	4436	476.366	18.409	0.992	19.401	4.405
Romania	4494	479.139	16.990	2.122	19.112	4.372
Russian Federation	4456	543.277	18.242	1.663	19.904	4.461
Saudi Arabia	5680	–	–	–	–	–
Singapore	5545	621.685	15.150	2.460	17.610	4.196
South Africa (9)	20829	359.304	7.069	1.717	8.786	2.964
Sweden	4556	518.613	9.335	2.125	11.460	3.385
Turkey	4662	513.004	10.776	0.730	11.506	3.392
United Arab Emirates	25539	474.300	4.948	1.200	6.148	2.480
United States	9942	529.806	22.738	0.262	23.000	4.796
Benchmarking Participants						
Ontario, Canada	4329	534.194	8.234	1.981	10.215	3.196
Quebec, Canada	3637	530.791	12.539	2.095	14.634	3.825
Moscow City, Russian Fed.	4323	565.458	9.405	0.285	9.690	3.113
Gauteng, RSA (9)	5633	415.939	12.742	2.831	15.573	3.946
Western Cape, RSA (9)	5351	432.123	24.629	2.448	27.076	5.203
Abu Dhabi, UAE	9380	416.753	13.073	1.843	14.916	3.862
Dubai, UAE	6544	553.965	4.321	0.439	4.761	2.182

Summary Statistics and Standard Errors for Proficiency in Chemistry—Grade 8

Country	Sample Size	Chemistry				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	9060	514.506	12.666	1.657	14.322	3.785
Bahrain	5725	480.402	2.890	2.902	5.791	2.407
Chile	4694	442.045	8.368	0.312	8.680	2.946
Chinese Taipei	5610	594.141	4.841	0.794	5.635	2.374
Cyprus	3521	478.263	3.482	0.862	4.344	2.084
Egypt	7210	397.186	32.980	1.716	34.697	5.890
England	3856	511.976	30.598	5.273	35.871	5.989
Finland	5565	545.296	12.485	1.776	14.262	3.776
France	4426	464.957	6.698	3.357	10.055	3.171
Georgia	3788	455.959	14.869	3.810	18.679	4.322
Hong Kong SAR	3730	484.988	28.086	2.570	30.656	5.537
Hungary	5217	527.441	6.799	5.486	12.284	3.505
Iran, Islamic Rep. of	5980	450.015	15.341	4.841	20.182	4.492
Ireland	4118	512.076	10.541	4.461	15.002	3.873
Israel	4268	518.484	20.026	0.974	20.999	4.583
Italy	4138	483.571	6.195	2.619	8.814	2.969
Japan	4446	560.362	4.542	2.917	7.459	2.731
Jordan	7176	454.397	24.938	2.873	27.811	5.274
Kazakhstan	4453	493.945	11.507	1.241	12.748	3.570
Korea, Rep. of	4409	550.703	4.256	1.929	6.185	2.487
Kuwait	4574	–	–	–	–	–
Lebanon	4730	412.428	15.729	5.397	21.126	4.596
Lithuania	4366	529.913	8.617	1.388	10.005	3.163
Malaysia	8077	434.060	14.491	3.287	17.778	4.216
Morocco	8458	402.319	6.530	2.696	9.226	3.037
New Zealand	6051	482.195	12.902	1.556	14.458	3.802
Norway (9)	5205	492.303	10.531	3.322	13.853	3.722
Oman	6751	443.230	8.173	1.675	9.847	3.138
Portugal	3867	512.342	10.020	2.485	12.505	3.536
Qatar	4436	474.420	17.492	1.616	19.108	4.371
Romania	4494	466.455	19.153	5.767	24.920	4.992
Russian Federation	4456	550.893	16.800	0.435	17.234	4.151
Saudi Arabia	5680	–	–	–	–	–
Singapore	5545	615.550	23.554	1.634	25.188	5.019
South Africa (9)	20829	371.508	7.764	9.825	17.589	4.194
Sweden	4556	508.858	12.530	1.115	13.644	3.694
Turkey	4662	515.692	18.874	4.148	23.022	4.798
United Arab Emirates	25539	475.265	4.816	1.151	5.967	2.443
United States	9942	508.918	23.437	3.593	27.030	5.199
Benchmarking Participants						
Ontario, Canada	4329	491.868	10.559	4.959	15.518	3.939
Quebec, Canada	3637	548.301	14.157	2.456	16.612	4.076
Moscow City, Russian Fed.	4323	561.054	7.224	1.303	8.527	2.920
Gauteng, RSA (9)	5633	422.547	14.015	3.242	17.258	4.154
Western Cape, RSA (9)	5351	441.693	27.272	24.978	52.250	7.228
Abu Dhabi, UAE	9380	420.980	13.472	3.504	16.976	4.120
Dubai, UAE	6544	554.126	4.168	0.526	4.693	2.166

Summary Statistics and Standard Errors for Proficiency in Physics—Grade 8

Country	Sample Size	Physics				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	9060	528.683	11.315	1.974	13.289	3.645
Bahrain	5725	479.952	3.163	3.576	6.739	2.596
Chile	4694	450.055	8.608	5.055	13.663	3.696
Chinese Taipei	5610	555.076	4.016	3.238	7.253	2.693
Cyprus	3521	479.543	3.599	9.476	13.076	3.616
Egypt	7210	394.105	24.596	0.754	25.350	5.035
England	3856	516.447	24.600	1.310	25.910	5.090
Finland	5565	539.426	10.843	4.351	15.194	3.898
France	4426	490.825	7.823	4.803	12.626	3.553
Georgia	3788	435.833	14.107	10.606	24.713	4.971
Hong Kong SAR	3730	509.783	25.585	5.535	31.120	5.579
Hungary	5217	527.718	7.045	1.217	8.263	2.874
Iran, Islamic Rep. of	5980	453.440	12.741	4.504	17.245	4.153
Ireland	4118	518.631	10.408	4.007	14.415	3.797
Israel	4268	519.817	22.836	1.532	24.368	4.936
Italy	4138	486.616	7.142	12.932	20.074	4.480
Japan	4446	570.344	4.609	1.581	6.190	2.488
Jordan	7176	449.168	18.079	2.898	20.977	4.580
Kazakhstan	4453	475.513	10.332	4.549	14.882	3.858
Korea, Rep. of	4409	569.202	6.092	1.361	7.453	2.730
Kuwait	4574	–	–	–	–	–
Lebanon	4730	377.899	21.752	2.387	24.139	4.913
Lithuania	4366	528.559	8.214	3.751	11.965	3.459
Malaysia	8077	474.958	11.559	0.304	11.863	3.444
Morocco	8458	402.182	6.404	1.812	8.216	2.866
New Zealand	6051	501.682	13.169	0.957	14.126	3.758
Norway (9)	5205	492.837	9.868	3.137	13.005	3.606
Oman	6751	449.401	7.569	1.902	9.471	3.077
Portugal	3867	496.848	7.757	4.579	12.337	3.512
Qatar	4436	469.447	16.890	2.095	18.985	4.357
Romania	4494	457.739	16.655	1.480	18.135	4.259
Russian Federation	4456	540.401	18.605	3.415	22.020	4.693
Saudi Arabia	5680	–	–	–	–	–
Singapore	5545	619.173	15.833	1.010	16.843	4.104
South Africa (9)	20829	380.984	6.056	3.243	9.299	3.049
Sweden	4556	520.180	11.597	2.933	14.530	3.812
Turkey	4662	518.323	13.869	2.517	16.385	4.048
United Arab Emirates	25539	469.252	4.050	1.267	5.317	2.306
United States	9942	514.753	23.995	0.857	24.853	4.985
Benchmarking Participants						
Ontario, Canada	4329	519.644	9.247	3.221	12.469	3.531
Quebec, Canada	3637	520.620	13.913	4.020	17.933	4.235
Moscow City, Russian Fed.	4323	576.193	10.004	2.841	12.844	3.584
Gauteng, RSA (9)	5633	427.661	10.484	9.881	20.365	4.513
Western Cape, RSA (9)	5351	442.032	21.447	16.666	38.113	6.174
Abu Dhabi, UAE	9380	420.119	10.447	3.967	14.414	3.797
Dubai, UAE	6544	539.472	4.125	2.663	6.788	2.605

Summary Statistics and Standard Errors for Proficiency in Earth Science—Grade 8

Country	Sample Size	Earth Science				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	9060	533.237	9.029	1.746	10.775	3.283
Bahrain	5725	475.399	3.212	4.544	7.756	2.785
Chile	4694	463.904	7.950	2.814	10.764	3.281
Chinese Taipei	5610	578.960	3.436	2.733	6.169	2.484
Cyprus	3521	472.965	3.805	3.154	6.959	2.638
Egypt	7210	366.965	28.926	1.731	30.657	5.537
England	3856	517.410	24.111	6.498	30.609	5.533
Finland	5565	558.354	8.022	4.033	12.054	3.472
France	4426	502.438	7.070	11.066	18.136	4.259
Georgia	3788	430.673	12.813	0.273	13.085	3.617
Hong Kong SAR	3730	511.648	27.205	4.483	31.688	5.629
Hungary	5217	534.533	8.234	6.956	15.190	3.897
Iran, Islamic Rep. of	5980	436.599	12.200	3.886	16.086	4.011
Ireland	4118	536.214	8.149	6.048	14.197	3.768
Israel	4268	495.001	19.745	2.780	22.525	4.746
Italy	4138	511.911	10.408	1.737	12.145	3.485
Japan	4446	571.658	5.488	4.938	10.426	3.229
Jordan	7176	427.790	18.879	2.787	21.666	4.655
Kazakhstan	4453	447.804	10.250	6.852	17.102	4.135
Korea, Rep. of	4409	561.780	6.155	4.372	10.527	3.245
Kuwait	4574	–	–	–	–	–
Lebanon	4730	337.152	22.864	3.231	26.095	5.108
Lithuania	4366	534.290	8.460	2.343	10.802	3.287
Malaysia	8077	451.624	14.389	4.105	18.493	4.300
Morocco	8458	356.995	6.791	3.887	10.678	3.268
New Zealand	6051	509.893	11.023	2.731	13.755	3.709
Norway (9)	5205	518.939	8.608	6.555	15.163	3.894
Oman	6751	448.638	8.353	0.753	9.106	3.018
Portugal	3867	530.868	8.940	2.433	11.373	3.372
Qatar	4436	464.897	17.171	7.688	24.859	4.986
Romania	4494	453.352	16.201	5.460	21.661	4.654
Russian Federation	4456	533.092	18.010	1.552	19.562	4.423
Saudi Arabia	5680	–	–	–	–	–
Singapore	5545	562.046	12.606	3.924	16.530	4.066
South Africa (9)	20829	366.122	7.113	3.051	10.164	3.188
Sweden	4556	530.078	9.728	0.292	10.020	3.165
Turkey	4662	509.225	13.614	1.141	14.755	3.841
United Arab Emirates	25539	465.275	4.367	1.189	5.555	2.357
United States	9942	529.535	25.115	0.862	25.977	5.097
Benchmarking Participants						
Ontario, Canada	4329	520.047	8.226	0.913	9.139	3.023
Quebec, Canada	3637	552.540	14.844	5.723	20.567	4.535
Moscow City, Russian Fed.	4323	564.850	11.372	1.569	12.941	3.597
Gauteng, RSA (9)	5633	418.724	13.926	3.015	16.941	4.116
Western Cape, RSA (9)	5351	442.016	25.595	19.063	44.658	6.683
Abu Dhabi, UAE	9380	412.635	10.134	6.623	16.756	4.093
Dubai, UAE	6544	538.338	4.359	1.113	5.472	2.339

Summary Statistics and Standard Errors for Proficiency in Knowing in Science—Grade 8

Country	Sample Size	Science Knowing				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	9060	514.697	11.148	0.951	12.098	3.478
Bahrain	5725	492.769	3.174	0.684	3.858	1.964
Chile	4694	463.248	7.552	3.666	11.218	3.349
Chinese Taipei	5610	600.076	4.174	1.414	5.588	2.364
Cyprus	3521	482.286	3.331	5.775	9.106	3.018
Egypt	7210	395.965	32.821	2.189	35.010	5.917
England	3856	519.604	25.021	0.477	25.498	5.050
Finland	5565	544.527	8.459	1.915	10.375	3.221
France	4426	480.468	7.041	1.590	8.631	2.938
Georgia	3788	458.503	12.601	4.612	17.214	4.149
Hong Kong SAR	3730	501.163	30.959	1.888	32.847	5.731
Hungary	5217	537.353	6.982	1.787	8.769	2.961
Iran, Islamic Rep. of	5980	448.880	13.450	3.581	17.031	4.127
Ireland	4118	512.853	6.763	2.371	9.134	3.022
Israel	4268	513.698	20.479	0.544	21.023	4.585
Italy	4138	507.148	5.817	1.133	6.950	2.636
Japan	4446	562.869	4.561	1.068	5.628	2.372
Jordan	7176	455.300	26.009	2.159	28.169	5.307
Kazakhstan	4453	463.327	11.416	2.468	13.884	3.726
Korea, Rep. of	4409	557.989	5.697	1.145	6.842	2.616
Kuwait	4574	–	–	–	–	–
Lebanon	4730	388.268	15.482	3.765	19.247	4.387
Lithuania	4366	527.349	7.771	1.621	9.392	3.065
Malaysia	8077	442.024	13.821	1.370	15.191	3.898
Morocco	8458	379.809	6.225	3.120	9.345	3.057
New Zealand	6051	479.617	11.951	0.754	12.705	3.564
Norway (9)	5205	497.137	6.226	0.207	6.433	2.536
Oman	6751	461.169	8.702	2.388	11.090	3.330
Portugal	3867	520.448	8.066	1.294	9.360	3.059
Qatar	4436	486.718	16.437	0.890	17.327	4.163
Romania	4494	474.791	13.262	6.281	19.543	4.421
Russian Federation	4456	543.424	19.161	2.730	21.890	4.679
Saudi Arabia	5680	–	–	–	–	–
Singapore	5545	620.671	16.190	1.816	18.006	4.243
South Africa (9)	20829	361.083	7.229	3.021	10.249	3.201
Sweden	4556	521.312	9.650	0.561	10.212	3.196
Turkey	4662	506.090	15.617	2.189	17.806	4.220
United Arab Emirates	25539	481.967	5.356	2.076	7.432	2.726
United States	9942	514.747	20.210	0.986	21.196	4.604
Benchmarking Participants						
Ontario, Canada	4329	504.869	8.302	2.440	10.742	3.278
Quebec, Canada	3637	528.916	10.172	2.440	12.612	3.551
Moscow City, Russian Fed.	4323	570.337	8.671	0.318	8.989	2.998
Gauteng, RSA (9)	5633	413.035	14.295	9.255	23.550	4.853
Western Cape, RSA (9)	5351	427.052	26.167	9.477	35.643	5.970
Abu Dhabi, UAE	9380	421.520	13.781	4.531	18.312	4.279
Dubai, UAE	6544	560.429	4.194	0.987	5.181	2.276

Summary Statistics and Standard Errors for Proficiency in Applying in Science—Grade 8

Country	Sample Size	Science Applying				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	9060	532.062	10.832	0.501	11.333	3.367
Bahrain	5725	480.744	2.327	4.659	6.986	2.643
Chile	4694	462.110	7.595	1.493	9.089	3.015
Chinese Taipei	5610	566.715	3.934	0.570	4.504	2.122
Cyprus	3521	477.274	3.532	0.083	3.615	1.901
Egypt	7210	383.807	29.476	3.244	32.720	5.720
England	3856	514.594	24.429	1.365	25.794	5.079
Finland	5565	536.516	9.724	0.993	10.717	3.274
France	4426	482.050	6.770	1.177	7.947	2.819
Georgia	3788	439.918	12.498	1.011	13.509	3.675
Hong Kong SAR	3730	501.484	26.043	1.264	27.306	5.226
Hungary	5217	528.071	6.543	3.065	9.608	3.100
Iran, Islamic Rep. of	5980	451.631	11.067	1.083	12.150	3.486
Ireland	4118	520.930	8.669	2.902	11.571	3.402
Israel	4268	509.029	16.149	2.549	18.698	4.324
Italy	4138	498.762	7.631	3.874	11.505	3.392
Japan	4446	575.729	5.300	0.184	5.484	2.342
Jordan	7176	453.120	21.217	2.457	23.675	4.866
Kazakhstan	4453	480.744	9.498	2.363	11.861	3.444
Korea, Rep. of	4409	560.072	4.389	1.522	5.911	2.431
Kuwait	4574	–	–	–	–	–
Lebanon	4730	375.493	18.175	8.935	27.110	5.207
Lithuania	4366	530.275	8.266	0.494	8.761	2.960
Malaysia	8077	473.026	11.239	0.554	11.793	3.434
Morocco	8458	393.257	6.044	2.538	8.582	2.930
New Zealand	6051	502.559	13.101	1.717	14.819	3.850
Norway (9)	5205	492.722	8.575	3.600	12.174	3.489
Oman	6751	456.420	7.941	3.445	11.386	3.374
Portugal	3867	514.456	8.510	1.330	9.839	3.137
Qatar	4436	469.461	19.769	0.780	20.549	4.533
Romania	4494	466.774	17.107	0.884	17.991	4.242
Russian Federation	4456	542.658	19.413	0.579	19.992	4.471
Saudi Arabia	5680	–	–	–	–	–
Singapore	5545	608.392	14.942	2.128	17.070	4.132
South Africa (9)	20829	377.249	6.356	2.226	8.582	2.930
Sweden	4556	518.079	10.366	0.457	10.823	3.290
Turkey	4662	514.607	12.038	3.101	15.139	3.891
United Arab Emirates	25539	472.285	4.106	0.759	4.865	2.206
United States	9942	523.273	22.961	0.509	23.471	4.845
Benchmarking Participants						
Ontario, Canada	4329	523.066	9.629	2.290	11.919	3.452
Quebec, Canada	3637	537.504	13.250	4.011	17.261	4.155
Moscow City, Russian Fed.	4323	561.854	9.341	4.456	13.797	3.714
Gauteng, RSA (9)	5633	428.295	11.867	2.169	14.036	3.746
Western Cape, RSA (9)	5351	446.415	22.504	2.318	24.822	4.982
Abu Dhabi, UAE	9380	420.513	11.783	0.487	12.270	3.503
Dubai, UAE	6544	544.602	3.706	2.390	6.096	2.469

Summary Statistics and Standard Errors for Proficiency in Reasoning in Science—Grade 8

Country	Sample Size	Science Reasoning				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	9060	535.575	9.289	0.190	9.479	3.079
Bahrain	5725	482.162	2.446	3.294	5.740	2.396
Chile	4694	457.606	8.248	1.671	9.919	3.149
Chinese Taipei	5610	558.643	3.127	1.093	4.220	2.054
Cyprus	3521	487.781	4.028	1.152	5.180	2.276
Egypt	7210	377.907	25.829	6.100	31.930	5.651
England	3856	513.271	23.905	1.489	25.394	5.039
Finland	5565	547.770	10.519	1.122	11.641	3.412
France	4426	502.326	7.421	1.829	9.250	3.041
Georgia	3788	436.249	13.558	4.198	17.755	4.214
Hong Kong SAR	3730	503.506	25.436	1.514	26.950	5.191
Hungary	5217	524.255	7.440	1.885	9.326	3.054
Iran, Islamic Rep. of	5980	444.029	13.204	6.466	19.670	4.435
Ireland	4118	534.107	8.920	2.771	11.691	3.419
Israel	4268	517.949	16.588	4.423	21.010	4.584
Italy	4138	495.372	7.469	8.182	15.651	3.956
Japan	4446	570.443	4.205	1.918	6.123	2.474
Jordan	7176	442.586	20.331	2.687	23.018	4.798
Kazakhstan	4453	482.152	8.548	3.371	11.919	3.452
Korea, Rep. of	4409	564.013	3.962	1.526	5.487	2.343
Kuwait	4574	–	–	–	–	–
Lebanon	4730	345.754	23.437	3.160	26.597	5.157
Lithuania	4366	540.721	8.969	1.266	10.235	3.199
Malaysia	8077	458.723	11.291	2.402	13.692	3.700
Morocco	8458	397.680	6.474	1.112	7.586	2.754
New Zealand	6051	509.600	11.543	0.848	12.391	3.520
Norway (9)	5205	494.183	10.094	2.675	12.769	3.573
Oman	6751	450.090	6.947	2.020	8.968	2.995
Portugal	3867	519.458	9.190	3.064	12.254	3.501
Qatar	4436	463.967	19.180	1.577	20.756	4.556
Romania	4494	463.921	17.999	1.439	19.438	4.409
Russian Federation	4456	543.015	13.746	6.447	20.193	4.494
Saudi Arabia	5680	–	–	–	–	–
Singapore	5545	594.629	14.739	1.179	15.918	3.990
South Africa (9)	20829	361.576	6.774	2.116	8.890	2.982
Sweden	4556	523.709	10.507	3.874	14.381	3.792
Turkey	4662	523.927	14.025	2.132	16.156	4.019
United Arab Emirates	25539	461.215	3.755	1.131	4.886	2.210
United States	9942	528.228	20.069	1.577	21.646	4.653
Benchmarking Participants						
Ontario, Canada	4329	532.629	7.628	5.991	13.619	3.690
Quebec, Canada	3637	540.499	15.457	1.567	17.024	4.126
Moscow City, Russian Fed.	4323	568.117	8.595	1.201	9.796	3.130
Gauteng, RSA (9)	5633	416.823	12.674	2.033	14.707	3.835
Western Cape, RSA (9)	5351	438.071	23.783	10.218	34.000	5.831
Abu Dhabi, UAE	9380	411.626	10.930	1.937	12.866	3.587
Dubai, UAE	6544	538.236	3.449	1.959	5.408	2.325

Appendix 14E: Summary Statistics and Standard Errors for Proficiency in Mathematics and Science for the Grade 4 Bridge Samples

Summary Statistics and Standard Errors for Proficiency in Overall Mathematics—Grade 4 Bridge Samples

Country	Sample Size	Overall Mathematics				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Austria	1964	533.830	10.461	0.138	10.599	3.256
Canada	1604	511.587	14.238	0.516	14.754	3.841
Chile	1612	436.314	16.937	2.111	19.048	4.364
Chinese Taipei	1663	603.363	6.653	0.153	6.806	2.609
Croatia	1472	511.055	9.897	1.739	11.636	3.411
Czech Republic	2030	518.892	53.219	0.405	53.623	7.323
Denmark	1432	527.810	12.155	0.602	12.757	3.572
England	1242	552.941	23.439	1.932	25.371	5.037
Finland	1983	533.235	10.513	0.850	11.364	3.371
France	1948	480.785	12.746	1.224	13.971	3.738
Georgia	1632	504.578	60.780	2.669	63.449	7.965
Germany	1505	519.271	16.177	1.559	17.737	4.211
Hong Kong SAR	1329	607.214	56.466	5.437	61.904	7.868
Hungary	1778	530.046	25.864	0.456	26.320	5.130
Italy	1921	510.804	23.657	0.680	24.337	4.933
Korea, Rep. of	1541	594.732	6.069	0.146	6.215	2.493
Lithuania	1587	547.329	7.085	0.520	7.605	2.758
Netherlands	1295	528.108	14.844	2.041	16.884	4.109
Norway (5)	1899	540.210	14.867	0.160	15.027	3.876
Portugal	1612	536.042	19.543	1.058	20.601	4.539
Qatar	1486	449.663	35.778	4.669	40.447	6.360
Russian Federation	2128	558.896	14.818	0.142	14.959	3.868
Singapore	1881	631.356	30.425	1.101	31.526	5.615
Slovak Republic	1610	505.109	20.817	1.546	22.363	4.729
Spain	1670	501.836	22.202	0.671	22.873	4.783
Sweden	1697	516.572	32.608	0.942	33.550	5.792
United Arab Emirates	2243	495.741	61.636	0.866	62.502	7.906
United States	1652	536.716	24.934	0.815	25.748	5.074

Summary Statistics and Standard Errors for Proficiency in Overall Science—Grade 4 Bridge Samples

Country	Sample Size	Overall Science				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Austria	1964	511.455	13.693	1.721	15.414	3.926
Canada	1604	512.234	18.296	1.829	20.124	4.486
Chile	1612	461.405	15.149	4.684	19.833	4.453
Chinese Taipei	1663	554.372	6.666	1.703	8.369	2.893
Croatia	1472	523.618	9.932	2.045	11.976	3.461
Czech Republic	2030	517.327	85.905	3.126	89.031	9.436
Denmark	1432	514.178	17.629	0.854	18.483	4.299
England	1242	542.768	20.677	1.224	21.901	4.680
Finland	1983	546.916	13.005	2.706	15.711	3.964
France	1948	478.221	14.022	1.591	15.613	3.951
Georgia	1632	476.959	64.134	1.311	65.445	8.090
Germany	1505	522.197	22.184	0.228	22.412	4.734
Hong Kong SAR	1329	542.344	52.295	0.818	53.112	7.288
Hungary	1778	532.838	39.090	0.089	39.179	6.259
Italy	1921	507.154	15.230	1.228	16.458	4.057
Korea, Rep. of	1541	588.049	5.571	1.274	6.845	2.616
Lithuania	1587	539.155	7.803	1.736	9.539	3.089
Netherlands	1295	510.988	19.207	0.937	20.144	4.488
Norway (5)	1899	535.958	10.133	1.776	11.909	3.451
Portugal	1612	508.763	9.556	2.595	12.150	3.486
Qatar	1486	463.042	71.953	2.215	74.168	8.612
Russian Federation	2128	567.273	16.354	1.026	17.380	4.169
Singapore	1881	599.129	26.139	0.209	26.348	5.133
Slovak Republic	1610	511.589	22.881	0.930	23.811	4.880
Spain	1670	514.435	18.437	0.418	18.855	4.342
Sweden	1697	522.709	40.072	1.384	41.455	6.439
United Arab Emirates	2243	484.822	71.684	1.845	73.529	8.575
United States	1652	535.459	28.397	0.828	29.224	5.406

Appendix 14F: Summary Statistics and Standard Errors for Proficiency in Mathematics and Science for the Grade 8 Bridge Samples

Summary Statistics and Standard Errors for Proficiency in Overall Mathematics—Grade 8 Bridge Samples

Country	Sample Size	Overall Mathematics				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Chile	1526	434.152	9.803	0.969	10.772	3.282
Chinese Taipei	1578	617.958	28.118	1.150	29.267	5.410
England	1592	525.666	36.296	0.134	36.430	6.036
Georgia	1314	452.235	46.725	3.380	50.105	7.078
Hong Kong SAR	1423	581.128	45.687	1.378	47.065	6.860
Hungary	1751	521.043	50.453	0.622	51.075	7.147
Israel	1863	511.454	48.390	1.688	50.078	7.077
Italy	2032	495.350	11.582	0.395	11.977	3.461
Korea, Rep. of	1693	613.440	12.601	0.315	12.917	3.594
Lithuania	1687	509.663	26.804	0.380	27.184	5.214
Malaysia	1560	473.342	95.148	3.416	98.565	9.928
Norway (9)	2018	509.421	14.338	0.741	15.079	3.883
Qatar	1490	452.250	33.261	6.955	40.216	6.342
Russian Federation	2083	542.804	55.946	0.125	56.071	7.488
Singapore	1871	630.347	41.455	1.303	42.758	6.539
Sweden	1582	513.370	22.833	0.604	23.437	4.841
Turkey	1819	486.938	49.058	1.597	50.654	7.117
United Arab Emirates	2089	481.636	74.089	0.733	74.822	8.650
United States	1484	511.773	40.469	0.422	40.892	6.395

Summary Statistics and Standard Errors for Proficiency in Overall Science—Grade 8 Bridge Samples

Country	Sample Size	Overall Science				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Chile	1526	458.329	11.511	2.527	14.038	3.747
Chinese Taipei	1578	584.185	20.317	4.531	24.849	4.985
England	1592	529.216	23.825	1.464	25.289	5.029
Georgia	1314	434.787	47.700	8.045	55.745	7.466
Hong Kong SAR	1423	530.579	25.505	2.365	27.870	5.279
Hungary	1751	523.286	40.031	1.998	42.029	6.483
Israel	1863	497.660	46.912	1.923	48.834	6.988
Italy	2032	487.430	17.480	0.908	18.389	4.288
Korea, Rep. of	1693	562.584	10.185	2.708	12.893	3.591
Lithuania	1687	522.169	17.423	2.561	19.984	4.470
Malaysia	1560	469.391	116.807	1.359	118.166	10.870
Norway (9)	2018	500.464	20.316	1.715	22.031	4.694
Qatar	1490	494.777	23.676	2.366	26.042	5.103
Russian Federation	2083	543.813	36.668	0.693	37.361	6.112
Singapore	1871	611.040	37.081	0.393	37.474	6.122
Sweden	1582	521.140	34.419	2.223	36.642	6.053
Turkey	1819	517.604	37.650	3.187	40.837	6.390
United Arab Emirates	2089	490.235	97.258	0.786	98.044	9.902
United States	1484	523.665	33.892	2.145	36.037	6.003

CHAPTER 15

Using Scale Anchoring to Interpret the TIMSS 2019 Achievement Scales

Ina V.S. Mullis
Bethany Fishbein

Introduction

As described in Chapter 12: Implementing the TIMSS 2019 Scaling Methodology, the TIMSS 2019 achievement results are summarized using item response theory (IRT) scaling. Countries' average achievement scores are reported on the TIMSS achievement scales for mathematics and science, with most average achievement scores ranging from 300 to 700. Average achievement provides data users with information about how achievement compares among countries and whether scores are improving or declining over time.

To provide as much information as possible for policy and curriculum reform, however, it is important to describe the mathematics and science competencies associated with different locations within the range of scores on the achievement scales. For example, in terms of levels of proficiency in mathematics and science, what does it mean for a country to have average achievement of 513 or 426, and how different are these scores?

The TIMSS 2019 International Benchmarks provide information about what students know and can do at different points along the achievement scales. More specifically, TIMSS has identified four points along the achievement scales to use as international benchmarks of achievement—Advanced International Benchmark (625), High International Benchmark (550), Intermediate International Benchmark (475), and Low International Benchmark (400). For each assessment and International Report, the TIMSS & PIRLS International Study Center works with the expert international committee that guides assessment development, Science and Mathematics Item Review Committee (SMIRC), to conduct a scale anchoring analysis to describe student competencies at the TIMSS International Benchmarks.

This chapter describes the scale anchoring procedures that were applied to update the descriptions of student performance at the international benchmarks from TIMSS 2015 to TIMSS 2019. The analysis was conducted separately for mathematics and for science at fourth and eighth grades. In brief, scale anchoring involved conducting a scale anchoring analysis to identify items that students scoring at the international benchmarks answered correctly, and then having experts examine the content of each item to determine the kind of knowledge, skill, or reasoning demonstrated by students who responded correctly to the item. The experts then summarized the detailed list of item competencies in a brief description of achievement at each international benchmark. Thus, the scale anchoring procedure yielded a content-referenced interpretation of the achievement results that can be considered in light of the TIMSS 2019 frameworks for assessing mathematics and science.

Classifying the Items

As the first step, students scoring within 5 scale-score points of each benchmark (i.e., the benchmark point plus or minus 5) were identified for the benchmark analysis. This 10-point range provided an adequate sample of students scoring at the benchmark, and yet was small enough so that performance at one international benchmark was still distinguishable from the next. The score ranges around each international benchmark and the number of students scoring in each range are shown in Exhibit 15.1.

Exhibit 15.1: Range Around Each TIMSS 2019 International Benchmark and Number of Students Within Each Range

		Low (400)	Intermediate (475)	High (550)	Advanced (625)
Range of Scale Scores		**395–405**	**470–480**	**545–555**	**620–630**
Grade 4 Mathematics (includes Less Difficult)	eTIMSS	3,255	6,798	8,222	4,248
	paperTIMSS	3,876	6,310	6,485	3,140
	Less Difficult	3,792	3,718	1,905	428
	Total	**10,923**	**16,826**	**16,612**	**7,816**
Grade 4 Science	eTIMSS	2,899	6,738	8,939	4,297
	paperTIMSS	3,513	6,062	6,736	2,805
	Total	**6,412**	**12,800**	**15,675**	**7,102**
Grade 8 Mathematics	eTIMSS	2,918	4,700	4,513	2,617
	paperTIMSS	4,468	4,464	3,513	1,831
	Total	**7,386**	**9,164**	**8,026**	**4,448**
Grade 8 Science	eTIMSS	2,485	4,531	5,066	3,092
	paperTIMSS	3,676	4,412	3,990	2,032
	Total	**6,161**	**8,943**	**9,056**	**5,124**

paperTIMSS sample sizes include bridge samples from eTIMSS countries.
Data analysis was conducted before achievement scaling completely finalized.

The second step involved computing the percentage of those students scoring in the range around each international benchmark that answered each item correctly. To compute these percentages, students in each country were weighted proportionally to the size of the student population in the country. For multiple-choice items and constructed response items worth 1 point, it was a straightforward matter of computing the percentage of students at each benchmark who answered each item correctly. For constructed response items scored for partial and full credit, percentages were computed for students receiving partial credit (1-point) as well as for students receiving full credit (2-points). Because the students in about half the countries took the items in the eTIMSS format and the other half in the paper format, first the percent corrects were computed separately for computer-based and paper, including the bridge data. A comparison showed these to be very similar (recall that the detailed comparison between modes found more than 80% of the items to be mode invariant, see Chapter 12). So for most the items, the two percentages, one for computer-based and one for paper, were averaged for the scale anchoring analysis. For the remaining 20 percent or fewer items with differences in the percentages between the eTIMSS and paper format, the eTIMSS percentage was used as the better bridge to future assessments.

Third, the criteria described below were applied to identify the items that anchored at each benchmark. An important feature of the scale anchoring method is that it yields descriptions of the performance demonstrated by students reaching each of the international benchmarks on the scales, and that the descriptions reflect demonstrably different accomplishments by students reaching each successively higher benchmark. Because the process entails the delineation of sets of items that students at each international benchmark are likely to answer correctly and that discriminate between one benchmark and the next, the criteria for identifying the anchor items considers performance at adjacent benchmarks.

For multiple-choice items, 65 percent answering correctly was used as the criterion for anchoring at each benchmark being analyzed, since students would be likely (about two-thirds of the time) to answer the item correctly. In addition, a criterion of less than 50 percent was used for the next lower benchmark, because with this response probability, students were more likely to have answered the item incorrectly than correctly. A somewhat less strict criterion was used for the constructed response items, because students had much less scope for guessing. For constructed response items, the criterion of 50 percent answering correctly was used for the benchmark without any discrimination criterion for the next lower benchmark.

Using a multiple-choice items as an example, the criteria for each benchmark are outlined below:

- A multiple-choice item anchored at the Low International Benchmark (400) if at least 65 percent of students scoring in the range around the Low International Benchmark (395–405) answered the item correctly. Because this was the lowest benchmark described, there were no further criteria.

- A multiple-choice item anchored at the Intermediate International Benchmark (475) if at least 65 percent of students scoring in the range answered the item correctly, and less than 50 percent of students at the Low International Benchmark answered the item correctly.
- A multiple-choice item anchored at the High International Benchmark (550) if at least 65 percent of students scoring in the range answered the item correctly, and less than 50 percent of students at the Intermediate International Benchmark answered the item correctly.
- A multiple-choice item anchored at the Advanced International Benchmark (625) if at least 65 percent of students scoring in the range answered the item correctly, and less than 50 percent of students at the High International Benchmark answered the item correctly.

To include all of the multiple-choice items in the anchoring process and provide information about content domains and cognitive processes that might not otherwise have been represented by many anchor items, the concept of items that "almost anchored" was introduced. These were items that met slightly less stringent criteria for being answered correctly. The criteria to identify multiple-choice items that "almost anchored" were that 60 to 65 percent of students scoring in the range answered the item correctly and less than 50 percent of students at the next lowest benchmark answered the item correctly. To be completely inclusive for all items, items that met only the criterion that 60 to 65 percent of the students answered correctly (regardless of the performance of students at the next lower point) were also identified. The categories of items were mutually exclusive, and ensured that all of the items were available to inform the descriptions of student achievement at the anchor levels. A multiple-choice item was considered to be "too difficult" to anchor if less than 60 percent of students at the advanced benchmark answered the item correctly. A constructed response item was considered to be "too difficult" to anchor if less than 50 percent of students at the advanced benchmark answered the item correctly.

Exhibit 15.2 presents the number of TIMSS 2019 mathematics and science items that anchored at each international benchmark. A description of the items for mathematics at the fourth grade, science at the fourth grade, mathematics at the eighth grade, and science at the eighth grade can be found in Appendices 15A, 15B, 15C, and 15D, respectively. It should be noted that a partial credit item can anchor twice, typically at a higher benchmark for full credit (2 of 2 points), and a lower benchmark for partial credit (1 of 2 points), but sometimes both anchored at the same level. Only the full credit anchoring results were used to write the benchmark descriptions. For the mathematics scale anchoring at the fourth grade, TIMSS took advantage of data from the less difficult assessment items in developing the descriptions for the Low and Intermediate Benchmarks.

Exhibit 15.2: Number of Items Anchoring and Almost Anchoring at Each TIMSS 2019 International Benchmark

Content Domain	Low (400)	Intermediate (475)	High (550)	Advanced (625)	Above Advanced	Total
Grade 4 Mathematics						
Number	28	39	42	27	6	142
Measurement and Geometry	10	15	14	32	5	76
Data	12	14	17	10	3	56
Grade 4 Mathematics Total*	**50**	**68**	**73**	**69**	**14**	**274**
Grade 4 Science						
Life Science	2	15	28	24	8	77
Physical Science	1	15	25	17	4	62
Earth Science	1	2	14	15	3	35
Grade 4 Science Total	**4**	**32**	**67**	**56**	**15**	**174**
Grade 8 Mathematics						
Number	1	7	26	29	3	63
Algebra	0	2	24	29	7	55
Geometry	0	3	16	20	10	39
Data and Probability	0	8	12	14	6	34
Grade 8 Mathematics Total	**1**	**20**	**78**	**92**	**26**	**217**
Grade 8 Science						
Biology	3	16	37	23	10	89
Chemistry	1	5	13	16	11	46
Physics	1	5	15	25	8	54
Earth Science	1	4	15	14	10	44
Grade 8 Science Total	**6**	**30**	**80**	**78**	**39**	**233**

* Grade 4 Mathematics includes less difficult items at the Low and Intermediate Benchmarks.

Writing the Scale Anchoring Descriptions

Due to COVID-19, an online scale anchoring for TIMSS 2019 was conducted in the spring of 2020, instead of an in-person meeting of the Mathematics and Science SMIRC. In preparation for review by SMIRC, staff at the TIMSS & PIRLS International Study Center used examples from previous assessments to draft short descriptions of the student competencies demonstrated by a correct (or partially correct) response to each mathematics and science item. Then, the mathematics and science items were organized separately by grade, grouped by international benchmark, and within each benchmark the items were sorted by content area. The final categorization was by the anchoring criteria the items met—items that anchored, followed by items that almost anchored, then by items that met only the 60 to 65 percent criteria. Also, in addition to the short draft descriptions, the following information was included for each item: framework classification, answer key or scoring guide, secure status, percent correct at each benchmark, and overall international percent correct. Beyond the item-by-item descriptions and anchoring data, SMIRC members were given the benchmark descriptions from TIMSS 2015.

The members of SMIRC committee 1) worked through each item to review/revise the description of the student competencies demonstrated by a correct (or a partially correct) response, 2) updated the TIMSS 2015 summaries of the proficiency demonstrated by students reaching each international benchmark for publication in the TIMSS 2019 report, and 3) selected example items from TIMSS 2019 that supported and illustrated the benchmark descriptions to illustrate the types of items answered correctly by students at each of the four benchmarks..

Following the SMIRC review, the descriptions and example items were included in the TIMSS 2019 International Report for review by the TIMSS 2019 National Research Coordinators at their 8th meeting in July 2020 (scheduled for Prague, but held virtually).

Appendix 15A: Grade 4 Mathematics Item Descriptions Developed During the TIMSS 2019 Benchmarking

Items at Low International Benchmark (400)	
Number	
N01_02	Solves a word problem involving subtraction of a one-digit number from a two-digit number
N01_05	Multiplies a one-digit number by a two-digit number
N03_01	Orders four three-digit numbers
N03_02	Solves a word problem involving division of a two-digit number by a one-digit number
N03_03A	Identifies the largest of four three-digit numbers in context
N05_01	Identifies a four-digit number represented in words
N07_03	Solves a word problem involving subtraction of a one-digit number from a three-digit number
N07_04	Writes a number between two two-digit numbers
N09_02	Solves a word problem involving addition of two two-digit numbers
N11_02	Solves a word problem involving addition of two- and three-digit numbers
N11_04	Multiplies a three-digit number by a one-digit number
N12_01	Identifies the smallest of four four-digit numbers
N12_02	Identifies a multiple of a one-digit number
N13_01	Adds two two-digit numbers
N13_02	Divides a two-digit number by a one-digit number
N14_01	Supplies one factor of a two-digit number (1 of 2 points)

Items beginning with "N" are items unique to less difficult mathematics.

N04_02	Divides a two-digit number by a one-digit number with a remainder
N07_02	Multiplies a one-digit number by a two-digit number
N04_04	Implements a one-step rule to generate the next number in a pattern
N07_07	Finds the missing term in an addition number sentence
N12_05	Identifies the missing term in an addition sentence
N12_06	Implements a one-step rule forward and in reverse to partially complete a table (1 of 2 points)
N14_04	Solves for the missing number in a multiplication sentence
N14_05	Identifies the operation for an expression that represents a situation
N01_07	Writes a fraction larger than a given unit fraction
N03_06	Recognizes a unit fraction represented pictorially
N04_07	Solves a word problem involving addition of one-place decimals
N09_08	Recognizes a non-unit fraction represented pictorially
Measurement and Geometry	
N03_09	Solves a word problem involving addition of three one-digit numbers
M03_10	Relates a specified face of a cube to its net
N01_08	Identifies a cylinder
N03_08	Writes the names of four common two-dimensional shapes
N05_10	Completes a rectangle on a square grid
N11_11A	Identifies the tallest of four rectangular prisms represented pictorially

Items beginning with "N" are items unique to less difficult mathematics.

N12_10	Completes a symmetric figure on a square grid given half the shape and the line of symmetry
N14_09	Determines the number of lateral faces of a given rectangular prism
N09_10	Identifies a cube
N11_11B	Identifies the greatest volume of four rectangular prisms represented pictorially
Data	
M08_12	Labels bars on a bar graph to represent data given in a table
M10_11	Represents data from a table in a bar graph
N01_04A	Reads data from a bar graph
N04_13	Represents data from a table in a bar graph (1 of 2 points)
N04_14	Evaluates statements about data in a pictograph (1 of 2 points)
N05_05A	Reads data from a table
N05_05B	Compares data presented in a table
N07_05	Uses data from a table to complete a bar graph (2 of 2 points)
N07_05	Uses data from a table to partially complete a bar graph (1 of 2 points)
N13_04A	Reads data from a bar graph
N14_13A	Reads data from a table
N13_04B	Compares data presented on a bar graph

Items beginning with “N” are items unique to less difficult mathematics.

Items at Intermediate International Benchmark (475)	
Number	
M02_01	Identifies a three-digit number given the place values of two of its digits
M05_01	Adds a four-digit and a three-digit number
M06_02	Divides a three-digit number by a one-digit number
M09_02	Determines a four-digit number given the place values of the digits
M11_03	Multiplies a one-digit number by a three-digit number
M12_01	Identifies a four-digit number given in words in expanded form
N01_01	Subtracts a two-digit number from a three-digit number
N01_03	Determines a four-digit number given the values of two of its digits
N03_03B	Justifies the greatest number if one of four numbers is increased by 100
N05_02	Solves a two-step word problem involving subtraction of one- and two-digit numbers
N05_03	Solves a word problem involving division of a two-digit number by a one-digit number
N07_06	Finds the missing term in an addition word problem
N09_04	Determines the smallest 3-digit number with three given digits
N12_04	Solves a multi-step word problem involving multiplication of one-digit numbers and comparison of two-digit numbers
N13_03	Compares representations of a value in numbers and in words and explains answer
N13_05	Solves a multi-step word problem involving addition and subtraction of one- and two-digit numbers and explains answer
M03_01	Identifies a four-digit number given in expanded form

Items beginning with "N" are items unique to less difficult mathematics.

M04_01	Subtracts a three-digit number from a four-digit number
N11_06	Solves a multi-step word problem involving multiplication of one-digit numbers and addition of two-digit numbers
M09_01	Adds a four-digit, three-digit, and two-digit number
N03_05	Solves a word problem involving subtraction of one- and two-digit numbers
N04_01	Solves a word problem involving subtraction of two three-digit numbers
N07_01	Solves a word problem involving multiplication of one- and two-digit numbers
N09_01	Subtracts a two-digit number from a three-digit number
N09_03	Solves a word problem involving multiplication of one- and two-digit numbers
N11_01	Identifies a four-digit number given the digits in two places
N11_03	Divides a two-digit number by a one-digit number with a remainder
M04_03	Identifies the missing number in two number sentences with inequalities
M13_05	Solves for the missing number in a subtraction sentence
N04_06	Identifies an expression with addition and subtraction that represents a situation
N09_06	Determines the missing number in a well defined number pattern
N12_06	Implements a one-step rule forward and in reverse to complete a table (2 of 2 points)
M11_06	Determines the operation to complete a number sentence
N05_04	Identifies an arithmetic operation that represents a situation
N11_05	Identifies an expression that represents a situation
N05_07	Supplies a fraction larger than one half

Items beginning with “N” are items unique to less difficult mathematics.

N05_08	Orders two decimals and a whole number
N13_07	Solves a word problem involving addition of decimals
M01_02	Identifies the representation of a non-unit fraction
Measurement and Geometry	
M08_06	Measures the vertical height of an object with a ruler
N04_08	Solves a multi-step word problem involving multiplication and addition of mass
N05_11	Determines the number of unit cubes to fill a rectangular prism
N05_12	Solves a word problem involving addition of hours and minutes
N13_10	Determines the perimeter of a triangle given the side lengths
N13_11	Solves a word problem involving addition of hours and minutes
N14_08	Identifies the appropriate metric unit of volume for an object
N03_10	Solves a word problem involving subtraction of hours and minutes
N09_09	Identifies the appropriate metric unit of measurement for an object
M01_07	Identifies the number of triangular faces in a given three-dimensional shape
M02_09	Completes a symmetric figure on a square grid given half the shape and the line of symmetry
N13_08	Identifies a square
M01_06B	Identifies a street perpendicular to a given street
N05_09	Identifies a shape with equal angles
N07_10	Identifies a common shape inside another common shape

Items beginning with “N” are items unique to less difficult mathematics.

Data	
M01_09A	Identifies the greatest value in a bar graph
M02_10A	Reads data from a line graph
M10_12A	Represents data in a table
M12_10	Identifies a title and axis labels for a bar graph
M13_10A	Reads data from a graph
N04_12	Reads data from a tally chart
N04_13	Represents data from a table in a bar graph (2 of 2 points)
N14_12	Labels a bar on a bar graph given data in a tally chart
M14_10	Solves a word problem involving reading data from a table
N12_11	Reads data from a table
N12_12A	Identifies the label for a bar in a bar graph given the data in a table
N14_11	Reads data from a table
M04_12	Determines one or two out of three missing values in a table given conditions for the data (1 of 2 points)
N09_05	Uses data from a bar graph to solve a problem

Items beginning with “N” are items unique to less difficult mathematics.

Items at High International Benchmark (550)	
Number	
M01_01	Identifies the set of numbers having a given number as a factor
M02_02	Solves a word problem involving division of a two-digit number by a one-digit number
M03_02	Multiplies a two-digit number by a two-digit number
M05_02	Identifies the number closest in size to a given four-digit number
M05_03	Solves a word problem involving division
M06_01	Classifies two- and three-digit numbers as even or odd
M06_03	Devises one way of grouping objects that satisfy two conditions (1 of 2 points)
M07_01	Solves a multi-step word problem involving multiplication and addition of whole numbers
M07_03	Solves a word problem involving multiplication of two-digit numbers
M08_01	Solves a word problem involving division of a two-digit number by a one-digit number with a remainder
M08_02	Solves a word problem involving subtraction of a two-digit number from a four-digit number
M09_03	Devises one way to allocate money in a given context (1 of 2 points)
M10_02	Solves a word problem involving multiplication of a three-digit number by a one-digit number
M11_01	Rounds a four-digit number to the thousands place
M11_04	Solves a problem set in a novel situation involving addition and comparison of whole numbers and justifies the solution
M11_09	Given two positions on a curved path, follows specified moves and labels another position (1 of 2 points)
M12_02	Solves a word problem involving addition of one-, two-, and three-digit numbers

M13_03	Identifies a true statement about comparison of two- and three-digit numbers
M03_09	Finds the distance between two positions on a number line
M11_02	Identifies a number that satisfies two conditions of multiples
M09_08	Finds the halfway point between two positions on a number line
M14_02	Solves a word problem involving subtraction of two three-digit numbers
M01_05	Follows a two-step rule to extend a number pattern
M03_06	Solves for a repeated missing number in a subtraction sentence
M06_06	Determines the operation to complete a number sentence with operations on both sides
M06_07	Identifies an expression that represents a situation
M07_05	Solves a number sentence involving division
M08_03	Determines the place of three numbers in a number sentence with operations on both sides
M09_06	Identifies an expression that represents a situation
M12_04A	Implements a two-step arithmetic rule to a given number
M13_02	Determines whether three pairs of numbers follow a given two-step rule
M13_07	Follows a two-step rule to generate the next number in a pattern
M14_03	Identifies an expression with multiplication and subtraction that represents a situation
M14_04	Determines the one-step rule that relates five numbers in a pattern
M02_04	Identifies the operation for an expression that represents a situation
M03_07	Identifies an expression that represents a situation

M01_04A	Solves a word problem involving rectangular representations of fractions
M01_04B	Solves a word problem involving rectangular representations of fractions
M06_04	Solves a word problem involving subtraction of a non-unit fraction from 1
M07_06	Adds a whole number and a two-place decimal
M09_05	Solves a word problem involving subtracting one-place decimals
M14_05	Solves a word problem involving addition of three two-place decimals
Measurement and Geometry	
M02_06	Identifies the appropriate metric unit of measurement for three objects
M04_06A	Solves a word problem involving multiplication of lengths
M05_04	Solves a word problem involving addition of time
M05_11	Solves a problem by filling a three-dimensional shape with rectangular solids
M14_07	Solves a word problem involving conversion of minutes to hours
M10_06	Estimates the total length of an object given the length of part of it
M12_06	Estimates the length of an object given the length of another object
M11_08	Classifies angle types in a figure
M12_05	Identifies a pair of shapes that make a rectangle
M13_09	Identifies a two-dimensional view of an irregular three-dimensional figure
M14_09	Completes a shape on a square grid to satisfy four conditions
M05_08	Identifies a shape with a right angle

M07_09	Identifies the two-dimensional view of a three-dimensional object
M09_09	Identifies a solid given two faces
Data	
M02_11	Represents data from a table in a pictograph
M04_10B	Compares data in a table
M05_12	Completes a bar graph using information from a pictograph
M07_12	Identifies a pie chart that has the same information as a bar graph
M08_11	Determines whether questions can be answered with data in a table
M08_13	Determines the key for a pictograph and uses it to complete the graph (1 of 2 points)
M10_10A	Reads data from a line graph
M11_11A	Uses a key to retrieve data from a pictograph
M12_09A	Uses a key to retrieve data from a pictograph
M12_11	Labels sections of a pie chart given three conditions for the data
M14_11A	Evaluates statements about data in a bar graph
M14_13	Completes a table for a given data series
M06_11	Identifies a pie chart that represents given data
M07_11	Uses information from a bar graph to solve a problem
M10_10B	Extrapolates a point on a line graph with constant slope
M10_12B	Interprets data in a table to solve a problem
M13_10B	Extrapolates from a graph to solve a problem

Items at Advanced International Benchmark (625)	
Number	
M01_03	Solves a multi-step problem involving division and gives a reason for their answer
M03_03	Solves a multi-step word problem involving addition and subtraction of two- and three-digit numbers
M04_02	Determines a number that meets two conditions of multiples and one condition of order
M06_03	Devises two ways of grouping objects that satisfy two conditions (2 of 2 points)
M09_03	Devises two ways to allocate money in a given context (2 of 2 points)
M10_03	Solves a multi-step word problem involving division of one- and two-digit numbers with remainders
M11_09	Given two positions on a curved path, follows specified moves and labels another position (2 of 2 points)
M13_01	Recognizes equivalent three-digit numbers written in expanded form
M14_01	Divides a three-digit number by a one-digit number
M09_07	Identifies the missing number in a number sentence with operations on both sides
M12_04B	Implements a two-step rule in reverse to generate the previous number in a pattern
M13_06	Identifies an operation that represents a situation
M04_04	Identifies an expression with division and addition that represents a situation
M10_04	Solves a word problem by extending a pattern
M11_07	Identifies number sentence that represents a situation
M02_03	Identifies non-unit fractions greater than a given unit fraction
M03_04	Solves a problem to identify a fraction that represents the shaded portion of a figure

M03_05	Solves a word problem involving division with a remainder (1 of 2 points)
M05_05	Identifies a fraction equivalent to a given fraction
M05_06	Solves a multi-step problem involving fractions
M06_05	Draws a complete shape on a grid given a picture of a fraction of the shape
M07_02	Identifies a fraction equivalent to a one place decimal
M08_05	Identifies a decimal given the place values of two of its digits
M10_05	Identifies the larger number among pairs of one- and two-place decimals
M11_05	Solves a word problem involving adding fractions with different denominators
M13_04	Identifies a number between a one-place decimal and two-place decimal
M07_04	Identifies a set of objects with a given fraction shaded
Measurement and Geometry	
M02_08	Determines the number of three different shapes that cover the area of a square (2 of 2 points)
M02_08	Determines the number of two different shapes that cover the area of a square (1 of 2 points)
M04_06B	Solves a word problem involving division of lengths
M05_10	Reads a ruler to find the length of one side of an equilateral triangle and finds its perimeter
M06_08	Reads a ruler to find the length of a line segment beginning and ending at half-units
M07_07	Reads a ruler to find the length of an object beginning at a half-unit
M08_07	Solves a word problem involving subtraction of volumes
M08_10	Determines two sides lengths of a hexagon given the other four side lengths and its perimeter

M10_07	Draws a rectangle with a given perimeter on a square grid
M10_09	Solves a multi-step word problem involving multiplication and division of lengths, with a remainder (1 of 2 points)
M12_07	Determines the number of unit cubes to fill a rectangular prism and explains method (2 of 2 points)
M12_07	Determines the number of unit cubes to fill a rectangular prism and explains method (1 of 2 points)
M12_08	Measures a horizontal object with a ruler and applies a scale to determine its length
M14_06	Identifies the speed shown on a speedometer
M02_07	Solves a multi-step word problem involving multiplication and division of weights
M04_05	Estimates the height of an object given the height of another object
M04_07	Determines the weight of an object given a series of three balanced scales
M07_10	Finds the area of a rectangle given its dimensions
M08_08	Analyzes information in a time table to solve a word problem
M14_08	Identifies the area of a rectangle drawn at an angle on a square grid
M03_08	Identifies parallel lines on a geometric shape
M04_08	Draws a line on a square grid that is parallel to a given line and passes through a specified point
M04_09	Identifies properties of two pentagons
M06_09	Recognizes acute angles in an irregular quadrilateral
M06_10	Determines the number of square and triangular faces of three-dimensional shapes (2 of 2 points)
M06_10	Determines the number of square and triangular faces of three-dimensional shapes (1 of 2 points)
M07_08	Given a line, draws another line to form an angle less than a right angle

M09_10A	Draws a parallel line on a square grid given conditions
M13_08	Draws an obtuse angle on a square grid given one side
M01_06A	Identifies a street parallel to a given street
M05_09	Identifies a shape that has both line and rotational symmetry
M11_10	Identifies a net of a hexagonal prism
Data	
M02_12	Determines the y-axis scale for a bar graph given the data in a table
M03_11	Represents data from a table in a pie chart
M08_13	Determines the key for a pictograph and uses it to complete the graph (2 of 2 points)
M14_12	Identifies the optimal data display for data given in a table
M01_09B	Interprets a bar graph to solve a two-step problem
M04_12	Determines three missing values in a table given conditions for the data (2 of 2 points)
M11_11B	Uses information in a pictograph to solve a problem
M12_09B	Uses data from a pictograph to solve a problem
M12_12	Compares the slope of two lines on a graph for a specific period and explains answer
M14_11B	Uses data from a bar graph to determine whether a conclusion is true and explains answer.

Items Above Advanced International Benchmark (625)	
Number	
M02_05	Solves a multi-step word problem involving unit price (2 of 2 points)
M02_05	Solves a multi-step word problem involving unit price (1 of 2 points)
M09_04	Determines the missing digit for a two-digit number that satisfies two conditions
M08_04	Solves a word problem by implementing a one-step rule to generate numbers in a pattern
M03_05	Solves a word problem involving division with a remainder and justifies the solution (2 of 2 points)
M12_03	Adds a one-place decimal and a two-place decimal
Measurement and Geometry	
M10_09	Solves a multi-step word problem involving multiplication and division of lengths, with a remainder (2 of 2 points)
M01_08	Identifies a net of a given object
M05_07	Identifies a rule to sort shapes into two sets
M09_10B	Draws a perpendicular line on a square grid given conditions
M10_08	Identifies properties of a parallelogram and a rectangle
Data	
M04_11	Analyzes data in two bar graphs to refute a conclusion
M08_14	Represents data from a table in a line graph (2 of 2 points)
M08_14	Represents data from a table in a line graph (1 of 2 points)

Appendix 15B: Grade 4 Science Item Descriptions Developed During the TIMSS 2019 Benchmarking

Items at Low International Benchmark (400)	
Life Science	
S06_01	Recognizes an animal that has a backbone
S10_01	Identifies 4 or 5 of 6 animals as birds, insects, mammals, or reptiles (1 of 2 points)
Physical Science	
S04_10	Identifies the most likely material making up a spoon that gets hot sitting in a pot of boiling soup
Earth Science	
S04_12	Identifies natural resources used to grow plants

Items at Intermediate International Benchmark (475)	
Life Science	
S14_05	Explains why a plant kept by a window is healthier than a plant kept in a closed closet
S12_05	Identifies the characteristic used to sort animals into two groups
S04_01	Recognizes the function in common between a hedgehog's spines and snail's shell
S13_03	Analyzes a diagram to explain which flower will grow better
S01_01	Recognizes that in mammals, a male and female of the same kind are needed to reproduce
S11_06	Describes one way a polar bear's fur helps it survive (1 of 2 points)
S03_05	Describes how human heart rate changes during exercise

S06_04	States two reasons why a plant will not survive by analyzing given conditions
S05_04	States one reason why plastic objects in the ocean are dangerous for sea animals
S03_03	Recognizes a living thing that produces its own food (1 of 2 points)
S04_03	Identifies a predator and its prey
S09_02	States two things that plants need from their environment to make their own food
S13_04	Evaluates two diagrams to explain which environment is better for sharks
S10_02	Explains why a person should wash their hands before eating even if they do not appear dirty
S09_01	Recognizes why milk is important in a balanced diet
Physical Science	
S06_06	Recognizes the states of matter of three different materials
S09_08	Classifies materials as solids, liquids, or gases
S10_10	Recognizes an object that could be used to complete a circuit to light a bulb
S13_08	Identifies the best material to complete a circuit
S07_09	Using a model of a flashlight, identifies an object that can be used to complete an electrical connection
S08_08	Predicts what will happen to magnets in plastic and iron cups when the cups are turned upside down
S14_08	Identifies which of six objects can be picked up using a magnet
S10_07	Explains why a flashlight needs batteries in order to turn on
S12_07	Describes what will happen to the temperature of a table where it touches the bottom of a hot cup of tea
S04_07	Identifies the diagram that shows a circuit where a bulb will be lit

S11_10	Identifies why a bulb will not light in a model of an electric circuit
S07_08	Gives a reason why two objects of the same shape and size travel different distances after a push
S09_09	Explains why one object requires more force to start its motion than another
S02_10	Recognizes the best explanation for why a box on a cart is easier to pull than a box resting directly on the floor
S08_07	Recognizes the ramp that will make it easiest to move a heavy box onto a table
Earth Science	
S05_10	Matches each item in a list of Earth's landscape features to its description
S14_11	Recognizes the cause of the movement of sand dunes in a desert

Items at High International Benchmark (550)	
Life Science	
S02_01	Lists two living things and two nonliving things shown in a picture of a desert ecosystem
S02_03	Identifies characteristics that describe either a toy duck and a living duck or only a living duck
S06_02	Describes two ways that a mammal helps its young survive
S03_04	Recognizes a feature of how snakes eat
S01_05	Identifies a function of a plant's stalk by interpreting an observation from an investigation
S02_02	Identifies the human organ with the same function as a fish's gills
S11_03	Completes a diagram describing the stages in the life cycle of a flowering plant
S12_01	Orders the life stages of a butterfly within a diagram

S08_04	Identifies the process that must have occurred in a strawberry plant from two pictures of the plant taken four weeks apart
S14_02	Recognizes another life stage of a caterpillar
S04_05	Identifies the characteristics that a female rabbit's mate must have
S10_05	Recognizes a characteristic of a seedling that will show whether it is a fir tree or a cherry tree
S08_02	Identifies the picture showing the seedling form of an adult plant
S02_06	Describes how holding its tail over its head helps a ground squirrel survive in hot, dry environments
S05_03	Recognizes an advantage of thin, pointed leaves compared to broad, flat leaves
S11_04A	Interprets data from an investigation to recognize the best condition for growing plants
S09_03	Identifies a reason that some mammals pant on hot days
S11_05	Relates factory pollution to its effect on farm fields
S14_01	Identifies which of six animals could live in a desert
S02_04	Explains how an increase in the number of bats in an area could lead to a decrease in the number of insects
S03_02	Uses a list of living things in an Arctic ecosystem to complete a food chain
S08_03	Completes a food chain using three given animals
S06_03A	Uses a food web to identify what a predator eats
S05_05	Provides a possible reason why some trees in a group do not grow as well as others
S01_02	Explains that germs can be transmitted even when people do not appear to be sick
S04_06	Describes how germs can still spread if a person covers their mouth with their hands when they cough
S12_02	Explains how one way of eating ice cream exposes a person to fewer germs than another way of eating ice cream

S09_06	Recognizes a way to avoid spreading the flu
Physical Science	
S05_07	Describes a difference between ice and water in addition to their physical states
S11_07	Identifies a physical property of metal pot that makes it good for boiling water
S01_09	Using a diagram, identifies which hidden object could complete an electric circuit
S01_07	Observes that two metal bars repel and determines whether they are magnets
S12_10	Identifies an explanation for why magnets push against each other when they are brought together
S03_06	Identifies a way to sort objects containing metals
S10_08	Recognizes a pair of carts carrying magnets that will move away from each other
S03_08	Recognizes what happens to the water when a puddle of water on a metal tray becomes smaller
S11_09A	Explains why boiling decreases the amount of water in a container
S11_09B	Predicts the effect on a cold window glass of boiling water nearby
S07_06	States a reason for the color change and surface roughening of a metal object over time
S03_10	Explains why pressing a guitar string stops the sound
S05_09A	Identifies from a diagram how a shadow is formed
S03_07	Analyzes a diagram to identify one way to make a shadow bigger
S04_09	Identifies a graph showing the relationship between increasing the force used to hit a drum and the loudness of the drum's sound
S01_08	Explains that heat in a metal object reaches the nearest point soonest
S12_08	Recognizes the energy change occurring in a circuit with a battery and a lightbulb

S14_09	Explains which of two circuits will have a lit bulb
S02_09	Recognizes the energy change that occurs when a flashlight is turned on
S14_07	Recognizes the energy change that occurs in an electric iron
S08_10	States the force that causes a skydiver and a book to fall
S14_10	Recognizes the force that makes it more difficult to move a sofa on a rug than on a wood floor
S02_08	Recognizes the force that causes a skydiver to fall to Earth
S12_06	Identifies the force that makes an open parachute fall more slowly than a crumpled parachute
S08_09	Identifies a description of how a pulley makes it easier to move a heavy box
Earth Science	
S08_11	Recognizes what covers most of Earth's surface
S14_12	Recognizes the best way to replenish a forest from which wood is taken
S12_12	Explains what a fish fossil reveals about the history of a desert area
S08_12	Interprets data in a table to identify which of two locations is a desert and explains reasoning
S06_10A	Interprets information from a graph to recognize which crops will grow best in an area with given precipitation
S03_12	Recognizes which step in a diagram of a water cycle shows evaporation
S11_01	Recognizes which place is likely to have weather that is hot and wet
S01_11	Recognizes that the solar system is made up of the Sun and its planets
S03_11	Using two pictures of the same location, explains that the Moon can look different at different times
S10_13	Recognizes pictures of shapes the Moon can have

S12_13	Identifies a diagram that represents the Solar System
S07_11	Recognizes a feature of the Moon from observations over a month
S11_02	Recognizes seasons north and south of the Equator
S04_11	Interprets a diagram of the Sun and the Earth to identify the season in a labeled city

Items at Advanced International Benchmark (625)	
Life Science	
S01_04	States two things in addition to water that animals need to survive
S05_01	States one difference between living things and nonliving things
S10_01	Identifies 6 of 6 animals as birds, insects, mammals, or reptiles (2 of 2 points)
S07_01	Recognizes the function of muscles attached to bones
S07_05	Draws a conclusion by relating one function of feathers to keeping a body warm in the case of dinosaurs
S03_01	Recognizes the plant part that produces seeds
S09_04	Identifies a difference in the life cycles of a grasshopper and a butterfly
S04_04	States either an advantage for a dandelion to make many seeds or an advantage for a dandelion to make light, fluffy seeds (1 of 2 points)
S07_03Z	Recognizes whether labeled features of a bird are inherited
S12_03	Identifies an explanation for why laying many eggs is helpful for insects' survival
S01_03B	Identifies a desert plant and describes one feature that helps it survive in the desert
S10_04	States why it is better for a lemur to sit in the sun with its arms outstretched rather than at its sides in order to get warm

S11_06	Describes two ways a polar bear's fur helps it survive (2 of 2 points)
S07_04	Evaluates three experimental designs and explains which is best to test if plants need light to grow
S08_06	States one way that the human body reacts to cold temperatures
S11_04B	Identifies a conclusion about plant growth using data from an investigation
S10_06	Recognizes the ecosystem where a set of living things is most likely to be found
S02_05	Recognizes how a Venus flytrap differs from most other plants
S09_05	Predicts the consequences of removing a predator from an animal's habitat
S12_04	Explains why the number of mice in a town increased after trees were cut down
S10_03	Identifies the food chain that best shows how energy is transferred from the Sun to an owl
S06_03B	Uses a food web to determine which animals are competitors
S13_02	States two ways to avoid catching illness in a crowded space
S13_05	Describes how boiling water makes it safe to drink
Physical Science	
S13_06	Recognizes one property of a liquid
S01_06	Identifies that two objects of the same size and shape have the same volume and, from a diagram, that they have different masses
S06_08	Explains how to separate a mixture of two types of solids of different sizes
S09_07	Predicts which of two objects is a better conductor of heat with supporting explanation
S11_08A	Evaluates the best way to separate a mixture of solids of similar size
S05_08	Identifies that the temperature at which an object melts depends on the material from which it is made

S11_08B	Evaluates the best way to separate a mixture of things that dissolve and things that do not dissolve
S04_08	Predicts how a train car with a magnet attached will move when another train car with a magnet attached is brought towards it
S02_07	Recognizes a change in which the materials in objects stay the same
S06_09A	Recognizes set-ups that will more quickly dissolve a solid in water
S06_09B	Explains the importance of controlling a variable in an experiment
S13_07	Evaluates the best set-up to investigate whether temperature affects the rate at which a solid dissolves in water
S14_06	Identifies a statement describing the change that occurs when water boils
S10_09	Identifies conclusions that are supported by the results of an experiment during which a gas is collected in a balloon
S05_09B	Recognizes that a shadow produced in colored light is black
S09_10	States one form of energy present in a model of an electric circuit (1 of 2 points)
S13_09	Recognizes a diagram that demonstrates motion due to gravity
Earth Science	
S01_10	Identifies the diagram that shows relative amounts of water and land on the Earth's surface
S02_11A	States one advantage of farming near a river
S02_11B	States one disadvantage of farming near a river
S05_11	Identifies how fish fossils are formed
S09_12	From pictures of rock formations, identifies how a given rock may have looked long ago
S10_12	Identifies the best explanation for finding a tropical plant fossil in a cold region
S13_10	Relates two different environments and weathering effects on rocks

S13_12	Interprets information from temperature graphs to identify which of two places has certain climate properties
S06_10B	Synthesizes precipitation information from a graph and diagram to recognize the best area to plant a crop in a given climate
S07_10	Identifies that clouds are made of water droplets
S09_11	Recognizes a diagram showing the correct relative positions of the Earth, Moon, and Sun
S02_12	Places the Earth in a model to show its position relative to the Sun when a labeled city is experiencing summer
S07_12	Interprets a diagram of the Earth and the Sun to describe how Earth turning on its axis causes day and night in a particular location
S08_13	Identifies a picture of a tree and its shadow in the afternoon based on a picture of the tree and its shadow in the morning
S12_11	Interprets a diagram of a man and his shadow to identify the Sun's relative position when his shadow will be shorter

Items Above Advanced International Benchmark (625)

Life Science

S04_02	Identifies which of four animals have backbones
S07_02	Recognizes the main function of leaves on a plant
S04_04	States both an advantage for a dandelion to make many seeds and an advantage for a dandelion to make light, fluffy seeds (2 of 2 points)
S14_04	Explains why a single elephant calf has a better chance of survival than a single frog egg
S01_03A	Explains that to test the survival of plants, they should be compared under different conditions
S05_06	Identifies that more use of public transportation will decrease air pollution in a large city
S03_03	Recognizes a living thing that produces its own food and describes the process (2 of 2 points)
S08_05	Identifies the animal that competes with giraffes for food

Physical Science	
S07_07	Explains why a metal spoon in hot soup feels hotter than a wooden spoon in hot soup
S06_07	Explains the process by which wet objects become dry
S03_09	States one source of energy other than sunlight that can be changed into electricity
S09_10	States two forms of energy present in a model of an electric circuit (2 of 2 points)
Earth Science	
S13_11	Recognizes four true statements about recycling metals
S14_13	Explains one benefit of using sunlight or wind to produce electricity compared to oil or natural gas
S10_11	Interprets data in a table to identify the place where is it most likely to rain

Appendix 15C: Grade 8 Mathematics Item Descriptions Developed During the TIMSS 2019 Benchmarking

Items at Low International Benchmark (400)	
Number	
M09_01	Recognizes a 7-digit number given in words

Items at Intermediate International Benchmark (475)	
Number	
M01_02A	Solves a word problem involving addition of time
M06_01	Solves a word problem involving subtraction of negative numbers
M07_03	Solves a two-step word problem involving whole numbers
M05_02	Solves a word problem involving subtraction of negative numbers
M04_03	Determines whether a series of decimals are greater than, less than, or equal to fractions (1 of 2 points)
M06_02	Identifies equivalent ratios
M10_05	Given a ratio, represents an equivalent ratio pictorially
Algebra	
M04_07A	Solves a word problem involving an inequality and explains answer
M08_08A	Extends a given geometric pattern to supply the value of the 7th term
Geometry	
M02_10	Determines the value of an angle in an irregular quadrilateral given the values of the other angles

M14_10	Identifies the reflections of irregular shapes
M11_11	Determines the total number of stacked unit cubes
Data and Probability	
M05_12	Finds and compares the unit prices of four objects
M05_14	Identifies the bar graph that matches the information shown in a table
M07_12A	Compares data from two line graphs to solve a problem
M07_12B	Reads data from a line graph
M09_12A	Calculates mean and median for one ordered lists of data (1 of 2 points)
M13_11	Evaluates information given by a time/distance graph
M01_12	Solves a problem given the chance of an outcome
M03_14A	Estimates an expected value given an observed sample

Items at High International Benchmark (550)	
Number	
M03_01	Identifies an expression equivalent to a given division expression
M05_01	Evaluates an expression involving negative whole numbers and parentheses
M05_05A	Solves a word problem involving multiplication and addition of whole numbers
M10_01	Adds two numbers with different exponents and bases
M11_02	Solves a word problem involving division of whole numbers with a remainder

M14_01	Identifies the number with the most factors
M01_03	Understands a property of adding multiples
M01_04	Writes a decimal with three places as a fraction
M03_03	Finds the missing value in an addition problem with both fractions and decimals
M04_03	Determines whether a series of decimals are greater than, less than, or equal to fractions (2 of 2 points)
M05_04	Given the two parts of a whole in a word problem, identifies the fraction which represents one part
M07_04	Determines what fraction of a 10X10 grid is shaded
M13_02	Solves a two-step word problem involving subtraction of whole numbers and multiplication of a fraction
M14_02	Determines the numerator that makes two fractions equivalent
M01_01	Identifies the representation of a fraction equivalent to a given representation of a fraction
M08_02	Adds two decimals represented in words
M11_01	Solves a word problem involving a fraction of a whole
M12_03	Identifies a decimal equivalent to the sum of two fractions with denominators that are powers of ten
M04_04	Solves a word problem involving a fraction of a whole
M02_04	Solves a word problem involving a three-part ratio
M03_04	Shades a percent of a figure
M10_04	Solves a word problem involving ratios
M13_04	Solves a word problem involving ratios and decimals
M14_04A	Determines a ratio to model a situation

M14_04B	Determines a ratio to model a situation
M07_01	In a word problem for dividing a given quantity in a given ratio, determines the quantity of one of the parts.
Algebra	
M01_06	Identifies the equivalent algebraic expression involving exponents and multiplication
M02_07	Solves a word problem involving evaluating a formula with exponents
M10_06	Evaluates an expression with two variables
M10_09	Solves a pair of simultaneous linear equations in two variables
M12_06	Evaluates a formula with an exponent
M12_08	Solves a word problem involving simultaneous linear equations in two variables
M14_05	Identifies an expression that represents a situation
M14_06	Solves a linear equation involving fractions
M14_07	Solves a word problem involving evaluating a formula with two variables
M05_07	Evaluates an algebraic expression involving fractions and integers
M08_05	Evaluates an expression with a square root and two variables with exponents
M08_07	Solves a multi-step word problem involving linear inequalities
M11_06	Identifies an equation that models a situation
M02_05	Evaluates an expression with two variables
M03_07	Identifies an algebraic expression that represents the perimeter of an irregular shape
M04_06	Evaluates an equation with three variables

M07_06	Evaluates the power of an expression given its value
M11_07	Identifies an expression for the area of part of a geometric figure
M13_05	Solves a linear equation in two-variables given the value of one variable
M01_07A	Extends a given geometric pattern to find the value of the 10th term
M03_08	Determines a missing coordinate for a linear relationship given in a table
M04_08A	Extends a given geometric pattern to supply the value of the 10th term
M11_08	Uses values for a linear function to determine an extrapolated value
M13_07	Identifies the true statement about a linear relationship given in a graph
Geometry	
M02_11	Compares properties of two open cylinders made by rolling the same rectangle in different directions
M02_12	Determines the coordinates of a trapezoid's missing vertex given a congruent trapezoid in the Cartesian plane
M03_10	Finds the coordinates of a midpoint given two points in the Cartesian plane
M03_12	Draws rectangle on square grid given area and perimeter (1 of 2 points)
M05_10	Identifies the value of an angle involving properties of corresponding and supplementary angles
M07_10	Solves a problem involving similar triangles
M08_09A	Determines the area of a parallelogram given its base and height
M08_09B	Uses the Pythagorean theorem to solve for a side length of a parallelogram and calculates the perimeter (1 of 2 points)
M08_10	Completes a parallelogram in the Cartesian plane given three of its vertices
M09_09	Recognizes congruent quadrilaterals

M13_08A	Solves a word problem involving the length around a hexagonal prism
M13_09	Determines the number of exposed faces for unit-cubes that make up a larger cube (1 of 2 points)
M01_11	Solves a problem involving angles of a triangle
M10_13	Identifies the net of a triangular prism
M11_09	Identifies the reflection of a partly shaded shape
M12_10	Solves a two-step word problem involving volume of a rectangular prism and cost
Data and Probability	
M02_13	Computes the mean of five positive and negative values
M02_14	Identifies an appropriate graph for three different types of data
M06_13A	Computes the mean of four given values
M08_11A	Determines the mean value of data represented by four bars in a bar graph
M10_14	Identifies relevant considerations for systematic data collection
M12_14A	Computes the mean of five six-digit numbers
M14_14	Estimates the value of a bar in a bar graph without a scale given the value of another bar
M01_13B	Uses and interprets data sets in pie charts to solve a problem involving percentages
M08_12	Estimates the probability of an event given an observed sample
M11_14	Draws a spinner that has given probabilities
M12_13	Estimates the probability of an event given an observed sample
M06_12	Estimates the number of objects in a given probability sample

Items at Advanced International Benchmark (625)	
Number	
M02_01	Recognizes true or false statements based on properties of operations
M03_02	Solves a two-step word problem involving whole numbers
M04_01	Identifies numbers that are perfect squares
M04_02	Analyzes truth of statements about the properties of a whole number
M05_05B	Solves a non-routine word problem involving whole numbers
M06_04	Uses four different digits to write two two-digit numbers with the smallest product
M08_01	Justifies that a given number satisfies a condition for its parity and factors
M12_01	Determines two integers that satisfy two conditions involving their sum and product
M07_02	Identifies a prime number
M13_01	Identifies an expression equivalent to a given multiplicative expression
M02_03	Solves a multi-step problem involving addition and subtraction of fractions
M08_03	Determines the missing value in a multiplication sentence involving fractions
M10_02	Determines the location of the product of two fractions on a number line
M12_02	Orders fractions and decimals
M12_04	Determines the denominator that makes the sum of a fraction and a whole number equivalent to a decimal
M14_03	Uses four different digits to write two fractions with the largest product

M02_02	Identifies the location of a fraction on a number line
M01_02B	Solves a word problem involving percentages and elapsed time
M05_03	Solves a two-step word problem involving percentages
M06_03	Determines the dimensions of a rectangle that is similar to a given rectangle
M08_04	Recognizes fractions and decimals equivalent to a given percentage
M09_02	Given the volume of a fraction of a container, determines the total volume for multiple containers of the same size
M09_03	Solves a word problem involving price per unit and explains reasoning
M11_03	Completes a table of equivalent proportions and percentages (2 of 2 points)
M11_03	Partially completes a table of equivalent proportions and percentages (1 of 2 points)
M11_04	Solves a word problem involving ratios
M12_05	Solves a word problem involving ratios
M13_03	Identifies a percentage using a given ratio
M07_05	Identifies a true statement about percentages of given numbers
Algebra	
M02_08	Constructs a linear equation for the perimeter of a triangle and solves for the length of one side
M05_06	Identifies an equivalent algebraic expression
M05_08	Uses a given formula involving fractions to solve a word problem
M06_08	Constructs a linear equation for the perimeter of a rectangle and finds the area (2 of 2 points)
M06_08	Constructs a linear equation for the perimeter of a rectangle and finds the area (1 of 2 points)

M08_06	Identifies an inequality that represents the relationship between the areas of two rectangles
M09_05	Simplifies an algebraic expression
M09_08	Constructs a linear equation for the perimeter of a triangle and solves for the length of one side
M10_07	Adds two expressions with two variables and simplifies the result
M12_07	Identifies a simplified expression equivalent to a given expression with parentheses
M14_08	Identifies a simplified expression equivalent to a given expression with fractions and two variables
M14_09	Identifies a pair of linear equations in two variables that represent a situation
M01_05	Identifies an algebraic expression that represents the area of a given rectangle
M02_06	Identifies an expression with parentheses equivalent to a given expression without parentheses
M04_07B	Identifies an inequality that represents a situation
M06_05	Identifies an expression that represents a situation
M10_08	Solves a word problem involving evaluating a formula with an exponent
M01_08	Identifies the graph of a linear equation
M04_08B	Constructs an expression for the nth term of a geometric pattern
M05_09	Demonstrates an understanding of slope by relating graphs and their equations
M06_06	Constructs a linear equation to represent a situation
M08_08B	Extends a given geometric pattern to supply the value of the 50th term
M09_06	Retrieves coordinate points from a graph of a function
M10_10	Constructs an equation to describe the relationship between two quantities

M12_09A	Extends a given geometric pattern to supply the value of the 5th term
M12_09B	Constructs an expression for the nth term of a geometric pattern
M13_06	Identifies the slope of a line given its equation
M02_09	Identifies a point that is collinear with three given collinear points
M03_06	Identifies a line with positive slope
Geometry	
M01_10	Determines the surface area of a prism given the dimensions of its net
M03_09	Uses properties of triangles and quadrilaterals to solve for an angle
M03_12	Draws rectangle on square grid given area and perimeter (2 of 2 points)
M04_09	Draws the image of a triangle translated horizontally and vertically on in the Cartesian plane
M04_10	Solves a two-step problem involving the area of a triangle inscribed in a square (2 of 2 points)
M04_10	Solves a two-step problem involving the area of a triangle inscribed in a square (1 of 2 points)
M05_11	Draws an angle of a given measure on a square grid
M06_10	Finds vertices of triangles created from trapezoids in the Cartesian plane (1 of 2 points)
M06_11	Uses properties of supplementary angles to solve for an angle
M07_09	Draws all lines of symmetry on a regular polygon
M10_11	Determines the area of a square given the side length of a regular hexagon with the same perimeter
M10_12	Solves a word problem involving circles and similar triangles
M11_10	Determines the number of faces of a solid with unit cubes removed

M12_11	Solves for a missing side length given two similar triangles in context
M13_08B	Solves a word problem involving the lateral surface area of a hexagonal prism
M13_10	Solves a word problem involving the Pythagorean theorem
M14_11	Uses properties of corresponding and supplementary angles to solve for an angle in a geometric figure
M14_12	Justifies that a right triangle and obtuse triangle with the same base and height have the same area
M14_13	Determines the surface area of a rectangular prism given its length, width, and height
M09_10	Finds the coordinates of a vertex of a rectangle given the other three vertices
Data and Probability	
M01_13A	Uses and interprets data sets in pie charts to solve a problem involving percentages
M04_11	Identifies relevant considerations for systematic data collection
M06_13B	Determines the change in a mean given changes in individual scores
M07_14	Justifies a conclusion resulting from comparing two distributions
M09_12A	Calculates mean and median for two ordered list of data (2 of 2 points)
M14_16	Explains the change in a mean given changes in individual values
M07_13	Interprets data in a pictograph to solve a multi-step problem
M11_13	Interprets a histogram to identify a proportion
M14_15	Identifies the optimal data display to answer a given question
M03_14B	Compares observed and expected values
M04_12A	Computes the probability of an event given the number of each type of object in a set

M04_12B	Computes the probability of an event given the number of each type of object in a set
M09_13	Solves a multi-step problem involving probability
M13_13	Identifies the conditional probability of an event

Items Above Advanced International Benchmark (625)	
Number	
M02_01	Recognizes equivalent expressions based on properties of operations (2 of 2 points)
M09_04	Given four different containers, identifies the container with the greatest fraction filled
M04_05	Given a ratio in a table, completes two equivalent ratios with one part missing in each
Algebra	
M03_05	Identifies the equivalent form of a linear inequality in one variable
M07_07	Identifies an algebraic expression involving parentheses and negative terms
M07_08	Solves a pair of simultaneous linear equations
M11_05	Identifies equivalent rational expressions
M01_07B	Gives a rule for the nth term of a geometric pattern
M08_08C	Constructs an expression for the nth term of a geometric pattern
M09_07	Determines a collinear point given another point on the line and the slope

Geometry	
M01_09	Identifies the image of a shape after rotation and reflection
M03_11	Solves for a missing side length given two similar triangles
M06_09	Estimates area of an irregular shape on a square grid
M06_10	Finds the coordinates of the vertices of triangles created from trapezoids in the Cartesian plane
M07_11	Solves a multi-step word problem involving ratios between volumes
M08_09B	Uses the Pythagorean theorem to solve for a side length of a parallelogram and calculates the perimeter (2 of 2 points)
M09_11	Explains why two shaded areas of overlapping congruent triangles are equal
M12_12	Solves a word problem involving a quarter of the circumference of a circle (2 of 2 points)
M12_12	Solves a word problem involving a quarter of the circumference of a circle (1 of 2 points)
M13_09	Determines the number of exposed faces for unit-cubes that make up a larger cube (2 of 2 points)
Data and Probability	
M02_15	Explains why a statement about data in a bar graph with a y-axis scale that does not start at 0 is incorrect
M03_13	Compares characteristics of two dot plots to justify a conclusion
M05_13	Explains why a data representation could be misleading
M08_11B	Converts the value of a bar in a bar graph to a percent
M10_15	Compares data in two pie charts with different totals to refute a conclusion
M11_12	Solves a word problem involving averages

Appendix 15D: Grade 8 Science Item Descriptions Developed During the TIMSS 2019 Benchmarking

Items at Low International Benchmark (400)	
Biology	
S11_04	Describes one characteristic of mammals that is advantageous for survival in cold weather (1 of 2 points)
S01_01	States one reason why male penguins' incubation behavior helps their eggs survive (1 of 2 points)
S13_01B	Uses a food web to identify which organisms eat only plants
Chemistry	
S10_10	Identifies the form of wood that will burn fastest based on its size (1 of 2 points)
Physics	
S11_15	Recognizes whether an electromagnet would attract objects made of various materials
Earth Science	
S10_15	States what must be removed from clean ocean water in order for a person to be able drink it

Items at Intermediate International Benchmark (475)	
Biology	
S12_03	Matches 4 of 5 organism groups to defining biological characteristics (1 of 2 points)
S09_04	Justifies an advantage of hollow bones for birds
S14_01	Evaluates a diagram to identify an advantage of a fish's field of vision
S06_04A	Identifies one way that plant and animal cells are similar (1 of 2 points)

S13_04	Recognizes the functions of 2 of 4 tissues found in the human stomach (1 of 2 points)
S06_02	States one substance plants obtain from their environment and use in photosynthesis (1 of 2 points)
S09_03	Recognizes characteristics inherited by rabbits in a given context
S05_03B	Reasons how a crocodile's angle of vision helps it to survive in the environment
S14_05	Identifies the rock layer containing the oldest fossils and justifies the choice
S10_05	Places four organisms in a model of an energy pyramid
S13_01A	Uses a food web to identify which organisms are producers
S08_01	Identifies the best description of the advantages to bird and crocodile in the symbiotic relationship formed when a bird picks food from around a crocodile's teeth
S05_02	Analyzes information about an ecosystem and explains the effect of introducing a new population
S04_04	Explains how reducing the number of vehicles in a city center affects air quality
S07_01A	Recognizes the agent that causes influenza
S14_02	Identifies diseases associated with 4 of 4 human behaviors
Chemistry	
S07_05Z	From a list of symbols and formulas, recognizes which are elements and which are compounds
S01_07	Applies knowledge of concentration to explain why one solution is paler than another solution
S12_09	Explains that volume is one factor that can be used to identify the solution with a higher concentration of solute (1 of 2 points)
S07_07	Recognizes an everyday occurrence that is an example of a chemical change
S03_01	Recognizes a chemical process that involves the absorption of light

Physics	
S14_12	Identifies the glass of ice cubes that will melt faster based on the ice cubes' size (1 of 2 points)
S03_09	Recognizes the type of energy change that occurs as a child slides down a slide
S10_13	Recognizes 5 of 5 materials as conductors or insulators based on a graph showing the electric current in circuits containing the materials
S12_13B	Identifies a statement describing the movement of a motorbike in a chronophotograph
S02_12	Recognizes why a vehicle has a different weight on Mars than it does on Earth
Earth Science	
S01_12	Recognizes the reason for cold temperatures outside an airplane in flight
S13_15	Synthesizes information in rainfall and temperature graphs to match 2 of 4 animals with the climate where they live (1 of 2 points)
S04_13	Identifies 5 or 6 of 7 activities as examples of reducing, reusing, or recycling (1 of 2 points)
S14_15	Interprets a diagram to identify the position of the Moon in orbit during a specific phase of the Moon

Items at High International Benchmark (550)	
Biology	
S04_01	Applies knowledge of mammals to identify how echidna differ from most mammals
S04_02	Identifies examples of animals belonging to 4 of 4 groups of organisms
S10_06	States one biological difference between fish and mammals (1 of 2 points)
S12_03	Matches 5 of 5 organism groups to defining biological characteristics (2 of 2 points)
S08_02	Identifies the body systems to which 4 of 4 organs belong

S09_02	Recognizes 2 of 3 major organs in a diagram (1 of 2 points)
S14_04	Explains why the percentage of oxygen differs in inhaled and exhaled air or why the percentage of nitrogen is the same in inhaled and exhaled air (1 of 2 points)
S11_04	Describes two characteristics of mammals that are advantageous for survival in cold weather (2 of 2 points)
S13_03	Predicts how heart rate changes in response to exercise, based on a set of given conditions
S04_03	Recognizes where new cells come from as an organism grows
S10_01	Identifies the functions of 5 of 5 human cell types (2 of 2 points)
S10_01	Identifies the functions of 4 of 5 human cell types (1 of 2 points)
S13_02	Explains how a fossil can be classified as plant or animal, based on its cellular structure
S01_02	Recognizes an organism that is made up of cells with cell walls
S08_03	Identifies an implication of removing a plant cell's chloroplasts
S09_01	Recognizes what happens to an animal's cells as it grows
S08_05	Identifies where DNA is located in a human body cell
S10_02	Identifies acquired characteristics of a pet bird
S06_03	Recognizes why rabbits inherit traits that their parents do not have
S14_03	Interprets a diagram to identify the source of DNA responsible for a plant's flower petal color
S12_05	Identifies the statement about python and boa evolution that is best supported by given information
S11_03	Identifies the conclusion best supported by a diagram of rock layers with embedded fossils
S01_03	Recognizes how decomposers get their energy
S02_03	Explains how roof gardens in cities help reduce the amount of carbon dioxide in the air

S12_01	States the part of tomato plant that releases the most water
S02_04	Recognizes an explanation for why the mass of leaves removed from a tree decreases over time
S01_04	Given a food chain, explains which organism competes most with humans in a farming community
S01_05Z	For pairs of animals, distinguishes between predatory and competitive relationships
S08_06	Uses information in a table to explain why the abundance of one specie in an ecosystem changed between two given years (1 of 2 points)
S04_05	Interprets a food web to identify a predator/prey relationship
S02_01	Recognizes the relationship that occurs when insects that feed on nectar pollinate flowering plants
S03_06A	Evaluates data from a table to draw a conclusion about the reason for a change in population of a species
S07_04	Explains how flooding leads to a shortage of drinking water or the spread of disease (1 of 2 points)
S11_02	Explains why it is unlikely for someone to get sick with the measles a second time
S03_05	Selects and classifies 3 of 4 foods from a list that comprise a balanced diet (1 of 2 points)
S05_01	Recognizes which food is the best source of carbohydrates
S11_01	Recognizes a list of food that comprises a healthy, balanced meal
Chemistry	
S02_06	Identifies the subatomic particle that is locates outside of an atom's nucleus
S11_06	Identifies the number of atoms of each element in nitric acid
S03_02	Recognizes a model of a carbon dioxide molecule
S08_10	Identifies an explanation of how carbon dioxide can extinguish a fire
S12_07	Recognizes a chemical property

S10_08	Identifies a necessary property for a liquid in a thermometer
S04_08B	States one variable to hold constant when investigating reactivity of different types of steel with water (1 of 2 points)
S07_10	Explains the effect of temperature on diffusion in the context of an investigation
S11_07	Uses data in a table to order set-ups according to the rate at which a solute will dissolve in water
S05_07	Recognizes a property that is common to both acids and bases
S05_08	Recognizes which process makes bronze dark and dull over time
S13_11	Explains whether a reaction between two solutions in a given context can occur a second time
S08_08	Interprets a diagram to identify the number of hydrogen atoms present before a chemical reaction
Physics	
S08_12	States that the amount of a substance present in its liquid form and present in its solid form is the same (1 of 2 points)
S02_11	Recognizes steps that should be taken to ensure an experiment will show whether iron or copper is the better conductor of heat
S09_05	Relates knowledge of heat transfer to recognize a graph that shows how two substances eventually reach temperature equilibrium
S13_07	Recognizes whether a red object will absorb or reflect different colors of light
S02_13	Applies knowledge of sound transmission to explain whether a ringing cell phone in a vacuum can be heard outside the vacuum chamber
S06_07	Recognizes which graph represents a musical note with given specifications for volume and pitch
S01_09B	Explains that in a parallel arrangement of two bulbs, one bulb failing does not affect the other bulb
S08_14	Recognizes for 5 statements about magnets whether they are true or false
S01_10	Recognizes the best explanation of why two bar magnets repel each other
S04_12	States the force represented by an arrow in a diagram of a falling object

S05_06	Recognizes and explains which substance will float on water using a table of densities
S05_10	Given the densities of two objects and three liquids, and diagrams showing the objects floating or sinking in the liquids, identifies each liquid
S10_11	Explains how deploying a parachute slows a skydiver's fall
S13_06	Relates knowledge of density to indicate the order in which three liquids will settle after being poured in a beaker
S06_08	Recognizes a free-body diagram that has a total force acting towards the right
Earth Science	
S13_14	Recognizes sources of fresh and salt water in a diagram
S02_16	Interprets a diagram to identify the natural resource that is formed during the process depicted
S06_01	Recognizes the process in the water cycle indicated in a diagram of an ecosystem
S02_15	Identifies evidence that the Earth is becoming warmer over time
S06_13A	Relates information in temperature graphs and maps to recognize climatic attributes of two cities
S11_11A	Interprets information in a climate graph to determine the warmest and driest month of the year
S13_13	Identifies how the melting of permafrost can affect the Earth's climate
S13_15	Synthesizes information in rainfall and temperature graphs to match 4 of 4 animals with the climates where they live (2 of 2 points)
S05_13	Uses a graph of average monthly temperature to identify the city most likely to be located at the equator
S09_13	Identifies a disadvantage of using solar energy
S04_14	Recognizes the best explanation for why a river floods more often after a forest is cleared
S01_11B	Synthesizes information from tables about revolution times around and distances from the Sun to infer relative distances of planets from the Sun
S12_14	Identifies the best explanation for why Saturn is visible from Earth

S01_11A	Uses information in a table with characteristics of planets to identify the planet with the shortest day length
S04_15	Recognizes a description of how the Sun produces its own light

Items at Advanced International Benchmark (625)	
Biology	
S02_05A	Classifies 7 of 7 animals as mammals or nonmammals
S10_06	States two biological differences between fish and mammals (2 of 2 points)
S14_04	Explains why the percentage of oxygen differs in inhaled and exhaled air and why the percentage of nitrogen is the same in inhaled and exhaled air (2 of 2 points)
S07_02	Interprets a diagram to identify what happens to biceps and triceps when an elbow bends
S02_02	Recognizes where DNA replication takes place in an animal cell
S06_04A	Identifies two ways that plant and animal cells are similar (2 of 2 points)
S06_04B	States one way that plant and animal cells are different (1 of 2 points)
S13_04	Recognizes the functions of 4 of 4 tissues found in the human stomach (2 of 2 points)
S10_04	Identifies an explanation for why plants in a tank with woodlice grow faster than plants in a tank without woodlice
S10_03	Identifies the tube containing two substances bacteria need for cellular respiration
S12_06	Identifies how fermentation differs from typical cellular respiration
S05_04	States one similarity between the life cycles of a bird and a frog
S07_03	Recognizes a human characteristic that is acquired
S01_01	States two reasons why male penguins' incubation behavior helps their eggs survive (2 of 2 points)

S05_03A	Justifies a statement about crocodiles' adaptation to their environment, based on given facts
S03_04	Applies knowledge about the theory of evolution to identify the best conclusion supported by a diagram of limbs from different animals
S08_04	Identifies where the largest energy transfer occurs in an energy pyramid
S11_05	Recognizes an example of a symbiotic relationship between two organisms
S03_06B	Selects and evaluates data from a table to draw a conclusion about the likely reason for a change in population of a species
S14_06	States two ways that planting trees is beneficial for the environment
S12_02	Identifies a human activity that can increase the amount of nutrients in a pond
S14_07	Recognizes the function of white blood cells in the human immune system
S03_05	Selects and classifies 4 of 4 foods from a list that comprise a balanced diet (2 of 2 points)
Chemistry	
S04_07	States the subatomic particle that is not included in a diagram of an atom
S06_11	Recognizes what happens to the atoms in an object pounded flat
S09_08	Recognizes whether 4 of 5 substances are elements, compounds, or mixtures (1 of 2 points)
S02_07	Uses a portion of the periodic table to order four elements from the smallest atomic number to the largest atomic number
S14_08	Uses atomic numbers to identify the position of 4 of 4 elements in a portion of the periodic table
S10_09	Identifies a similarity between two elements in the same group of the periodic table
S04_08A	Explains how measuring the amount of rust on discs made from different types of steel will show which type of steel is more reactive with water
S07_06	Identifies an element as a metal or a nonmetal, based on a list of physical properties and predicts one additional property
S14_09	Compares/contrasts substances in a table to identify the property used to sort them into two groups

S03_03	Applies knowledge of density to identify and explain which liquid will leave a dropper first after a mixture separates
S12_08	Identifies pieces of equipment that could be used to separate and collect substances from 4 of 4 mixtures
S12_09	Applies knowledge of concentration to identify the cup of tea with the higher concentration of sugar (2 of 2 points)
S09_11	Explains whether a reaction took place after a pH indicator is added to a solution based on information provided about the indicator
S09_10	Identifies and explains whether a described change is physical or chemical
S06_10	Recognizes which model best illustrates the results of a chemical reaction
S12_10	Identifies the statement that best describes what occurs when iron sulfide is formed
Physics	
S02_10	Recognizes a diagram of what happens to gas molecules inside a balloon when the balloon expands
S06_09	Explains the difference between a solid and air in terms of particle spacing in context
S11_13	Draws a conclusion about the states of substances in two pistons, based on the different amounts of compression that occurred
S05_09	Recognizes why gases are easier to compress than solids and liquids
S12_11	Recognizes what happens to water molecules in an ice cube when the ice cube melts
S14_11	Interprets a temperature graph to identify the process happening in a given section of the graph
S08_12	Applies the law of conservation of mass to compare the mass of a substance before and after a state change (2 of 2 points)
S01_08	Recognizes an everyday process that is an example of a physical change
S06_05	Recognizes how the mass of a metal ball will change as it cools down
S11_14	Recognizes the type of energy transformation that occurs when a car begins to move from rest
S10_12	Recognizes an experimental design that will determine whether an aluminum, iron, or ceramic bar conducts heat the fastest

S09_07	Recognizes an explanation for why a ball appears a certain color in a given context
S04_11	Uses a diagram to determine a position where an observer's shadow would not fall on a monument
S03_07	Recognizes which property of sound allows animals to navigate and find food
S14_13	Identifies a description of the relationship between sounds made by the longest and shortest bars on a xylophone
S01_09A	States one reason why a bulb in a diagram of an electrical circuit does not light
S08_13	Identifies the components that must be included in a circuit that will turn a bell on and off
S13_08	Indicates whether parts of a light bulb are electrical conductors or insulators
S01_09C	Recognizes a correct statement about battery life and bulb brightness in two given electrical circuits
S07_09	Recognizes how to increase the strength of an electromagnet
S08_11	States the two measurements needed to calculate average speed in an everyday context
S12_13A	Identifies the movement of a motorbike in a chronophotograph and explains how the chronophotograph reveals the motorbike's movement
S03_10	Identifies and explains which of three methods will require the smallest force to move a heavy box onto a truck
S12_12	Explains why a person slides down a waterslide faster when the water is turned on
S07_11	Applies knowledge about the relationship between depth and water pressure to recognize a conclusion about the pressure at different depths
Earth Science	
S05_12	States one condition below Earth's crust that can be inferred from volcanic eruptions
S03_11	Recognizes a major source of water for desalinization plants
S09_12	Recognizes the gas that makes up most of Earth's atmosphere
S02_14	Recognizes why a balloon gets bigger as its height above the ground increases

S09_14A	Recognizes the process that forms rock layers
S14_14	Recognizes climatic conditions that cause rock to erode the fastest
S03_13	Uses a diagram of a mountain range on the ocean and a given wind direction to recognize which location will have the greatest rainfall
S07_12	Recognizes the source of energy for the water cycle
S04_13	Identifies 7 of 7 activities as examples of reducing, reusing, or recycling (2 of 2 points)
S06_12	Describes one geographic factor to consider when selecting a safe location for a nuclear power plant
S07_13	Explains one way trees protect soil from erosion
S13_12	Recognizes a negative effect that fertilizer can have on the environment
S08_17	Recognizes the main cause of water level changes in a harbor over the course of 24 hours
S07_14	Justifies a claim that the Moon travels around the Sun

Items Above Advanced International Benchmark (625)

Biology

S02_05B	States the biological characteristic used to distinguish vertebrates from invertebrates
S09_02	Recognizes 3 of 3 major organs in a diagram (2 of 2 points)
S06_04B	States two ways that plant and animal cells are different (2 of 2 points)
S06_02	States two substances plants obtain from their environment and use in photosynthesis (2 of 2 points)
S05_05	Identifies an explanation for disappearance of a trait over generations
S04_06	Identifies where the carbon in wood comes from

S08_06	Uses information in a table to explain why the abundance of two species in an ecosystem changed between two given years (2 of 2 points)
S03_06C	Predicts which species would best survive in a given environment, using information in a table, and provides a supporting explanation
S07_04	Explains how flooding leads to a shortage of drinking water and the spread of disease (2 of 2 points)
S07_01B	Explains how influenza can be spread rapidly around the world
Chemistry	
S13_10	Recognizes a true statement about neutral atoms
S09_08	Recognizes whether each of five substances is an element, a compound, or a mixture (2 of 2 points)
S04_08B	States two variables to hold constant when investigating reactivity of different types of steel with water (2 of 2 points)
S08_07	Evaluates whether a series of steps will separate a mixture of salt, sand, and iron
S02_08	Interprets information in a table to determine if 3 of 3 solutions are acidic, basic, or neutral
S02_09	Recognizes the reason for a temperature increase when an acid and base are combined
S04_09	Identifies and explains the solution that should be combined with an acidic solution to neutralize it
S11_08	Recognizes a property of a basic solution
S14_10	Predicts the color of flowers that are produced when peat moss is added to soil with a given pH
S13_09	Explains how painting a metal prevents rust from forming
S10_10	Identifies the form of wood that will burn fastest based on its surface area (2 of 2 points)
Physics	
S07_08	Recognizes the property of a gas in a dented ping pong ball that stays constant if the ball is heated
S09_09	Explains how a substance can be in two different states in a container at one time in a given context

S14_12	Identifies the glass of ice cubes that will melt faster based on the ice cubes' surface area (2 of 2 points)
S10_14	Recognizes the position in a diagram where a thrown stone has the greatest kinetic energy
S03_08	Recognizes how the temperature of water changes over time when heated
S04_10	Interprets a graph to identify the description of how heat is transferred between a substance and its surroundings
S06_06	Uses a diagram to explain one way to increase the strength of an electromagnet
S09_06	Explains why a vehicle with tires is more likely to sink in the mud than a vehicle with treads
Earth Science	
S14_16	Recognizes the diagram that best represents the structure of the Earth
S11_10	Recognizes the relative composition of gases in Earth's atmosphere
S09_14B	Given a diagram, explains a process that shaped a rock formation in the ocean
S11_09	Recognizes how oil is formed on Earth
S06_13B	Synthesizes information in temperature graphs and maps to recognize an explanation for the difference in seasonal climates of two cities at similar latitudes
S10_16	Identifies best explanation for why temperatures are hotter in a city center than in a meadow
S11_11B	Evaluates a conclusion about climate data, based on one week of weather observations
S08_16	Explains why oil, gas, and coal are nonrenewable resources
S08_15	Evaluates what kind of area would benefit most from a desalination plant
S05_14	Identifies an explanation for why a constellation visible one night is no longer visible six months later

CHAPTER 16

Creating and Interpreting the TIMSS 2019 Context Questionnaire Scales

Liqun Yin
Bethany Fishbein

Overview

As described in Chapter 2, many of the TIMSS 2019 context questionnaire items were developed and updated to be combined into scales measuring a single underlying latent construct. While some of these scales measuring particular factors or constructs have been of interest for several TIMSS assessment cycles, others were developed in TIMSS 2019 to reflect current research or collect useful information of interest in important areas, such as enhancing the measures of teacher instructional quality. For reporting, the scales were constructed using item response theory (IRT) scaling methods, specifically the Rasch partial credit model (Masters, 1982; Masters & Wright, 1997). For certain scales that maintained many of the same items across TIMSS 2015 and TIMSS 2019, the scales were linked to allow for trend measurement on the background construct.

As a parallel to the TIMSS International Benchmarks of achievement, each context scale allowed students to be classified into regions corresponding to high, middle, and low values on the construct. To facilitate interpretation of the regions, the cutpoints on the scale delimiting the regions were described in terms of combinations of response categories.

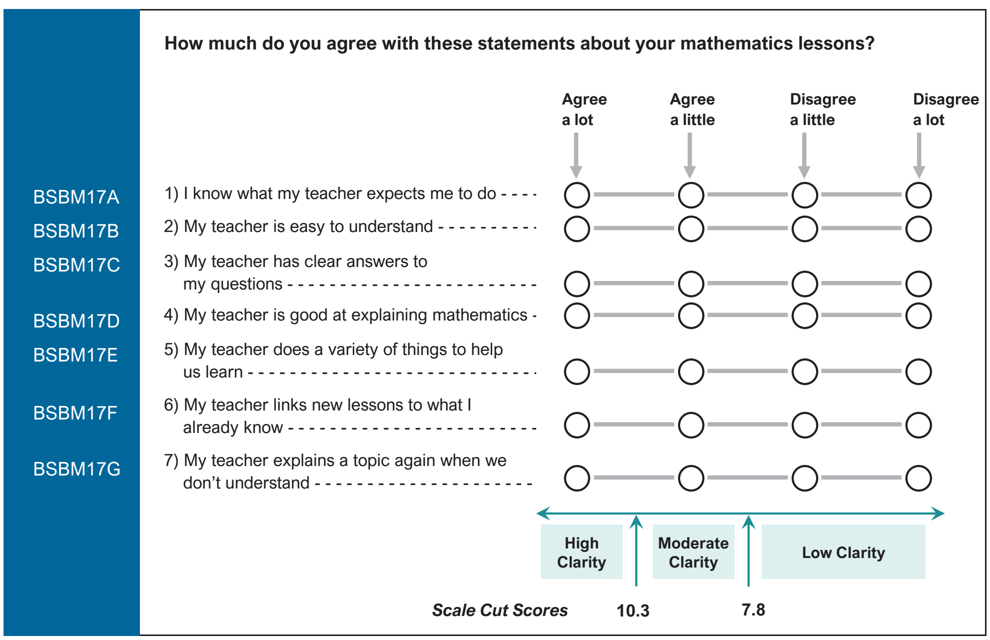
This chapter describes the procedures for constructing, interpreting, and validating scales based on responses to student, home, school, and teacher questionnaires, and then details the process for linking and reporting trend scales.

Reporting TIMSS 2019 Context Questionnaire Scales

As an example illustrating the TIMSS approach to reporting context questionnaire data, Exhibit 16.1 presents the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* scale at the eighth grade, a scale that was reported for the first time for the TIMSS 2019 assessment. As the name suggests, this scale seeks to measure students' perceptions about the clarity of instruction in their mathematics lessons based on their responses to seven statements. For each of the seven statements, students were asked to indicate the degree of their agreement with the statement: agree a lot, agree a little, disagree a little, or disagree a lot. Using the IRT partial credit model, the data from student responses were placed on a scale constructed so that the scale centerpoint of 10 was located at the mean score of the combined distribution of all TIMSS 2019 eighth grade countries. The units of the scale were chosen so that 2 scale score points corresponded to the standard deviation of the distribution.

Students who were classified as having "high clarity of instruction" in their mathematics lessons are those with a scale score greater than or equal to a cutpoint defined for the scale, 10.3 in this case, corresponding to "agreeing a lot" with four of the seven statements and "agreeing a little" with the other three statements, on average. Students who were classified as having "low clarity of instruction" in their mathematics lessons had a score no higher than a cutpoint of 7.8, corresponding to "disagreeing a little" with four of the statements, and "agreeing a little" with the other three, on average. All other students, i.e., those with scores greater than 7.8 but lower than 10.3, were considered to report "moderate clarity of instruction."

Exhibit 16.1: Items in the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* Scale—Grade 8



Scaling Procedure for Context Questionnaire Scales

Partial credit IRT scaling is based on a latent variable model that relates the probability that a person will choose a particular response to an item to that person's location on the underlying construct.

The partial credit model (Masters, 1982) is shown below:

$$P_i(x|\theta_n) = \frac{e^{\sum_{j=0}^{x}\left(\theta_n - \delta_i + \tau_{ij}\right)}}{\sum_{h=0}^{m_i} e^{\sum_{j=0}^{h}\left(\theta_n - \delta_i + \tau_{ij}\right)}} \qquad x = 0, 1, \ldots, m_i \tag{16.1}$$

where $P_i(x|\theta_n)$ denotes the probability that person n with location θ_n on the latent construct would choose response category x on item i out of the $(m_i + 1)$ possible response levels for the item. The item parameter δ_i gives the location of the item on the latent construct and the τ_{ij} denote step parameters for the response levels. Masters (1982) uses a somewhat different parameterization, namely $b_{ij} = \delta_i - \tau_{ij}$ (see also Chapter 11 in this volume). For each scale, the scaling procedure involves first estimating the δ_i and τ_{ij} item parameters at the international level, and then using the model with these parameters to estimate θ_n, the score on the latent construct, for each on the n respondents. Depending on the scale, respondents may be students, parents, teachers, or school principals.

In the TIMSS 2019 eighth grade *Instructional Clarity in Mathematics Lessons* scale, the underlying construct is students' perceptions about the clarity of instruction in their mathematics lessons, and students who agree in general with the seven statements are assumed to perceive more instructional clarity, and students who disagree with the statements are assumed to perceive less clarity.

The TIMSS 2019 context questionnaire scaling was conducted using the ConQuest 2.0 software (Wu, Adams, Wilson, & Haldane, 2007). In preparation for the context questionnaire scaling effort, the TIMSS & PIRLS International Study Center developed a system of production programs for calibration of the items on each scale using ConQuest and produce scale scores for each scale respondent. Each assessment population (TIMSS fourth grade, TIMSS eighth grade) consisted of approximately 300,000 students, as well as their parents, teachers, and school principals. The estimation of the item parameters, a procedure also known as item calibration, was conducted on the combined data from all countries, regardless of whether they participated in eTIMSS or paperTIMSS in 2019, with each country contributing equally to the calibration. This was achieved by assigning 'senate weights', i.e., sampling weights that sum to 500 for each country's student data.

Exhibit 16.2 shows the international item parameters for the *Instructional Clarity in Mathematics Lessons* scale at the eighth grade. For each item, the delta parameter δ_i shows the estimated overall location of the item on the scale, and the tau parameters τ_{ij} show the location of the steps, expressed as deviations from delta. In ConQuest, there are two options on the constraints, either on items or on cases, to remove scale indeterminacy in calibration. In TIMSS 2019 context scale calibrations, the constraints on items were used to identify the model by setting the mean of the item difficulty parameters to zero. As it is shown in Exhibit 16.2, the sum of all δ_i is 0 for the *Instructional Clarity in Mathematics Lessons* scale.

In addition, included in the right column is the Rasch infit item statistic, which is a measure of how accurately the model fits the data, with values ranging from 0 to infinity with a value of 1.0 corresponding to optimal fit. Bond & Fox (2001) provide rules of thumb from a practical perspective and consider a range 0.6 to 1.4 for this measure to support an appropriate level of fit for the measurement of survey data using Likert and other rating scales. As can be seen in this exhibit, the infit values obtained for the scale support appropriate fit of the model for the *Instructional Clarity in Mathematics Lessons* scale.

Exhibit 16.2: Item Parameters for the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* Scale—Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBM17A	0.05459	-1.49723	-0.59435	2.09158	1.43
BSBM17B	0.32080	-1.45404	-0.41713	1.87117	0.92
BSBM17C	0.00334	-1.36560	-0.33731	1.70291	0.81
BSBM17D	-0.11300	-1.01045	-0.39445	1.40490	0.77
BSBM17E	-0.03564	-1.22811	-0.35992	1.58803	0.93
BSBM17F	0.11754	-1.42524	-0.44389	1.86913	1.04
BSBM17G	-0.34763	-0.88348	-0.46134	1.34482	0.99

Once the calibration was completed and international item parameters were estimated, individual scores for each respondent (students, teachers, principals, or parents) were estimated using weighted maximum likelihood estimation (Warm, 1989). All cases with valid responses to at least two items on a scale were included in the calibration and scoring processes.

The scale scores produced by the weighted likelihood estimation are in the logit metric with estimated values ranging from approximately -5 to +5. To convert to a more convenient reporting metric for the newly constructed scales, a linear transformation was applied to the international distribution of logit scores for each scale, so that the resulting distribution across all countries had a mean of 10 and a standard deviation of 2. As an example, Exhibit 16.3 presents the scale transformation constants applied to the international distribution of logit scores for the *Instructional Clarity in Mathematics Lessons* scale at the eighth grade to transform them to the (10, 2) reporting metric. Each scale was transformed using a different set of transformation constants (not reported here as they do not carry specific meaning) in order to put the estimates on the reporting scale to match the targets of an average of 10 and standard deviation of 2.

Exhibit 16.3: Scale Transformation Constants for the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* Scale—Grade 8

Scale Transformation Constants	
A = 8.053331	Transformed Scale Score = 8.053331 + 1.109981 • Logit Scale Score
B = 1.109981	

To provide an approach to reporting the context questionnaire scales analogous to the TIMSS International Benchmarks for the TIMSS achievement scales, a method was developed to divide each scale into high, middle, and low regions and provide a content-referenced interpretation for these regions. For the TIMSS achievement scales, the Low, Intermediate, High, and Advanced International Benchmarks are specific reference points on the scale that can be used to monitor progress in student achievement. Using a scale anchoring procedure, as described in Chapter 15, student performance at each benchmark is described in terms of the mathematics and science (depending on the subject) that students reaching that benchmark know and can do. The percentage of students reaching each of these international benchmarks can serve as a profile of student achievement in a country.

For the high, middle, and low regions of the context questionnaire scales, the interpretation is content-referenced to the extent that the boundaries of the regions were described in terms of identifiable combinations of response categories. The particular response combinations that defined the regions boundaries, or cutpoints, were based on a judgment by TIMSS & PIRLS International Study Center staff of what constituted a high or low region on each individual scale. For example, based on a consideration of the questions making up the *Instructional Clarity in Mathematics Lessons* scale, it was determined that in order to be in the high region of the scale and labeled "high clarity of instruction," a student would have to "agree a lot," on average, to at least four of the seven statements and "agree a little" with the other three. Similarly, it was determined that a student who, on average, at most "disagreed a little" with four of the statements and "agreed a little" with the other three would be labeled to have "low clarity of instruction."

The scale region cutpoints were quantified by assigning a numeric value to each response category, such that each respondent's responses to the scale's questions could be expressed as a "raw score." Assigning 0 to "Disagree a lot," 1 to "Disagree a little," 2 to "Agree a little," and 3 to "Agree a lot," results in raw scores on the *Instructional Clarity in Mathematics Lessons* scale ranging from 0 (disagree a lot with all seven statements) to 21 (agree a lot to all seven). A student who "agreed a lot" with four statements and "agreed a little" with the other three would have a raw score of 18 (4×3 + 3×2). Following this approach, a student with a raw score of 18 or more would be in the "high clarity of instruction" region of the scale. Similarly, "agreeing a little" with three statements and "disagreeing a little" with four statements would result in a raw score of 10 (3×2 + 4×1), so that a student with a raw score less than or equal to 10 would be in the "low clarity of instruction" region. Students with raw scores between 10 and 18 were assigned to the "moderate clarity of instruction" region.

A property of a Rasch scale is that each raw score has a unique scale score associated with it. Exhibit 16.4 presents a raw score-scale score equivalence table for the *Instructional Clarity in Mathematics Lessons* scale at the eighth grade. From this table, it can be seen that, for this particular scale, a raw score of 10 corresponds to a scale score of 7.8 (rounding up) and a raw score of 18 corresponds to a scale score of 10.3 (rounding down) for this particular scale.[1] These scale scores were the cutpoints used to divide the scale into the three regions.

Exhibit 16.4: Equivalence Table of Raw and Transformed Scale Scores for the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* Scale—Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	3.69567	
1	4.93373	
2	5.53063	
3	5.94336	
4	6.27085	
5	6.55198	
6	6.80608	
7	7.04464	
8	7.27558	
9	7.50516	
10	7.73983	7.8
11	7.98224	
12	8.24029	
13	8.51896	
14	8.82274	
15	9.15605	
16	9.52219	
17	9.92675	
18	10.38390	10.3
19	10.92487	
20	11.65115	
21	13.01885	

1 The reason for rounding was to facilitate reporting, and it was decided that the highest cutpoint would be rounded down to ensure that those with an unrounded scale score (e.g., 10.38390 for the Instructional Clarity scale) at the cutpoint were included within the highest region. For a similar reason, the lower cutpoint was rounded up.

Linking Procedures for Trend Context Questionnaire Scales

Using context questionnaire IRT scales to measure contextual constructs started with TIMSS 2011, and during the development phase of the TIMSS 2015 questionnaires, a conscious effort was made to increase the number of items contributing to each scale in order to enhance scale reliability. As described in Chapter 2, many of the scales included in the TIMSS 2019 context questionnaires were brought forward from 2015 because they addressed home and school factors that have been of interest for several TIMSS assessments and were relatively stable across successive cycles. These scales were either brought forward in their entirety or modified for the TIMSS 2019 assessment cycle. The scales with no changes or with minor to moderate modifications were considered as trend scales. Linking procedures were then implemented on these trend scales to place the context questionnaire data from the two or three cycles on a common metric, depending on whether the common reporting metric was established in 2011 or in 2015. This section describes the linking procedures for placing data for the TIMSS 2019 context questionnaire scales onto the trend reporting metric and validating this process.

The trend context scales in TIMSS 2019 have items common to both TIMSS 2015 and TIMSS 2019, called trend items, and some new items unique to TIMSS 2019. Generally, a context questionnaire scale was considered as trend in 2019 if it had a sufficient number of items in common with TIMSS 2015: a minimum of 5 common items and more than half of the TIMSS 2019 items being common items. However, all separate science scales at the eighth grade (biology, chemistry, physics, earth sciences) were treated as new scales in 2019 even if most items were common to the TIMSS 2015 versions of those scales. The number of trend countries for separate science scales was relatively small. Linking the scales using limited trend countries across cycles was deemed less beneficial than providing scale cut points that best reflect the distribution of all countries using separate science scales for each cycle separately.

In TIMSS 2019, 18 fourth grade and 16 eighth grade context questionnaire scales were specified as trend scales. Exhibit 16.5 lists the trend context scales in TIMSS 2019 for both the fourth grade and the eighth grade.

Exhibit 16.5: Trend Context Scales in TIMSS 2019

Trend Scales	Year of Trend Metric Established	
	Grade 4	Grade 8
Students Confident in Mathematics	2011	2011
Students Confident in Science	2011	2011
Students Like Learning Mathematics	2011	2011
Students Like Learning Science	2011	2011
Students' Sense of School Belonging	2015	2015
Students Value Mathematics	~	2011
Students Value Science	~	2011
Home Educational Resources	~	2011
Home Resources for Learning	2011	~
Home Early Literacy and Numeracy Activities Before Primary School	2011	~
Could Do Literacy and Numeracy Tasks When Beginning Primary School	2015	~
Parents' Perceptions of Their Child's School	2015	~
Instruction Affected by Mathematics Resource Shortages - Principals' Reports	2011	2011
Instruction Affected by Science Resource Shortages - Principals' Reports	2011	2011
School Discipline - Principals' reports	2011	2011
School Emphasis on Academic Success - Principals' Reports	2015	2015
Schools Where Students Enter the Primary Grades with Literacy and Numeracy Skills	2015	~
Classroom Teaching Limited by Students Not Ready for Instruction	2015	2015
Safe and Orderly Schools - Teachers' Reports	2011	2011
Teachers' Emphasis on Science Investigation	2015	2015
Teachers' Job Satisfaction	2015	2015

As an example, Exhibit 16.6 shows the TIMSS 2019 *Students Confident in Mathematics* scale for fourth grade students. This scale was originally established in TIMSS 2011, and measures how confident students feel about their ability in mathematics, in terms of their level of agreement with nine statements about mathematics. Statements expressing negative sentiment were reverse coded during the scaling. All of the nine statements were common to the TIMSS 2015 and TIMSS 2019 versions of this scale.

Exhibit 16.6: Items in the TIMSS 2019 *Students Confident in Mathematics* Trend Scale—Grade 4

How much do you agree with these statements about mathematics?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
ASBM05A T	1) I usually do well in mathematics	○	○	○	○
ASBM05B T	2) Mathematics is harder for me than for many of my classmates R	○	○	○	○
ASBM05C T	3) I am just not good at mathematics R	○	○	○	○
ASBM05D T	4) I learn things quickly in mathematics	○	○	○	○
ASBM05E T	5) Mathematics makes me nervous R	○	○	○	○
ASBM05F T	6) I am good at working out difficult mathematics problems	○	○	○	○
ASBM05G T	7) My teacher tells me I am good at mathematics	○	○	○	○
ASBM05H T	8) Mathematics is harder for me than any other subject R	○	○	○	○
ASBM05I T	9) Mathematics makes me confused R	○	○	○	○

Very Confident | Somewhat Confident | Not Confident

Scale Cut Scores 10.7 8.5

R Reverse coded

The IRT calibration and scoring procedures for trend scales were the same as those used for the newly developed context scales. The data for these nine items were calibrated across all TIMSS 2019 countries using the Rasch partial credit model, and, through this calibration, item parameters were estimated on a logit scale that was unique to the 2019 assessment cycle. Following calibration, weighted maximum likelihood estimation was used to derive Rasch logit scale scores based on these estimated item parameters for all countries and benchmarking participants, and as such, student scores were placed on this 2019 logit metric. Although similar, the TIMSS 2019 logit metric is not identical to the TIMSS 2015 logit metric, especially for the scales with items modified or new items added, and thus the TIMSS 2019 scores needed to be adjusted to the 2015 metric to allow for trend linking.

This linking was achieved through a two-step transformation process. The first transformation—with linear constants A_1 and B_1—placed the TIMSS 2019 logit scale scores on the TIMSS 2015 logit metric, and the second transformation—with linear constants A_2 and B_2—transformed the TIMSS 2015 logit metric to the TIMSS scale reporting metric, which uses the (10, 2) metric established in 2011. To increase the efficiency of this transformation process and reduce rounding errors, both transformations were combined into one calculation using the equations below to create a set of final scale transformation constants, A and B:

$$B = B_2 \bullet B_1 \tag{16.2}$$

$$A = A_2 + B_2 \bullet A_1 \tag{16.3}$$

The first set of transformation parameters, A_1 and B_1, were obtained by applying the mean/sigma method (Kolen & Brennan, 2004; Marco, 1977) to the two sets of common item parameters: one from the current calibration of TIMSS 2019 context data and the other from the previous calibration of TIMSS 2015 data. The mean and standard deviation of the estimates of the threshold parameters (Masters, 1982), that is, the difference between item location and item step parameter, ($b_{ij} = \delta_i - \tau_{ij}$), were first found over all common items and all categories for each calibration. The transformation parameters A_1 and B_1 were calculated based on these two sets of means and standard deviations:

$$B_1 = \frac{SD_{c15}}{SD_{c19}} \tag{16.4}$$

$$A_1 = MN_{c15} - \frac{SD_{c15}}{SD_{c19}} \bullet MN_{c19} \tag{16.5}$$

where MN_{c19} and SD_{c19} are the mean and standard deviation of the estimates of ($b_{ij} = \delta_i - \tau_{ij}$) of all common items and categories from the current calibration on TIMSS 2019 data; MN_{c15} and SD_{c15} are the mean and standard deviation of the estimates of ($b_{ij} = \delta_i - \tau_{ij}$) of all common items and categories from the previous calibration on TIMSS 2015 data.

The second set of transformation parameters, A_2 and B_2, were retrieved from the scale transformations produced in 2015. This transformation aimed to place the resulting Rasch scores on the TIMSS reporting metric (10, 2). When the scale started in TIMSS 2015, the second set of transformation parameters were produced in 2015 by simply transforming TIMSS 2015 Rasch logit scores on the (10, 2) scale reporting metric.

The Exhibit 16.7 presents the final trend scale transformation constants applied to the TIMSS 2019 international distribution of logit scale scores for the fourth grade *Students Confident in Mathematics* trend scale to transform them to the (10, 2) reporting metric.

Exhibit 16.7: Scale Transformation Constants for the TIMSS 2019 *Students Confident in Mathematics* Scale—Grade 4

Scale Transformation Constants	
A = 8.556200	Transformed Scale Score = 8.556200 + 1.689256 • Logit Scale Score
B = 1.689256	

To validate the trend measurement on the scales, the TIMSS & PIRLS International Study center conducted extensive analysis to examine item behavior in both cycles. To assess the accuracy of the linking, item parameter estimates for the trend items were compared across the two cycles by examining the differences between the TIMSS 2019 item parameter estimates after being transformed to the TIMSS 2015 logit metric, and the TIMSS 2015 item parameter estimates on the 2015 logit scale. The transformation parameters used for transforming item parameters and this evaluation purpose were the same as those for linking the TIMSS 2019 Rasch scale scores to the TIMSS 2015 logit metric, A_1 and B_1, as described above. Exhibit 16.8 presents the differences between these estimates for the fourth grade *Students Confident in Mathematics* trend scale. As can be seen in the exhibit, the differences were at an acceptable level for both location and step parameters, with most deviations being less than 0.1.

Exhibit 16.8: Differences in Parameter Estimates for Common Items on the TIMSS 2015 Logit Metric, *Students Confident in Mathematics* Scale—Grade 4

TIMSS 2019 Variable	TIMSS 2015 Variable	Difference in delta	Difference in tau_1	Difference in tau_2	Difference in tau_3
ASBM05A	ASBM03A	–0.05448	0.01395	0.02991	–0.04386
ASBM05B R	ASBM03B	0.00360	–0.03255	0.04038	–0.00784
ASBM05C R	ASBM03C	0.01797	–0.04531	0.05472	–0.00941
ASBM05D	ASBM03D	0.04765	–0.02641	0.07533	–0.04891
ASBM05E R	ASBM03E	–0.09879	0.05639	–0.01716	–0.03923
ASBM05F	ASBM03F	0.02092	–0.06191	0.06295	–0.00104
ASBM05G	ASBM03G	0.03794	0.00123	0.01043	–0.01166
ASBM05H R	ASBM03H	–0.01459	–0.02508	0.04345	–0.01836
ASBM05I R	ASBM03I	0.03978	–0.03662	0.01018	0.02644

R Reverse coded

Analyzing Trends on the TIMSS 2019 Context Questionnaire Scales

The resulting trend scale scores are comparable to the scales scores in previous cycles, TIMSS 2015, or both TIMSS 2015 and TIMSS 2011 if the trend metric was established in TIMSS 2011. However, trend analyses with the TIMSS context questionnaire scales should be conducted with caution, as observed differences in the scale scores across TIMSS cycles may not be meaningful for all countries. The psychometric behavior of the contextual scales may vary when the group of participating countries varies between successive cycles (as it is the case for the eighth grade separate science scales), and the scales may behave differently across cultures. In addition, the psychometric behavior of the scales within any given country may not remain stable, as linking the TIMSS 2019 scales to TIMSS 2015 did not account for any differences in national translations or adaptations between the two cycles.

Trend results cannot be analyzed using the percentage of students in each benchmark region. For trend scales, the raw cut points were re-defined in each TIMSS assessment cycle because the number of component variables in the scales would likely change when the scales were modified across cycles. To facilitate interpretation of the region boundaries in terms of combinations of response categories, trend scales followed the same procedure as non-trend scales in setting cutpoints for classification into regions. As such, the procedure was primarily dependent on similarities in response patterns without taking into account variations in difficulty across the items that were unique to 2015 or 2019.

Reliability and Validity of the TIMSS 2019 Context Questionnaire Scales

As evidence that the context questionnaire scales provide comparable measurement across countries, reliability coefficients were computed for each scale for every country and benchmarking participant, and a principal component analysis (Hotelling, 1933) of the scale items was conducted. Exhibit 16.9 presents the results of this analysis for the fourth grade *Students Confident in Mathematics* scale. The Cronbach's Alpha reliability coefficients generally were at an acceptable level (Peterson, 1994; Taber, 2017), with almost all above 0.7 and many above 0.8. The exhibit also shows the percentage of variance among the scale items accounted for by the first principal component in each country. In most cases this was acceptably high, indicating that the items could be adequately represented by a single scale. The component loadings of each questionnaire item from the principal component analysis are positive and substantial, indicating a strong correlation between each item and the scale in every country.

Exhibit 16.9: Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Confident in Mathematics* Scale—Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			ASBM03A	ASBM03B[R]	ASBM03C[R]	ASBM03D	ASBM03E[R]	ASBM03F	ASBM03G	ASBM03H[R]	ASBM03I[R]
Albania	0.82	42	0.65	0.68	0.74	0.61	0.58	0.65	0.65	0.69	0.55
Armenia	0.81	40	0.65	0.60	0.73	0.60	0.57	0.60	0.63	0.69	0.63
Australia	0.87	51	0.76	0.73	0.79	0.72	0.62	0.76	0.39	0.78	0.77
Austria	0.89	54	0.77	0.75	0.77	0.76	0.61	0.70	0.69	0.80	0.76
Azerbaijan	0.81	40	0.57	0.64	0.74	0.59	0.60	0.58	0.63	0.69	0.65
Bahrain	0.81	39	0.39	0.68	0.71	0.55	0.67	0.54	0.50	0.76	0.70
Belgium (Flemish)	0.88	52	0.79	0.77	0.82	0.68	0.55	0.73	0.61	0.79	0.70
Bosnia and Herzegovina	0.86	47	0.72	0.69	0.72	0.67	0.63	0.65	0.68	0.73	0.71
Bulgaria	0.88	51	0.75	0.72	0.75	0.75	0.61	0.71	0.67	0.76	0.72
Canada	0.87	50	0.76	0.75	0.77	0.71	0.64	0.71	0.39	0.77	0.77
Chile	0.82	41	0.63	0.60	0.72	0.63	0.61	0.60	0.47	0.73	0.74
Chinese Taipei	0.85	46	0.73	0.67	0.64	0.74	0.56	0.69	0.64	0.73	0.72
Croatia	0.89	52	0.75	0.73	0.72	0.77	0.66	0.69	0.61	0.79	0.77
Cyprus	0.87	49	0.70	0.74	0.77	0.67	0.65	0.67	0.57	0.74	0.75
Czech Republic	0.86	48	0.73	0.70	0.78	0.73	0.64	0.57	0.57	0.76	0.72
Denmark	0.88	50	0.76	0.73	0.81	0.76	0.55	0.66	0.51	0.75	0.75
England	0.88	52	0.77	0.74	0.81	0.74	0.64	0.76	0.35	0.80	0.76
Finland	0.86	48	0.76	0.71	0.80	0.75	0.64	0.70	0.48	0.76	0.57
France	0.88	52	0.77	0.73	0.79	0.74	0.58	0.70	0.55	0.75	0.82
Georgia	0.81	40	0.63	0.69	0.75	0.58	0.60	0.46	0.49	0.73	0.68
Germany	0.90	54	0.76	0.77	0.80	0.74	0.64	0.70	0.65	0.78	0.75
Hong Kong SAR	0.84	45	0.70	0.65	0.78	0.67	0.59	0.70	0.56	0.73	0.63
Hungary	0.88	51	0.78	0.73	0.78	0.73	0.55	0.70	0.72	0.74	0.69
Iran, Islamic Rep. of	0.79	38	0.62	0.52	0.52	0.63	0.64	0.63	0.58	0.69	0.68
Ireland	0.86	50	0.75	0.75	0.78	0.74	0.60	0.75	0.31	0.80	0.73
Italy	0.85	46	0.71	0.70	0.64	0.72	0.65	0.59	0.55	0.75	0.73
Japan	0.87	49	0.72	0.67	0.83	0.70	0.65	0.76	0.43	0.72	0.75
Kazakhstan	0.86	47	0.68	0.71	0.79	0.73	0.58	0.69	0.58	0.73	0.67
Korea, Rep. of	0.88	53	0.84	0.76	0.76	0.75	0.46	0.80	0.53	0.80	0.75
Kosovo	0.74	33	0.53	0.69	0.71	0.41	0.64	0.34	0.49	0.73	0.51
Kuwait	0.76	34	0.29	0.70	0.72	0.41	0.71	0.35	0.38	0.74	0.69

R Reverse coded

Exhibit 16.9: Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Confident in Mathematics* Scale—Grade 4 (continued)

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			ASBM03A	ASBM03B[R]	ASBM03C[R]	ASBM03D	ASBM03E[R]	ASBM03F	ASBM03G	ASBM03H[R]	ASBM03I[R]
Latvia	0.89	53	0.79	0.73	0.79	0.77	0.61	0.71	0.66	0.77	0.68
Lithuania	0.87	50	0.78	0.70	0.70	0.73	0.62	0.73	0.66	0.76	0.64
Malta	0.87	49	0.71	0.73	0.78	0.69	0.67	0.64	0.51	0.77	0.75
Montenegro	0.84	44	0.66	0.69	0.70	0.61	0.66	0.57	0.60	0.74	0.72
Morocco	0.74	32	0.39	0.67	0.69	0.39	0.67	0.33	0.37	0.72	0.68
Netherlands	0.88	53	0.78	0.74	0.84	0.71	0.49	0.75	0.51	0.79	0.83
New Zealand	0.83	43	0.71	0.64	0.75	0.66	0.55	0.70	0.38	0.74	0.69
North Macedonia	0.82	41	0.59	0.72	0.72	0.55	0.65	0.50	0.58	0.74	0.70
Northern Ireland	0.87	51	0.77	0.73	0.80	0.74	0.63	0.74	0.34	0.78	0.77
Norway (5)	0.87	50	0.75	0.73	0.79	0.74	0.54	0.76	0.43	0.76	0.74
Oman	0.71	31	0.16	0.73	0.65	0.34	0.73	0.19	0.25	0.76	0.72
Pakistan	0.70	30	0.21	0.66	0.71	0.19	0.72	0.21	0.20	0.73	0.74
Philippines	0.54	26	−0.31	0.67	0.60	−0.25	0.57	−0.32	−0.34	0.67	0.61
Poland	0.87	49	0.75	0.71	0.71	0.75	0.58	0.72	0.57	0.77	0.71
Portugal	0.85	46	0.70	0.62	0.77	0.65	0.62	0.68	0.60	0.71	0.73
Qatar	0.79	37	0.37	0.72	0.75	0.43	0.71	0.43	0.39	0.77	0.74
Russian Federation	0.88	52	0.75	0.71	0.79	0.79	0.61	0.71	0.62	0.75	0.73
Saudi Arabia	0.76	36	0.12	0.75	0.74	0.37	0.76	0.24	0.33	0.79	0.76
Serbia	0.90	55	0.77	0.76	0.81	0.72	0.66	0.71	0.68	0.77	0.76
Singapore	0.87	51	0.77	0.75	0.79	0.70	0.63	0.70	0.47	0.78	0.75
Slovak Republic	0.88	50	0.76	0.69	0.80	0.68	0.65	0.65	0.58	0.76	0.75
South Africa (5)	0.68	28	0.20	0.69	0.70	0.21	0.66	0.16	0.12	0.74	0.69
Spain	0.84	44	0.66	0.69	0.73	0.61	0.64	0.61	0.54	0.72	0.75
Sweden	0.87	50	0.76	0.70	0.77	0.76	0.58	0.73	0.43	0.80	0.74
Turkey (5)	0.84	45	0.71	0.69	0.76	0.64	0.50	0.67	0.61	0.75	0.70
United Arab Emirates	0.80	39	0.50	0.70	0.73	0.51	0.66	0.48	0.46	0.75	0.71
United States	0.86	48	0.69	0.75	0.77	0.68	0.63	0.65	0.37	0.80	0.76
Benchmarking Participants											
Ontario, Canada	0.87	50	0.76	0.77	0.79	0.69	0.63	0.71	0.39	0.77	0.78
Quebec, Canada	0.87	50	0.78	0.73	0.71	0.74	0.66	0.74	0.39	0.75	0.78
Moscow City, Russian Fed.	0.90	56	0.76	0.77	0.82	0.81	0.65	0.74	0.60	0.78	0.77

[R] Reverse coded

Exhibit 16.9: Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Confident in Mathematics* Scale—Grade 4 (continued)

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			ASBM03A	ASBM03B[R]	ASBM03C[R]	ASBM03D	ASBM03E[R]	ASBM03F	ASBM03G	ASBM03H[R]	ASBM03I[R]
Madrid, Spain	0.84	44	0.68	0.66	0.71	0.62	0.65	0.61	0.54	0.70	0.78
Abu Dhabi, UAE	0.78	36	0.42	0.71	0.74	0.45	0.66	0.42	0.39	0.75	0.70
Dubai, UAE	0.82	42	0.62	0.68	0.75	0.59	0.62	0.58	0.51	0.73	0.70

[R] Reverse coded

As indicators of effective environments for learning, a positive relationship with achievement is an important aspect of validity for the TIMSS context questionnaire scales. For the fourth grade *Students Confident in Mathematics* scale, Exhibit 16.10 presents the Pearson correlation with mathematics achievement in TIMSS 2019 for each country, together with *r*-squared—the proportion of variance in mathematics achievement attributable to the scale. These figures show a moderate relationship with achievement across participating countries. Also shown is the proportion of variance in achievement attributable to differences between the regions of the *Students Confident in Mathematics* scale. This is very similar to the proportion of variance explained by the scale as a whole, indicating that dividing the scale into regions loses little of its power to account for achievement differences.

Exhibit 16.10: Relationship Between the TIMSS 2019 *Students Confident in Mathematics* Scale—Grade 4 and TIMSS 2019 Mathematics Achievement

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Albania	0.39	0.15	0.16
Armenia	0.37	0.14	0.14
Australia	0.44	0.20	0.19
Austria	0.48	0.23	0.22
Azerbaijan	0.38	0.15	0.15
Bahrain	0.26	0.07	0.06
Belgium (Flemish)	0.46	0.21	0.21
Bosnia and Herzegovina	0.40	0.16	0.16
Bulgaria	0.42	0.18	0.19

Exhibit 16.10: Relationship Between the TIMSS 2019 *Students Confident in Mathematics* Scale—Grade 4 and TIMSS 2019 Mathematics Achievement (continued)

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Canada	0.44	0.20	0.20
Chile	0.41	0.17	0.17
Chinese Taipei	0.44	0.19	0.18
Croatia	0.44	0.20	0.19
Cyprus	0.48	0.24	0.22
Czech Republic	0.41	0.17	0.17
Denmark	0.47	0.22	0.20
England	0.45	0.20	0.19
Finland	0.42	0.18	0.18
France	0.43	0.18	0.18
Georgia	0.34	0.12	0.12
Germany	0.47	0.22	0.22
Hong Kong SAR	0.39	0.15	0.17
Hungary	0.50	0.25	0.25
Iran, Islamic Rep. of	0.30	0.09	0.09
Ireland	0.41	0.17	0.15
Italy	0.29	0.09	0.08
Japan	0.44	0.19	0.19
Kazakhstan	0.26	0.07	0.06
Korea, Rep. of	0.48	0.23	0.22
Kosovo	0.37	0.13	0.15
Kuwait	0.28	0.08	0.09
Latvia	0.51	0.26	0.25
Lithuania	0.48	0.23	0.21
Malta	0.39	0.16	0.16
Montenegro	0.42	0.17	0.17
Morocco	0.33	0.11	0.10
Netherlands	0.54	0.29	0.28
New Zealand	0.40	0.16	0.15
North Macedonia	0.43	0.19	0.18

Exhibit 16.10: Relationship Between the TIMSS 2019 *Students Confident in Mathematics* Scale—Grade 4 and TIMSS 2019 Mathematics Achievement (continued)

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Northern Ireland	0.43	0.19	0.19
Norway (5)	0.41	0.17	0.18
Oman	0.37	0.13	0.14
Pakistan	0.24	0.06	0.05
Philippines	0.33	0.11	0.10
Poland	0.45	0.20	0.20
Portugal	0.48	0.23	0.22
Qatar	0.30	0.09	0.09
Russian Federation	0.38	0.15	0.14
Saudi Arabia	0.35	0.12	0.14
Serbia	0.47	0.22	0.21
Singapore	0.49	0.24	0.25
Slovak Republic	0.43	0.18	0.17
South Africa (5)	0.41	0.17	0.15
Spain	0.46	0.21	0.21
Sweden	0.37	0.13	0.12
Turkey (5)	0.41	0.17	0.17
United Arab Emirates	0.24	0.06	0.06
United States	0.46	0.21	0.23
International Median	**0.42**	**0.17**	**0.17**
Benchmarking Participants			
Ontario, Canada	0.43	0.19	0.18
Quebec, Canada	0.48	0.23	0.24
Moscow City, Russian Fed.	0.48	0.23	0.23
Madrid, Spain	0.48	0.23	0.23
Abu Dhabi, UAE	0.27	0.08	0.08
Dubai, UAE	0.19	0.04	0.04

Item parameter estimates and item and scale statistics similar to those above are available in Appendix 16A for each of the fourth grade TIMSS 2019 context questionnaire scales and in Appendix 16B for each of the eighth grade context questionnaire scales. In both Appendices, scales based on students' reports are listed first, followed by principals' reports, then teachers' reports.

References

Adams, R. J., Wu, M. L., & Wilson, M. (2012). The Rasch rating model and the disordered threshold controversy. *Educational and Psychological Measurement, 72*(4), 547–573. Retrieved from http://epm.sagepub.com/content/72/4/547

Bond, T. J. & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental measurement in the human sciences* (1st Edition). Mahwah, NJ: Lawrence Erlbaum Assoc.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417–441, 498–520.

Kolen. M. J. & Brennan, R. L. (2004). *Test equating, scaling, and linking* (2nd Edition). New York: Springer Science + Business media, Inc.

Marco, G. L. (1977). Item characteristic curves solutions to three intractable testing problems. *Journal of Educational Measurement, 14*, 139–160.

Masters, G. N. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika, 47*(2), 149-174.

Masters, G. N., & Wright, B. D. (1997). The partial credit model. In: W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory*. Berlin: Springer.

Peterson, A. R. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research, 21*(2), 381-391.

Taber, K. (2017). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*, 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*(3), 427–450.

Wu, M. L., Adams, R. J, Wilson, M. R., & Haldane, S. (2007). Conquest 2.0 [computer software]. Camberwell, Australia: Australian Council for Educational Research.

Appendix 16A: TIMSS 2019 Context Questionnaire Scales—Grade 4

Scales Based on Students' and Parents' Reports

Disorderly Behavior During Mathematics Lessons – Grade 4

About the Scale

The *Disorderly Behavior During Mathematics Lessons* scale was created based on students' responses to six items listed below.

How often do these things happen in your mathematics lessons?

		Never	Some lessons	About half the lessons	Every or almost every lesson
ASBM04A	1) Students don't listen to what the teacher says	○	○	○	○
ASBM04B	2) There is disruptive noise	○	○	○	○
ASBM04C	3) It is too disorderly for students to work well	○	○	○	○
ASBM04D	4) My teacher has to wait a long time for students to quiet down	○	○	○	○
ASBM04E	5) Students interrupt the teacher	○	○	○	○
ASBM04F	6) My teacher has to keep telling us to follow the classroom rules	○	○	○	○

Few or No Lessons | **Some Lessons** | **Most Lessons**

Scale Cut Scores **11.6** **8.0**

Item Parameters for the TIMSS 2019 *Disorderly Behavior During Mathematics Lessons* Scale – Grade 4

Item	delta	tau_1	tau_2	tau_3	Infit
ASBM04A	0.06260	-0.67190	-0.64661	1.31851	1.06
ASBM04B	0.15853	-0.46978	-0.54643	1.01621	0.94
ASBM04C	-0.43410	-0.26438	-0.23380	0.49818	1.01
ASBM04D	0.06350	-0.10998	-0.60607	0.71605	0.91
ASBM04E	-0.09436	-0.22110	-0.59219	0.81329	0.90
ASBM04F	0.24383	0.10650	-0.79706	0.69056	1.29

Scale Transformation Constants for the TIMSS 2019 *Disorderly Behavior During Mathematics Lessons* Scale – Grade 4

Scale Transformation Constants	
A = 9.816113	Transformed Scale Score = 9.816113 + 1.569359 • Logit Scale Score
B = 1.569359	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Disorderly Behavior During Mathematics Lessons* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	5.44760	
1	6.85775	
2	7.51217	
3	7.96009	8.0
4	8.31475	
5	8.61844	
6	8.89281	
7	9.15042	
8	9.40001	
9	9.65042	
10	9.90339	
11	10.17203	
12	10.46699	
13	10.79976	
14	11.19445	
15	11.68787	11.6
16	12.34583	
17	13.31363	
18	15.22889	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Disorderly Behavior During Mathematics Lessons* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item					
			ASBM04A	ASBM04B	ASBM04C	ASBM04D	ASBM04E	ASBM04F
Albania	0.79	51	0.72	0.80	0.80	0.70	0.76	0.42
Armenia	0.84	55	0.75	0.78	0.66	0.77	0.76	0.73
Australia	0.86	58	0.71	0.77	0.71	0.81	0.81	0.75
Austria	0.85	57	0.67	0.78	0.76	0.82	0.79	0.70
Azerbaijan	0.80	50	0.70	0.79	0.75	0.71	0.72	0.54
Bahrain	0.79	49	0.69	0.74	0.67	0.75	0.76	0.59
Belgium (Flemish)	0.78	48	0.66	0.66	0.56	0.76	0.77	0.71
Bosnia and Herzegovina	0.84	55	0.67	0.75	0.62	0.81	0.80	0.76
Bulgaria	0.85	57	0.57	0.80	0.81	0.83	0.74	0.72
Canada	0.84	56	0.71	0.74	0.71	0.80	0.79	0.73
Chile	0.84	55	0.64	0.72	0.77	0.79	0.81	0.72
Chinese Taipei	0.84	56	0.72	0.78	0.81	0.76	0.78	0.60
Croatia	0.87	61	0.69	0.81	0.77	0.84	0.82	0.75
Cyprus	0.79	50	0.67	0.74	0.53	0.80	0.79	0.66
Czech Republic	0.89	64	0.74	0.82	0.78	0.83	0.84	0.80
Denmark	0.85	58	0.67	0.82	0.81	0.79	0.77	0.68
England	0.85	58	0.73	0.77	0.71	0.81	0.82	0.72
Finland	0.88	62	0.73	0.80	0.82	0.83	0.80	0.75
France	0.82	52	0.61	0.75	0.76	0.78	0.72	0.72
Georgia	0.84	55	0.67	0.77	0.62	0.81	0.81	0.74
Germany	0.85	58	0.68	0.79	0.79	0.80	0.78	0.70
Hong Kong SAR	0.88	63	0.78	0.81	0.82	0.82	0.82	0.71
Hungary	0.86	58	0.66	0.79	0.71	0.82	0.81	0.80
Iran, Islamic Rep. of	0.84	55	0.65	0.77	0.77	0.80	0.78	0.68
Ireland	0.83	54	0.71	0.74	0.69	0.78	0.77	0.70
Italy	0.81	52	0.61	0.77	0.76	0.74	0.76	0.67
Japan	0.86	58	0.73	0.76	0.78	0.75	0.78	0.77
Kazakhstan	0.77	50	0.69	0.79	0.76	0.73	0.80	0.35
Korea, Rep. of	0.83	55	0.77	0.80	0.75	0.73	0.81	0.54
Kosovo	0.80	52	0.71	0.79	0.76	0.76	0.79	0.45
Kuwait	0.76	45	0.69	0.76	0.67	0.73	0.75	0.37
Latvia	0.86	58	0.73	0.79	0.72	0.81	0.80	0.72
Lithuania	0.85	58	0.76	0.82	0.78	0.82	0.74	0.63
Malta	0.79	49	0.66	0.73	0.65	0.76	0.78	0.62
Montenegro	0.82	52	0.70	0.76	0.57	0.78	0.79	0.71
Morocco	0.75	45	0.64	0.72	0.72	0.71	0.72	0.47
Netherlands	0.84	55	0.60	0.79	0.79	0.81	0.78	0.67
New Zealand	0.84	56	0.67	0.75	0.76	0.79	0.80	0.70
North Macedonia	0.84	56	0.73	0.78	0.75	0.80	0.79	0.60

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Disorderly Behavior During Mathematics Lessons* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item					
			ASBM04A	ASBM04B	ASBM04C	ASBM04D	ASBM04E	ASBM04F
Northern Ireland	0.81	52	0.64	0.68	0.76	0.78	0.76	0.68
Norway (5)	0.87	60	0.70	0.81	0.78	0.83	0.80	0.74
Oman	0.80	49	0.64	0.74	0.74	0.74	0.73	0.61
Pakistan	0.77	47	0.68	0.66	0.69	0.74	0.69	0.65
Philippines	0.64	36	0.58	0.60	0.63	0.60	0.67	0.49
Poland	0.89	63	0.67	0.82	0.83	0.84	0.80	0.81
Portugal	0.86	58	0.64	0.78	0.77	0.79	0.81	0.79
Qatar	0.83	54	0.71	0.77	0.72	0.77	0.79	0.62
Russian Federation	0.88	63	0.73	0.82	0.82	0.82	0.79	0.77
Saudi Arabia	0.80	51	0.69	0.78	0.77	0.75	0.76	0.50
Serbia	0.83	53	0.68	0.75	0.66	0.76	0.80	0.72
Singapore	-	-	-	-	-	-	-	-
Slovak Republic	0.87	61	0.67	0.80	0.77	0.82	0.82	0.78
South Africa (5)	0.74	43	0.59	0.67	0.68	0.72	0.72	0.55
Spain	0.77	48	0.67	0.76	0.65	0.79	0.79	0.45
Sweden	0.87	61	0.71	0.80	0.81	0.84	0.81	0.71
Turkey (5)	0.83	54	0.72	0.78	0.66	0.76	0.77	0.70
United Arab Emirates	0.86	59	0.76	0.80	0.76	0.80	0.81	0.67
United States	0.87	60	0.74	0.79	0.71	0.81	0.82	0.76
Benchmarking Participants								
Ontario, Canada	0.84	56	0.71	0.74	0.68	0.80	0.82	0.73
Quebec, Canada	0.84	55	0.71	0.72	0.75	0.79	0.75	0.72
Moscow City, Russian Fed.	0.87	61	0.72	0.81	0.83	0.83	0.78	0.73
Madrid, Spain	0.77	48	0.67	0.74	0.65	0.79	0.79	0.46
Abu Dhabi, UAE	0.84	56	0.72	0.78	0.75	0.76	0.79	0.65
Dubai, UAE	0.87	62	0.78	0.79	0.73	0.82	0.83	0.74

A dash (–) indicates comparable data not available.

Relationship Between the TIMSS 2019 *Disorderly Behavior During Mathematics Lessons* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Albania	0.11	0.01	0.02
Armenia	0.06	0.00	0.01
Australia	0.24	0.06	0.04
Austria	0.21	0.04	0.03
Azerbaijan	0.04	0.00	0.00
Bahrain	0.10	0.01	0.01
Belgium (Flemish)	0.15	0.02	0.01
Bosnia and Herzegovina	0.09	0.01	0.02
Bulgaria	0.21	0.04	0.06
Canada	0.16	0.03	0.02
Chile	0.12	0.01	0.01
Chinese Taipei	0.01	0.00	0.00
Croatia	0.10	0.01	0.01
Cyprus	0.19	0.04	0.02
Czech Republic	0.18	0.03	0.03
Denmark	0.08	0.01	0.01
England	0.22	0.05	0.03
Finland	0.07	0.00	0.00
France	0.22	0.05	0.03
Georgia	0.14	0.02	0.02
Germany	0.11	0.01	0.01
Hong Kong SAR	0.07	0.01	0.00
Hungary	0.15	0.02	0.02
Iran, Islamic Rep. of	-0.06	0.00	0.01
Ireland	0.19	0.04	0.02
Italy	0.09	0.01	0.00
Japan	0.06	0.00	0.01
Kazakhstan	0.15	0.02	0.02
Korea, Rep. of	0.01	0.00	0.00
Kosovo	0.10	0.01	0.04
Kuwait	0.06	0.00	0.00
Latvia	0.16	0.02	0.02
Lithuania	0.12	0.01	0.01
Malta	0.12	0.01	0.01
Montenegro	0.10	0.01	0.02
Morocco	0.02	0.00	0.01
Netherlands	0.08	0.01	0.00
New Zealand	0.25	0.06	0.04

Relationship Between the TIMSS 2019 *Disorderly Behavior During Mathematics Lessons* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
North Macedonia	0.16	0.03	0.04
Northern Ireland	0.22	0.05	0.03
Norway (5)	0.10	0.01	0.00
Oman	0.13	0.02	0.01
Pakistan	0.19	0.04	0.03
Philippines	0.11	0.01	0.01
Poland	0.11	0.01	0.01
Portugal	0.16	0.02	0.02
Qatar	0.11	0.01	0.01
Russian Federation	0.14	0.02	0.02
Saudi Arabia	0.16	0.03	0.02
Serbia	0.06	0.00	0.01
Singapore	-	-	-
Slovak Republic	0.17	0.03	0.02
South Africa (5)	0.03	0.00	0.00
Spain	0.19	0.03	0.02
Sweden	0.12	0.01	0.01
Turkey (5)	0.14	0.02	0.01
United Arab Emirates	0.12	0.01	0.01
United States	0.19	0.03	0.03
International Median	**0.12**	**0.01**	**0.01**
Benchmarking Participants			
Ontario, Canada	0.17	0.03	0.02
Quebec, Canada	0.15	0.02	0.01
Moscow City, Russian Fed.	0.18	0.03	0.04
Madrid, Spain	0.18	0.03	0.02
Abu Dhabi, UAE	0.13	0.02	0.01
Dubai, UAE	0.07	0.01	0.00

A dash (–) indicates comparable data not available.

Early Literacy and Numeracy Activities – Grade 4

About the Scale

The *Early Literacy and Numeracy Activities* scale was created based on parents' responses to eighteen items listed below.

Before your child began primary/elementary school, how often did you or someone else in your home do the following activities with him/her?

		Often	Sometimes	Never or almost never
ASBH01A [T]	1) Read books	◯	◯	◯
ASBH01B [T]	2) Tell stories	◯	◯	◯
ASBH01C [T]	3) Sing songs	◯	◯	◯
ASBH01D [T]	4) Play with alphabet toys (e.g., blocks with letters of the alphabet)	◯	◯	◯
ASBH01E [T]	5) Talk about things you had done	◯	◯	◯
ASBH01F [T]	6) Talk about what you had read	◯	◯	◯
ASBH01G [T]	7) Play word games	◯	◯	◯
ASBH01H [T]	8) Write letters or words	◯	◯	◯
ASBH01I [T]	9) Read aloud signs and labels	◯	◯	◯
ASBH01J [T]	10) Say counting rhymes or sing counting songs	◯	◯	◯
ASBH01K [T]	11) Play with number toys (e.g., blocks with numbers)	◯	◯	◯
ASBH01L [T]	12) Count different things	◯	◯	◯
ASBH01M [T]	13) Play games involving shapes (e.g., shape sorting toys, puzzles)	◯	◯	◯
ASBH01N [T]	14) Play with building blocks or construction toys	◯	◯	◯
ASBH01O [T]	15) Play board or card games	◯	◯	◯
ASBH01P [T]	16) Write numbers	◯	◯	◯
ASBH01Q	17) Draw shapes	◯	◯	◯
ASBH01R	18) Measure or weigh things (e.g., when cooking)	◯	◯	◯

	Often		Sometimes		Never or Almost Never
Scale Cut Scores		10.6		6.5	

[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Early Literacy and Numeracy Activities* Scale – Grade 4

Item	delta	tau_1	tau_2	Infit
ASBH01A	-0.34476	-1.40303	1.40303	1.05
ASBH01B	-0.32936	-1.33519	1.33519	1.01
ASBH01C	-0.07756	-0.92757	0.92757	1.18
ASBH01D	0.18438	-0.96013	0.96013	0.95
ASBH01E	-0.72634	-1.12221	1.12221	1.08
ASBH01F	0.22725	-1.30680	1.30680	1.02
ASBH01G	0.31745	-1.17398	1.17398	0.94
ASBH01H	-0.10842	-1.07733	1.07733	0.97
ASBH01I	0.14535	-0.96132	0.96132	1.00
ASBH01J	0.41173	-0.98655	0.98655	1.01
ASBH01K	0.26656	-0.99356	0.99356	0.90
ASBH01L	-0.68847	-1.15621	1.15621	0.90
ASBH01M	-0.38096	-0.97230	0.97230	0.94
ASBH01N	-0.36931	-0.70955	0.70955	1.02
ASBH01O	0.20194	-1.08281	1.08281	1.08
ASBH01P	-0.25815	-1.11519	1.11519	0.96
ASBH01Q	-0.17558	-1.16158	1.16158	0.97
ASBH01R	1.70425	-1.11057	1.11057	1.15

Scale Transformation Constants for the TIMSS 2019 *Early Literacy and Numeracy Activities* Scale – Grade 4

Scale Transformation Constants	
A = 8.584625	Transformed Scale Score = 8.584625 + 1.556418 • Logit Scale Score
B = 1.556418	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Early Literacy and Numeracy Activities* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	1.07265	
1	2.86711	
2	3.74829	
3	4.36060	
4	4.84206	
5	5.24617	
6	5.60111	
7	5.92077	
8	6.21476	
9	6.48948	6.5
10	6.74944	
11	6.99730	
12	7.23716	
13	7.46992	
14	7.69717	
15	7.92028	
16	8.14041	
17	8.35859	
18	8.57563	
19	8.79290	
20	9.01090	
21	9.23074	
22	9.45351	
23	9.68045	
24	9.91301	
25	10.15289	
26	10.40107	
27	10.66194	10.6
28	10.93840	
29	11.23530	
30	11.55952	
31	11.92129	
32	12.33545	
33	12.83153	
34	13.46485	
35	14.37589	
36	16.21238	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Early Literacy and Numeracy Activities* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item																	
			ASBH01A	ASBH01B	ASBH01C	ASBH01D	ASBH01E	ASBH01F	ASBH01G	ASBH01H	ASBH01I	ASBH01J	ASBH01K	ASBH01L	ASBH01M	ASBH01N	ASBH01O	ASBH01P	ASBH01Q	ASBH01R
Albania	0.88	34	0.55	0.53	0.40	0.64	0.50	0.57	0.60	0.63	0.58	0.59	0.66	0.64	0.65	0.59	0.51	0.62	0.59	0.50
Armenia	0.85	28	0.47	0.51	0.39	0.60	0.39	0.48	0.58	0.56	0.57	0.58	0.64	0.58	0.60	0.44	0.46	0.57	0.60	0.50
Australia	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Austria	0.84	28	0.44	0.50	0.43	0.53	0.45	0.55	0.57	0.52	0.55	0.61	0.62	0.61	0.51	0.48	0.41	0.54	0.60	0.48
Azerbaijan	0.85	28	0.48	0.49	0.36	0.59	0.36	0.53	0.56	0.59	0.59	0.54	0.63	0.53	0.60	0.55	0.42	0.59	0.60	0.48
Bahrain	0.84	28	0.42	0.46	0.35	0.58	0.40	0.43	0.59	0.58	0.53	0.56	0.65	0.60	0.60	0.57	0.55	0.60	0.58	0.36
Belgium (Flemish)	0.85	29	0.48	0.48	0.42	0.58	0.40	0.49	0.63	0.56	0.59	0.62	0.65	0.64	0.50	0.41	0.50	0.56	0.57	0.47
Bosnia and Herzegovina	0.84	28	0.45	0.47	0.45	0.60	0.38	0.50	0.54	0.54	0.53	0.61	0.65	0.58	0.57	0.48	0.48	0.53	0.57	0.46
Bulgaria	0.92	43	0.62	0.59	0.46	0.73	0.63	0.69	0.70	0.74	0.66	0.65	0.74	0.71	0.72	0.68	0.59	0.74	0.64	0.49
Canada	0.89	36	0.55	0.55	0.47	0.65	0.47	0.62	0.68	0.67	0.64	0.60	0.69	0.65	0.63	0.51	0.50	0.67	0.63	0.53
Chile	0.87	31	0.47	0.52	0.44	0.62	0.37	0.49	0.63	0.64	0.52	0.63	0.69	0.59	0.62	0.61	0.47	0.62	0.60	0.43
Chinese Taipei	0.90	39	0.59	0.58	0.51	0.63	0.51	0.58	0.62	0.55	0.65	0.70	0.73	0.73	0.69	0.65	0.52	0.67	0.69	0.50
Croatia	0.87	31	0.51	0.55	0.40	0.62	0.41	0.55	0.61	0.62	0.58	0.55	0.68	0.62	0.61	0.51	0.45	0.63	0.61	0.39
Cyprus	0.88	34	0.53	0.54	0.45	0.67	0.45	0.58	0.63	0.62	0.57	0.61	0.71	0.65	0.64	0.56	0.51	0.63	0.64	0.48
Czech Republic	0.84	27	0.35	0.41	0.38	0.59	0.42	0.54	0.55	0.63	0.53	0.52	0.67	0.58	0.53	0.42	0.45	0.65	0.59	0.48
Denmark	0.86	30	0.47	0.52	0.43	0.59	0.43	0.58	0.66	0.65	0.57	0.59	0.61	0.60	0.51	0.40	0.45	0.62	0.51	0.54
England	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Finland	0.86	30	0.42	0.49	0.40	0.63	0.41	0.58	0.60	0.66	0.58	0.54	0.64	0.62	0.53	0.41	0.45	0.66	0.59	0.53
France	0.84	28	0.49	0.49	0.40	0.59	0.37	0.47	0.59	0.61	0.53	0.55	0.64	0.61	0.56	0.49	0.47	0.57	0.55	0.47
Georgia	0.85	29	0.53	0.43	0.37	0.66	0.42	0.50	0.58	0.58	0.59	0.55	0.68	0.52	0.56	0.57	0.36	0.59	0.59	0.44
Germany	0.84	27	0.36	0.44	0.37	0.59	0.40	0.50	0.62	0.61	0.60	0.58	0.63	0.56	0.49	0.38	0.43	0.61	0.61	0.49
Hong Kong SAR	0.89	35	0.52	0.55	0.52	0.65	0.43	0.49	0.65	0.57	0.57	0.67	0.73	0.71	0.72	0.65	0.48	0.62	0.60	0.42
Hungary	0.85	28	0.43	0.45	0.37	0.63	0.36	0.46	0.60	0.59	0.58	0.58	0.65	0.57	0.52	0.42	0.47	0.61	0.58	0.56
Iran, Islamic Rep. of	0.88	34	0.53	0.52	0.52	0.62	0.47	0.55	0.64	0.60	0.59	0.61	0.66	0.60	0.63	0.61	0.60	0.58	0.59	0.43
Ireland	0.89	35	0.53	0.55	0.48	0.64	0.44	0.55	0.65	0.62	0.63	0.64	0.70	0.66	0.65	0.55	0.49	0.65	0.65	0.52
Italy	0.83	26	0.40	0.44	0.37	0.62	0.35	0.49	0.56	0.61	0.54	0.59	0.65	0.59	0.49	0.43	0.34	0.58	0.54	0.41
Japan	0.88	33	0.47	0.51	0.49	0.59	0.49	0.56	0.63	0.66	0.61	0.61	0.63	0.68	0.58	0.50	0.47	0.68	0.65	0.49
Kazakhstan	0.84	28	0.48	0.35	0.37	0.55	0.43	0.52	0.56	0.56	0.53	0.56	0.64	0.51	0.58	0.54	0.58	0.56	0.60	0.45
Korea, Rep. of	0.90	39	0.49	0.43	0.45	0.64	0.47	0.55	0.72	0.68	0.69	0.72	0.74	0.74	0.68	0.56	0.51	0.72	0.68	0.56
Kosovo	0.82	26	0.42	0.41	0.38	0.59	0.34	0.41	0.52	0.56	0.54	0.55	0.62	0.53	0.59	0.53	0.48	0.52	0.54	0.47
Kuwait	0.85	29	0.43	0.46	0.31	0.60	0.39	0.40	0.58	0.60	0.55	0.58	0.69	0.60	0.64	0.58	0.53	0.63	0.59	0.34
Latvia	0.86	30	0.51	0.52	0.40	0.63	0.43	0.57	0.59	0.62	0.51	0.59	0.65	0.61	0.56	0.48	0.48	0.63	0.59	0.47
Lithuania	0.86	30	0.51	0.48	0.42	0.63	0.46	0.59	0.62	0.61	0.49	0.57	0.65	0.60	0.57	0.52	0.44	0.62	0.56	0.47
Malta	0.88	34	0.54	0.54	0.45	0.62	0.46	0.58	0.63	0.62	0.63	0.59	0.69	0.65	0.66	0.55	0.49	0.58	0.63	0.48
Montenegro	0.86	30	0.50	0.50	0.47	0.61	0.47	0.54	0.57	0.54	0.55	0.59	0.66	0.59	0.61	0.53	0.53	0.54	0.57	0.49
Morocco	0.93	43	0.61	0.62	0.54	0.70	0.59	0.65	0.69	0.74	0.70	0.64	0.73	0.61	0.70	0.67	0.69	0.73	0.71	0.46
Netherlands	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
New Zealand	0.90	38	0.53	0.60	0.53	0.61	0.51	0.64	0.67	0.67	0.67	0.62	0.68	0.67	0.68	0.52	0.55	0.66	0.66	0.60
North Macedonia	0.88	33	0.55	0.54	0.44	0.64	0.45	0.58	0.63	0.62	0.57	0.64	0.66	0.62	0.62	0.55	0.51	0.59	0.61	0.54
Northern Ireland	0.90	37	0.50	0.58	0.47	0.67	0.46	0.59	0.67	0.64	0.64	0.62	0.70	0.67	0.67	0.58	0.54	0.63	0.64	0.54

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Early Literacy and Numeracy Activities* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item																	
			ASBH01A	ASBH01B	ASBH01C	ASBH01D	ASBH01E	ASBH01F	ASBH01G	ASBH01H	ASBH01I	ASBH01J	ASBH01K	ASBH01L	ASBH01M	ASBH01N	ASBH01O	ASBH01P	ASBH01Q	ASBH01R
Norway (5)	0.87	31	0.46	0.47	0.40	0.64	0.43	0.61	0.71	0.64	0.61	0.63	0.69	0.59	0.50	0.33	0.43	0.65	0.54	0.54
Oman	0.83	27	0.47	0.45	0.23	0.57	0.38	0.49	0.58	0.59	0.59	0.48	0.62	0.52	0.57	0.49	0.58	0.57	0.55	0.41
Pakistan	0.86	31	0.53	0.53	0.27	0.56	0.53	0.61	0.53	0.68	0.60	0.62	0.47	0.59	0.57	0.61	0.42	0.65	0.56	0.49
Philippines	0.88	32	0.54	0.53	0.44	0.61	0.46	0.56	0.53	0.60	0.62	0.62	0.66	0.62	0.64	0.58	0.50	0.59	0.63	0.44
Poland	0.86	30	0.42	0.45	0.47	0.62	0.49	0.57	0.58	0.61	0.57	0.59	0.66	0.57	0.47	0.40	0.44	0.65	0.64	0.50
Portugal	0.86	30	0.52	0.53	0.40	0.61	0.41	0.49	0.61	0.57	0.51	0.58	0.69	0.64	0.62	0.55	0.48	0.59	0.58	0.45
Qatar	0.86	31	0.45	0.46	0.41	0.61	0.44	0.47	0.59	0.58	0.60	0.62	0.68	0.63	0.64	0.59	0.52	0.60	0.61	0.39
Russian Federation	0.88	33	0.56	0.49	0.47	0.61	0.54	0.57	0.60	0.63	0.52	0.57	0.66	0.58	0.63	0.58	0.59	0.64	0.63	0.49
Saudi Arabia	0.86	30	0.42	0.47	0.35	0.62	0.46	0.45	0.62	0.61	0.57	0.57	0.68	0.55	0.59	0.56	0.57	0.63	0.61	0.36
Serbia	0.86	30	0.53	0.54	0.51	0.61	0.43	0.53	0.57	0.58	0.58	0.59	0.64	0.60	0.59	0.49	0.49	0.55	0.58	0.47
Singapore	0.91	40	0.56	0.58	0.52	0.70	0.50	0.59	0.63	0.65	0.68	0.71	0.75	0.74	0.73	0.63	0.50	0.70	0.69	0.50
Slovak Republic	0.88	34	0.48	0.47	0.32	0.64	0.52	0.59	0.63	0.63	0.62	0.59	0.70	0.65	0.60	0.54	0.49	0.67	0.62	0.54
South Africa (5)	0.88	32	0.57	0.50	0.41	0.61	0.47	0.54	0.56	0.60	0.60	0.57	0.65	0.61	0.64	0.59	0.53	0.58	0.61	0.46
Spain	0.85	29	0.48	0.49	0.41	0.64	0.34	0.46	0.63	0.60	0.53	0.57	0.68	0.55	0.60	0.52	0.43	0.62	0.56	0.41
Sweden	0.87	32	0.41	0.46	0.44	0.67	0.37	0.58	0.69	0.64	0.57	0.63	0.67	0.61	0.57	0.44	0.44	0.65	0.60	0.53
Turkey (5)	0.94	50	0.69	0.68	0.54	0.71	0.72	0.74	0.75	0.72	0.75	0.69	0.75	0.73	0.77	0.68	0.61	0.74	0.75	0.65
United Arab Emirates	0.88	34	0.48	0.48	0.39	0.61	0.49	0.54	0.60	0.60	0.60	0.62	0.69	0.67	0.69	0.62	0.56	0.64	0.64	0.47
United States	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Benchmarking Participants																				
Ontario, Canada	0.90	38	0.54	0.57	0.50	0.65	0.50	0.65	0.69	0.69	0.68	0.61	0.70	0.68	0.67	0.54	0.53	0.68	0.65	0.52
Quebec, Canada	0.87	32	0.53	0.51	0.40	0.63	0.45	0.57	0.66	0.65	0.58	0.57	0.68	0.61	0.59	0.43	0.47	0.64	0.58	0.52
Moscow City, Russian Fed.	0.88	34	0.52	0.49	0.49	0.60	0.59	0.62	0.64	0.65	0.52	0.58	0.67	0.60	0.58	0.53	0.56	0.64	0.64	0.54
Madrid, Spain	0.85	29	0.43	0.46	0.38	0.65	0.39	0.45	0.61	0.61	0.52	0.60	0.69	0.55	0.59	0.52	0.42	0.61	0.59	0.44
Abu Dhabi, UAE	0.89	35	0.48	0.50	0.42	0.61	0.48	0.53	0.62	0.62	0.59	0.62	0.70	0.69	0.70	0.63	0.59	0.67	0.66	0.45
Dubai, UAE	0.88	33	0.50	0.48	0.40	0.61	0.47	0.52	0.58	0.59	0.61	0.61	0.70	0.67	0.68	0.59	0.50	0.63	0.62	0.46

A dash (–) indicates comparable data not available.

Relationship Between the TIMSS 2019 *Early Literacy and Numeracy Activities* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Albania	0.23	0.05	0.05	0.22	0.05	0.05
Armenia	0.09	0.01	0.01	0.11	0.01	0.01
Australia	-	-	-	-	-	-
Austria	0.15	0.02	0.01	0.16	0.03	0.02
Azerbaijan	0.12	0.02	0.02	0.14	0.02	0.03
Bahrain	0.11	0.01	0.01	0.19	0.04	0.03
Belgium (Flemish)	0.15	0.02	0.01	0.14	0.02	0.01
Bosnia and Herzegovina	0.13	0.02	0.01	0.14	0.02	0.02
Bulgaria	0.40	0.16	0.14	0.44	0.19	0.18
Canada	0.09	0.01	0.01	0.15	0.02	0.02
Chile	0.22	0.05	0.03	0.23	0.05	0.03
Chinese Taipei	0.22	0.05	0.03	0.22	0.05	0.03
Croatia	0.20	0.04	0.03	0.22	0.05	0.03
Cyprus	0.19	0.04	0.03	0.22	0.05	0.04
Czech Republic	0.09	0.01	0.01	0.07	0.00	0.00
Denmark	0.12	0.01	0.01	0.17	0.03	0.03
England	-	-	-	-	-	-
Finland	0.14	0.02	0.01	0.12	0.01	0.01
France	0.21	0.05	0.03	0.22	0.05	0.04
Georgia	-0.02	0.00	0.00	-0.04	0.00	0.00
Germany	0.12	0.01	0.01	0.11	0.01	0.01
Hong Kong SAR	0.18	0.03	0.03	0.19	0.03	0.03
Hungary	0.12	0.02	0.02	0.11	0.01	0.02
Iran, Islamic Rep. of	0.16	0.03	0.02	0.22	0.05	0.04
Ireland	0.21	0.04	0.03	0.22	0.05	0.03
Italy	0.15	0.02	0.01	0.17	0.03	0.01
Japan	0.17	0.03	0.03	0.16	0.03	0.02
Kazakhstan	0.12	0.02	0.01	0.17	0.03	0.03
Korea, Rep. of	0.21	0.04	0.03	0.19	0.04	0.03
Kosovo	0.12	0.01	0.01	0.16	0.03	0.03
Kuwait	0.12	0.01	0.02	0.14	0.02	0.02
Latvia	0.11	0.01	0.01	0.13	0.02	0.01
Lithuania	0.09	0.01	0.01	0.09	0.01	0.01
Malta	0.19	0.04	0.03	0.16	0.03	0.02
Montenegro	0.18	0.03	0.03	0.20	0.04	0.04
Morocco	0.20	0.04	0.04	0.20	0.04	0.04
Netherlands	-	-	-	-	-	-

Relationship Between the TIMSS 2019 *Early Literacy and Numeracy Activities* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
New Zealand	0.12	0.01	0.01	0.16	0.03	0.02
North Macedonia	0.20	0.04	0.03	0.23	0.05	0.05
Northern Ireland	0.15	0.02	0.01	0.16	0.02	0.02
Norway (5)	0.13	0.02	0.01	0.16	0.03	0.02
Oman	0.22	0.05	0.03	0.25	0.06	0.04
Pakistan	0.11	0.01	0.02	0.08	0.01	0.01
Philippines	0.18	0.03	0.03	0.20	0.04	0.04
Poland	0.10	0.01	0.00	0.11	0.01	0.00
Portugal	0.17	0.03	0.02	0.17	0.03	0.03
Qatar	0.19	0.04	0.03	0.21	0.04	0.03
Russian Federation	0.08	0.01	0.00	0.11	0.01	0.01
Saudi Arabia	0.12	0.02	0.01	0.15	0.02	0.02
Serbia	0.25	0.06	0.05	0.26	0.07	0.06
Singapore	0.15	0.02	0.02	0.18	0.03	0.03
Slovak Republic	0.12	0.01	0.02	0.17	0.03	0.04
South Africa (5)	0.20	0.04	0.03	0.20	0.04	0.04
Spain	0.18	0.03	0.02	0.18	0.03	0.02
Sweden	0.10	0.01	0.01	0.08	0.01	0.00
Turkey (5)	0.41	0.17	0.15	0.46	0.21	0.19
United Arab Emirates	0.14	0.02	0.02	0.16	0.02	0.02
United States	-	-	-	-	-	-
International Median	**0.15**	**0.02**	**0.02**	**0.17**	**0.03**	**0.03**
Benchmarking Participants						
Ontario, Canada	0.09	0.01	0.01	0.12	0.01	0.01
Quebec, Canada	0.13	0.02	0.02	0.17	0.03	0.02
Moscow City, Russian Fed.	0.07	0.00	0.01	0.08	0.01	0.01
Madrid, Spain	0.14	0.02	0.01	0.14	0.02	0.01
Abu Dhabi, UAE	0.16	0.03	0.02	0.16	0.03	0.02
Dubai, UAE	0.12	0.02	0.01	0.15	0.02	0.02

A dash (–) indicates comparable data not available.

Early Literacy and Numeracy Tasks – Grade 4

About the Scale

The *Early Literacy and Numeracy Tasks* scale was created based on parents' responses to twelve items listed below.

How well could your child do the following when he/she began the first grade of primary/elementary school?

		Very well	Moderately well	Not very well	Not at all
ASBH06A T	1) Recognize most of the letters of the alphabet	◯	◯	◯	◯
ASBH06B T	2) Read some words	◯	◯	◯	◯
ASBH06C T	3) Read sentences	◯	◯	◯	◯
ASBH06D T	4) Read a story	◯	◯	◯	◯
ASBH06E T	5) Write letters of the alphabet	◯	◯	◯	◯
ASBH06F	6) Write his/her name	◯	◯	◯	◯
ASBH06G	7) Write words other than his/her name	◯	◯	◯	◯

Could your child do the following when he/she began the first grade of primary/elementary school?

		Up to 100 or higher	Up to 20	Up to 10	Not at all
ASBH07A T	1) Count by himself/herself	◯	◯	◯	◯
ASBH07B T	2) Recognize written numbers	◯	◯	◯	◯
ASBH07C T	3) Write numbers	◯	◯	◯	◯

Could your child do the following when he/she began the first grade of primary/elementary school?

		Yes	No
ASBH07D T	1) Do simple addition	◯	◯
ASBH07E T	2) Do simple subtraction	◯	◯

	Very Well		Moderately Well		Not Well
Scale Cut Scores		11.2		8.6	

T Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Early Literacy and Numeracy Tasks* Scale – Grade 4

Item	delta	tau_1	tau_2	tau_3	Infit
ASBH06A	-0.89706	-1.55492	-0.05316	1.60808	0.91
ASBH06B	0.10783	-1.38397	-0.17270	1.55667	0.82
ASBH06C	0.99903	-1.17817	-0.18774	1.36591	0.87
ASBH06D	1.56339	-1.15006	-0.22907	1.37913	1.04
ASBH06E	-0.47596	-1.49821	-0.13281	1.63102	0.86
ASBH06F	-1.14427	-0.70993	-0.32096	1.03089	1.07
ASBH06G	0.18193	-1.37169	-0.14906	1.52075	0.87
ASBH07A	-0.21367	-1.91023	0.10567	1.80456	1.46
ASBH07B	0.20174	-2.00916	0.29670	1.71246	1.17
ASBH07C	0.41193	-1.91406	0.33064	1.58342	1.11
ASBH07D	-0.76002				1.00
ASBH07E	0.02513				1.03

Scale Transformation Constants for the TIMSS 2019 *Early Literacy and Numeracy Tasks* Scale – Grade 4

Scale Transformation Constants	
A = 8.738315	Transformed Scale Score = 8.738315 + 1.194342 • Logit Scale Score
B = 1.194342	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Early Literacy and Numeracy Tasks* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	3.06734	
1	4.45634	
2	5.14320	
3	5.62117	
4	5.99829	
5	6.31846	
6	6.60154	
7	6.85988	
8	7.10112	
9	7.33022	
10	7.55012	
11	7.76411	
12	7.97308	
13	8.17826	
14	8.38068	
15	8.58119	8.6
16	8.78057	
17	8.97964	
18	9.17925	
19	9.38042	
20	9.58432	
21	9.79237	
22	10.00637	
23	10.22851	
24	10.46028	
25	10.70746	
26	10.97468	
27	11.26987	11.2
28	11.60550	
29	12.00147	
30	12.50191	
31	13.21514	
32	14.63421	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Early Literacy and Numeracy Tasks* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item											
			ASBH06A	ASBH06B	ASBH06C	ASBH06D	ASBH06E	ASBH06F	ASBH06G	ASBH07A	ASBH07B	ASBH07C	ASBH07D	ASBH07E
Albania	0.91	52	0.78	0.81	0.79	0.67	0.79	0.73	0.80	0.62	0.67	0.65	0.61	0.65
Armenia	0.89	47	0.76	0.85	0.83	0.74	0.83	0.81	0.83	0.37	0.49	0.55	0.43	0.44
Australia	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Austria	0.89	44	0.76	0.81	0.78	0.70	0.75	0.45	0.76	0.48	0.67	0.71	0.49	0.51
Azerbaijan	0.91	52	0.76	0.81	0.80	0.72	0.82	0.80	0.82	0.51	0.61	0.65	0.60	0.63
Bahrain	0.88	45	0.71	0.79	0.80	0.71	0.73	0.65	0.76	0.57	0.62	0.62	0.52	0.54
Belgium (Flemish)	0.89	46	0.75	0.79	0.77	0.73	0.78	0.53	0.79	0.49	0.65	0.70	0.50	0.55
Bosnia and Herzegovina	0.90	49	0.73	0.82	0.81	0.76	0.77	0.65	0.81	0.52	0.62	0.66	0.59	0.60
Bulgaria	0.93	59	0.83	0.85	0.81	0.74	0.83	0.81	0.84	0.67	0.69	0.72	0.68	0.68
Canada	0.92	53	0.71	0.81	0.81	0.77	0.76	0.68	0.81	0.67	0.74	0.75	0.58	0.58
Chile	0.91	50	0.77	0.83	0.83	0.78	0.79	0.70	0.80	0.57	0.61	0.62	0.55	0.53
Chinese Taipei	0.86	43	0.78	0.83	0.82	0.78	0.80	0.73	0.79	0.24	0.30	0.31	0.49	0.51
Croatia	0.90	49	0.73	0.81	0.81	0.74	0.76	0.58	0.78	0.59	0.69	0.70	0.58	0.59
Cyprus	0.90	48	0.74	0.82	0.82	0.74	0.78	0.64	0.81	0.53	0.60	0.63	0.57	0.59
Czech Republic	0.90	49	0.80	0.84	0.81	0.75	0.79	0.58	0.83	0.51	0.66	0.69	0.46	0.52
Denmark	0.88	44	0.75	0.78	0.71	0.61	0.75	0.61	0.75	0.57	0.67	0.71	0.48	0.42
England	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Finland	0.91	52	0.76	0.84	0.82	0.77	0.78	0.63	0.82	0.67	0.70	0.73	0.52	0.56
France	0.88	44	0.67	0.76	0.74	0.68	0.73	0.60	0.73	0.60	0.66	0.68	0.55	0.51
Georgia	0.89	46	0.77	0.83	0.82	0.72	0.79	0.76	0.80	0.36	0.49	0.57	0.51	0.53
Germany	0.88	44	0.74	0.79	0.76	0.67	0.75	0.49	0.72	0.50	0.66	0.73	0.50	0.51
Hong Kong SAR	0.82	38	0.76	0.83	0.83	0.71	0.80	0.75	0.76	0.15	0.17	0.16	0.37	0.36
Hungary	0.90	48	0.78	0.84	0.79	0.70	0.81	0.64	0.81	0.48	0.65	0.70	0.46	0.47
Iran, Islamic Rep. of	0.91	51	0.77	0.82	0.80	0.74	0.81	0.74	0.78	0.48	0.66	0.68	0.60	0.57
Ireland	0.89	49	0.75	0.81	0.82	0.79	0.81	0.75	0.82	0.52	0.55	0.57	0.53	0.56
Italy	0.90	49	0.73	0.80	0.78	0.71	0.77	0.62	0.79	0.57	0.69	0.72	0.57	0.57
Japan	0.86	47	0.84	0.85	0.87	0.82	0.87	0.77	0.86	0.19	0.28	0.33	0.52	0.51
Kazakhstan	0.89	47	0.70	0.80	0.82	0.77	0.74	0.73	0.79	0.53	0.59	0.61	0.49	0.51
Korea, Rep. of	0.85	47	0.87	0.89	0.92	0.90	0.89	0.70	0.89	0.13	0.12	0.16	0.42	0.43
Kosovo	0.87	43	0.72	0.76	0.75	0.64	0.74	0.64	0.75	0.52	0.60	0.60	0.52	0.54
Kuwait	0.88	44	0.74	0.82	0.80	0.72	0.77	0.67	0.77	0.48	0.54	0.56	0.46	0.49
Latvia	0.88	44	0.69	0.79	0.81	0.75	0.72	0.64	0.76	0.58	0.61	0.62	0.42	0.44
Lithuania	0.90	48	0.74	0.81	0.81	0.78	0.75	0.62	0.79	0.58	0.65	0.64	0.51	0.52
Malta	0.91	51	0.71	0.80	0.79	0.75	0.78	0.74	0.80	0.60	0.67	0.70	0.59	0.58
Montenegro	0.90	49	0.77	0.82	0.81	0.74	0.80	0.68	0.81	0.49	0.59	0.64	0.57	0.58
Morocco	0.93	56	0.77	0.86	0.84	0.73	0.81	0.82	0.83	0.62	0.70	0.71	0.68	0.56
Netherlands	-	-	-	-	-	-	-	-	-	-	-	-	-	-
New Zealand	0.91	50	0.72	0.80	0.80	0.76	0.76	0.62	0.80	0.61	0.68	0.75	0.56	0.54
North Macedonia	0.91	51	0.78	0.83	0.77	0.70	0.79	0.70	0.80	0.56	0.65	0.65	0.63	0.65
Northern Ireland	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Early Literacy and Numeracy Tasks* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item											
			ASBH06A	ASBH06B	ASBH06C	ASBH06D	ASBH06E	ASBH06F	ASBH06G	ASBH07A	ASBH07B	ASBH07C	ASBH07D	ASBH07E
Norway (5)	0.90	48	0.78	0.82	0.77	0.66	0.79	0.62	0.78	0.54	0.70	0.72	0.57	0.51
Oman	0.90	48	0.73	0.81	0.80	0.73	0.75	0.66	0.78	0.59	0.63	0.63	0.57	0.59
Pakistan	0.88	44	0.65	0.71	0.70	0.65	0.74	0.79	0.78	0.45	0.63	0.65	0.54	0.56
Philippines	0.87	42	0.70	0.75	0.77	0.73	0.75	0.67	0.73	0.51	0.55	0.55	0.42	0.49
Poland	0.91	52	0.76	0.82	0.82	0.76	0.79	0.69	0.80	0.59	0.69	0.71	0.55	0.57
Portugal	0.89	46	0.73	0.79	0.76	0.70	0.77	0.63	0.77	0.53	0.63	0.66	0.52	0.53
Qatar	0.89	46	0.70	0.80	0.79	0.73	0.76	0.68	0.77	0.54	0.58	0.61	0.53	0.56
Russian Federation	0.91	52	0.76	0.82	0.83	0.76	0.77	0.76	0.80	0.66	0.70	0.71	0.50	0.51
Saudi Arabia	0.90	49	0.71	0.83	0.80	0.71	0.77	0.72	0.78	0.56	0.64	0.65	0.57	0.60
Serbia	0.91	52	0.77	0.82	0.80	0.73	0.80	0.71	0.82	0.56	0.65	0.68	0.62	0.61
Singapore	0.89	47	0.75	0.81	0.82	0.77	0.77	0.72	0.79	0.56	0.59	0.59	0.45	0.47
Slovak Republic	0.90	47	0.76	0.80	0.75	0.65	0.79	0.60	0.79	0.52	0.68	0.70	0.55	0.56
South Africa (5)	0.85	39	0.66	0.74	0.76	0.73	0.71	0.60	0.71	0.43	0.50	0.52	0.48	0.49
Spain	0.91	50	0.75	0.81	0.83	0.78	0.76	0.68	0.79	0.62	0.66	0.67	0.55	0.52
Sweden	0.91	52	0.79	0.82	0.82	0.74	0.80	0.68	0.81	0.62	0.69	0.71	0.55	0.57
Turkey (5)	0.95	64	0.84	0.87	0.85	0.81	0.87	0.83	0.88	0.67	0.75	0.78	0.71	0.71
United Arab Emirates	0.90	48	0.68	0.78	0.80	0.74	0.74	0.69	0.77	0.66	0.69	0.71	0.50	0.53
United States	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Benchmarking Participants														
Ontario, Canada	0.92	54	0.70	0.81	0.83	0.79	0.77	0.71	0.81	0.69	0.72	0.75	0.57	0.58
Quebec, Canada	0.90	48	0.70	0.79	0.74	0.70	0.76	0.66	0.78	0.63	0.72	0.73	0.55	0.53
Moscow City, Russian Fed.	0.90	50	0.74	0.81	0.81	0.75	0.77	0.75	0.81	0.61	0.66	0.69	0.46	0.48
Madrid, Spain	0.90	49	0.71	0.79	0.81	0.77	0.77	0.68	0.79	0.61	0.64	0.65	0.58	0.51
Abu Dhabi, UAE	0.91	51	0.72	0.81	0.81	0.74	0.77	0.72	0.79	0.67	0.70	0.71	0.51	0.55
Dubai, UAE	0.90	49	0.65	0.77	0.79	0.74	0.76	0.71	0.78	0.67	0.70	0.73	0.52	0.55

A dash (–) indicates comparable data not available.

Relationship Between the TIMSS 2019 *Early Literacy and Numeracy Tasks* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Albania	0.31	0.09	0.08	0.27	0.07	0.06
Armenia	0.12	0.01	0.01	0.13	0.02	0.01
Australia	-	-	-	-	-	-
Austria	0.13	0.02	0.01	0.03	0.00	0.00
Azerbaijan	0.24	0.06	0.04	0.19	0.04	0.03
Bahrain	0.17	0.03	0.02	0.28	0.08	0.06
Belgium (Flemish)	0.11	0.01	0.01	0.03	0.00	0.00
Bosnia and Herzegovina	0.27	0.07	0.06	0.22	0.05	0.04
Bulgaria	0.46	0.21	0.17	0.48	0.23	0.18
Canada	0.32	0.11	0.09	0.28	0.08	0.06
Chile	0.37	0.14	0.12	0.33	0.11	0.10
Chinese Taipei	0.29	0.09	0.07	0.27	0.07	0.06
Croatia	0.35	0.13	0.10	0.30	0.09	0.07
Cyprus	0.26	0.07	0.06	0.24	0.06	0.06
Czech Republic	0.24	0.06	0.05	0.19	0.04	0.03
Denmark	0.28	0.08	0.06	0.21	0.05	0.04
England	-	-	-	-	-	-
Finland	0.46	0.21	0.17	0.34	0.11	0.09
France	0.30	0.09	0.07	0.20	0.04	0.04
Georgia	0.14	0.02	0.02	0.15	0.02	0.02
Germany	0.15	0.02	0.02	0.06	0.00	0.00
Hong Kong SAR	0.34	0.11	0.09	0.31	0.10	0.08
Hungary	0.25	0.06	0.04	0.19	0.04	0.02
Iran, Islamic Rep. of	0.15	0.02	0.02	0.14	0.02	0.01
Ireland	0.37	0.14	0.11	0.34	0.12	0.09
Italy	0.16	0.03	0.03	0.12	0.01	0.02
Japan	0.36	0.13	0.11	0.30	0.09	0.08
Kazakhstan	0.19	0.04	0.03	0.18	0.03	0.03
Korea, Rep. of	0.33	0.11	0.09	0.30	0.09	0.07
Kosovo	0.21	0.05	0.04	0.20	0.04	0.04
Kuwait	0.30	0.09	0.09	0.28	0.08	0.08
Latvia	0.41	0.17	0.14	0.36	0.13	0.11
Lithuania	0.44	0.19	0.16	0.40	0.16	0.13
Malta	0.30	0.09	0.08	0.24	0.06	0.05
Montenegro	0.28	0.08	0.06	0.25	0.06	0.05
Morocco	0.30	0.09	0.10	0.29	0.08	0.09
Netherlands	-	-	-	-	-	-

Relationship Between the TIMSS 2019 *Early Literacy and Numeracy Tasks* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		*r*	r^2	
New Zealand	0.24	0.06	0.05	0.16	0.03	0.02
North Macedonia	0.25	0.06	0.04	0.23	0.05	0.04
Northern Ireland	-	-	-	-	-	-
Norway (5)	0.31	0.10	0.08	0.24	0.06	0.05
Oman	0.31	0.09	0.08	0.38	0.14	0.12
Pakistan	0.11	0.01	0.02	0.08	0.01	0.01
Philippines	0.37	0.13	0.11	0.32	0.10	0.09
Poland	0.33	0.11	0.09	0.27	0.07	0.06
Portugal	0.22	0.05	0.04	0.19	0.04	0.03
Qatar	0.27	0.07	0.07	0.28	0.08	0.07
Russian Federation	0.32	0.10	0.09	0.30	0.09	0.08
Saudi Arabia	0.21	0.05	0.04	0.23	0.05	0.05
Serbia	0.43	0.18	0.15	0.40	0.16	0.13
Singapore	0.43	0.19	0.17	0.42	0.17	0.14
Slovak Republic	0.23	0.05	0.04	0.22	0.05	0.02
South Africa (5)	0.25	0.06	0.06	0.24	0.06	0.05
Spain	0.37	0.13	0.11	0.33	0.11	0.10
Sweden	0.36	0.13	0.12	0.26	0.07	0.06
Turkey (5)	0.32	0.10	0.07	0.35	0.12	0.09
United Arab Emirates	0.16	0.03	0.02	0.17	0.03	0.03
United States	-	-	-	-	-	-
International Median	**0.29**	**0.09**	**0.07**	**0.25**	**0.06**	**0.06**
Benchmarking Participants						
Ontario, Canada	0.42	0.18	0.14	0.32	0.10	0.08
Quebec, Canada	0.27	0.07	0.06	0.20	0.04	0.03
Moscow City, Russian Fed.	0.38	0.14	0.13	0.33	0.11	0.09
Madrid, Spain	0.33	0.11	0.08	0.30	0.09	0.07
Abu Dhabi, UAE	0.22	0.05	0.05	0.21	0.04	0.05
Dubai, UAE	0.16	0.03	0.02	0.15	0.02	0.02

A dash (–) indicates comparable data not available.

Home Resources for Learning – Grade 4

About the Scale

The *Home Resources for Learning* scale was created based on students' and parents' reports regarding the availability of five resources listed below.

ASBG04 [T]

Number of books in the home (students):
1) 0-10
2) 11-25
3) 26-100
4) 101-200
5) More than 200

ASBH11 [T]

Number of children's books in the home (parents):
1) 0-10
2) 11-25
3) 26-50
4) 51-100
5) More than 100

ASDG05S [T 1]

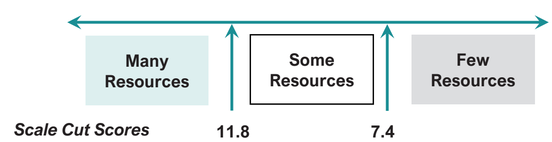
Number of home study supports (students):
1) None
2) Internet connection or own room
3) Both internet connection and own room

ASDHEDUP [T 1]

Highest level of education of either parent (parents):
1) Finished some primary or lower secondary or did not go to school
2) Finished lower secondary
3) Finished upper secondary
4) Finished post-secondary education
5) Finished university or higher

ASDHOCCP [T 1]

Highest level of occupation of either parent (parents):
1) Has never worked outside home for pay, general laborer, or semi-professional (skilled agricultural or fishery worker, craft or trade worker, plant or machine operator)
2) Clerical (clerk or service or sales worker)
3) Small business owner
4) Professional (corporate manager or senior official, professional, or technician or associate professional)



[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.
[1] Derived variable. For details, see Supplement 3 of the *TIMSS 2019 User Guide for the International Database* .

Item Parameters for the TIMSS 2019 *Home Resources for Learning* Scale – Grade 4

Item	delta	tau_1	tau_2	tau_3	tau_4	Infit
ASBG04	0.66771	-1.10357	-0.44084	0.87876	0.66565	1.00
ASBH11	0.73132	-0.67187	-0.46465	0.41378	0.72274	0.95
ASDG05S	-0.95838	-0.70755	0.70755			1.09
ASDHEDUP	-0.48782	-0.47629	-0.91551	0.96162	0.43018	0.94
ASDHOCCP	0.04717	-0.19695	0.76018	-0.56323		1.04

Scale Transformation Constants for the TIMSS 2019 *Home Resources for Learning* Scale – Grade 4

Scale Transformation Constants	
A = 9.451157	Transformed Scale Score = 9.451157 + 1.787338 • Logit Scale Score
B = 1.787338	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Home Resources for Learning* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	3.81993	
1	5.75677	
2	6.70211	
3	7.39134	7.4
4	7.95508	
5	8.43520	
6	8.85649	
7	9.23298	
8	9.58052	
9	9.91481	
10	10.25828	
11	10.61050	
12	10.98891	
13	11.40836	
14	11.89316	11.8
15	12.47492	
16	13.27846	
17	14.87955	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Home Resources for Learning* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item				
			ASBG04	ASBH11	ASDG05S	ASDHEDUP	ASDHOCCP
Albania	0.77	53	0.78	0.80	0.38	0.79	0.78
Armenia	0.59	39	0.53	0.70	0.24	0.77	0.74
Australia	-	-	-	-	-	-	-
Austria	0.73	49	0.73	0.80	0.28	0.79	0.75
Azerbaijan	0.57	36	0.55	0.57	0.41	0.71	0.72
Bahrain	0.56	36	0.60	0.69	0.23	0.72	0.63
Belgium (Flemish)	0.72	48	0.70	0.73	0.42	0.77	0.77
Bosnia and Herzegovina	0.69	46	0.64	0.75	0.18	0.81	0.79
Bulgaria	0.82	58	0.79	0.83	0.39	0.88	0.82
Canada	0.59	39	0.67	0.69	0.29	0.66	0.69
Chile	0.65	42	0.53	0.66	0.34	0.81	0.79
Chinese Taipei	0.71	48	0.77	0.79	0.25	0.78	0.71
Croatia	0.69	46	0.69	0.74	0.14	0.80	0.78
Cyprus	0.68	45	0.67	0.77	0.24	0.77	0.75
Czech Republic	0.64	43	0.69	0.72	0.03	0.78	0.74
Denmark	0.63	42	0.68	0.70	0.26	0.74	0.73
England	-	-	-	-	-	-	-
Finland	0.63	41	0.65	0.68	0.23	0.77	0.74
France	0.75	51	0.76	0.79	0.36	0.81	0.76
Georgia	0.68	45	0.70	0.75	0.37	0.75	0.71
Germany	0.69	45	0.71	0.75	0.28	0.76	0.72
Hong Kong SAR	0.76	51	0.73	0.80	0.36	0.81	0.78
Hungary	0.78	54	0.76	0.82	0.22	0.86	0.82
Iran, Islamic Rep. of	0.71	46	0.69	0.71	0.51	0.77	0.66
Ireland	0.69	46	0.74	0.78	0.06	0.76	0.74
Italy	0.68	45	0.69	0.76	0.11	0.79	0.74
Japan	0.61	40	0.71	0.76	0.20	0.69	0.63
Kazakhstan	0.54	36	0.63	0.63	0.47	0.65	0.61
Korea, Rep. of	0.62	41	0.66	0.73	0.32	0.75	0.64
Kosovo	0.62	39	0.69	0.69	0.30	0.71	0.66
Kuwait	0.41	31	0.46	0.56	0.05	0.74	0.68
Latvia	0.66	44	0.69	0.75	0.15	0.79	0.71
Lithuania	0.72	48	0.72	0.77	0.31	0.79	0.78
Malta	0.64	42	0.57	0.70	0.01	0.82	0.77
Montenegro	0.66	43	0.69	0.75	0.22	0.77	0.70
Morocco	0.68	44	0.72	0.71	0.55	0.73	0.59
Netherlands	-	-	-	-	-	-	-
New Zealand	0.66	43	0.69	0.69	0.38	0.74	0.71
North Macedonia	0.72	48	0.64	0.79	0.15	0.83	0.79

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Home Resources for Learning* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item				
			ASBG04	ASBH11	ASDG05S	ASDHEDUP	ASDHOCCP
Northern Ireland	0.70	46	0.73	0.75	0.08	0.81	0.75
Norway (5)	0.63	48	0.76	0.81	0.43	0.71	-
Oman	0.58	38	0.47	0.62	0.29	0.79	0.76
Pakistan	0.41	31	0.56	0.58	0.58	0.64	0.36
Philippines	0.47	33	0.27	0.51	0.47	0.74	0.74
Poland	0.71	47	0.70	0.74	0.22	0.80	0.78
Portugal	0.72	48	0.69	0.76	0.12	0.81	0.81
Qatar	0.50	34	0.48	0.65	0.32	0.71	0.67
Russian Federation	0.67	44	0.68	0.74	0.34	0.74	0.70
Saudi Arabia	0.47	32	0.35	0.50	0.22	0.79	0.77
Serbia	0.73	49	0.71	0.77	0.27	0.82	0.76
Singapore	0.65	43	0.70	0.73	0.23	0.76	0.71
Slovak Republic	0.78	54	0.79	0.81	0.32	0.82	0.79
South Africa (5)	0.59	36	0.52	0.54	0.54	0.70	0.69
Spain	0.69	46	0.68	0.76	-0.02	0.81	0.77
Sweden	0.70	48	0.72	0.76	0.41	0.78	0.72
Turkey (5)	0.75	50	0.68	0.77	0.55	0.80	0.70
United Arab Emirates	0.48	33	0.52	0.63	0.23	0.67	0.69
United States	-	-	-	-	-	-	-
Benchmarking Participants							
Ontario, Canada	0.58	38	0.67	0.68	0.32	0.64	0.69
Quebec, Canada	0.62	41	0.68	0.71	0.21	0.73	0.69
Moscow City, Russian Fed.	0.59	39	0.65	0.72	0.27	0.70	0.68
Madrid, Spain	0.71	47	0.69	0.76	0.05	0.82	0.80
Abu Dhabi, UAE	0.51	34	0.58	0.63	0.21	0.68	0.68
Dubai, UAE	0.46	33	0.50	0.65	0.24	0.68	0.67

A dash (–) indicates comparable data not available.

Relationship Between the TIMSS 2019 *Home Resources for Learning* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Albania	0.39	0.15	0.09	0.38	0.15	0.09
Armenia	0.26	0.07	0.03	0.27	0.08	0.03
Australia	-	-	-	-	-	-
Austria	0.46	0.21	0.14	0.53	0.29	0.17
Azerbaijan	0.28	0.08	0.03	0.27	0.08	0.03
Bahrain	0.19	0.04	0.02	0.31	0.09	0.04
Belgium (Flemish)	0.42	0.18	0.10	0.48	0.23	0.13
Bosnia and Herzegovina	0.35	0.12	0.06	0.31	0.10	0.04
Bulgaria	0.56	0.31	0.22	0.61	0.38	0.28
Canada	0.27	0.07	0.05	0.32	0.10	0.07
Chile	0.39	0.15	0.07	0.42	0.18	0.08
Chinese Taipei	0.40	0.16	0.11	0.44	0.19	0.13
Croatia	0.36	0.13	0.06	0.37	0.14	0.07
Cyprus	0.35	0.12	0.09	0.38	0.14	0.11
Czech Republic	0.41	0.17	0.11	0.42	0.18	0.11
Denmark	0.32	0.10	0.08	0.37	0.14	0.10
England	-	-	-	-	-	-
Finland	0.37	0.13	0.09	0.38	0.15	0.09
France	0.49	0.24	0.16	0.52	0.27	0.18
Georgia	0.23	0.05	0.03	0.20	0.04	0.03
Germany	0.47	0.22	0.13	0.49	0.24	0.15
Hong Kong SAR	0.37	0.14	0.09	0.38	0.14	0.11
Hungary	0.53	0.28	0.19	0.52	0.27	0.18
Iran, Islamic Rep. of	0.38	0.15	0.10	0.43	0.19	0.12
Ireland	0.43	0.19	0.12	0.45	0.20	0.13
Italy	0.31	0.09	0.04	0.35	0.12	0.06
Japan	0.35	0.12	0.06	0.36	0.13	0.06
Kazakhstan	0.22	0.05	0.02	0.27	0.07	0.05
Korea, Rep. of	0.40	0.16	0.13	0.40	0.16	0.12
Kosovo	0.29	0.08	0.04	0.29	0.09	0.04
Kuwait	0.20	0.04	0.02	0.19	0.03	0.01
Latvia	0.36	0.13	0.07	0.34	0.12	0.07
Lithuania	0.48	0.23	0.10	0.48	0.23	0.10
Malta	0.40	0.16	0.09	0.44	0.19	0.11
Montenegro	0.36	0.13	0.06	0.37	0.14	0.07
Morocco	0.26	0.07	0.05	0.21	0.05	0.04
Netherlands	-	-	-	-	-	-

Relationship Between the TIMSS 2019 *Home Resources for Learning* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		*r*	r^2	
New Zealand	0.45	0.20	0.14	0.49	0.24	0.16
North Macedonia	0.47	0.22	0.13	0.49	0.24	0.14
Northern Ireland	0.44	0.19	0.14	0.44	0.19	0.14
Norway (5)	0.36	0.13	0.09	0.43	0.19	0.11
Oman	0.29	0.08	0.04	0.34	0.11	0.05
Pakistan	0.09	0.01	0.00	0.07	0.01	0.00
Philippines	0.26	0.07	0.03	0.31	0.10	0.05
Poland	0.42	0.18	0.11	0.42	0.18	0.12
Portugal	0.42	0.18	0.11	0.41	0.17	0.11
Qatar	0.38	0.15	0.06	0.35	0.12	0.04
Russian Federation	0.28	0.08	0.03	0.33	0.11	0.05
Saudi Arabia	0.17	0.03	0.01	0.18	0.03	0.02
Serbia	0.51	0.26	0.17	0.51	0.26	0.18
Singapore	0.45	0.20	0.13	0.50	0.25	0.16
Slovak Republic	0.51	0.26	0.18	0.57	0.32	0.25
South Africa (5)	0.36	0.13	0.09	0.39	0.15	0.10
Spain	0.40	0.16	0.10	0.40	0.16	0.10
Sweden	0.43	0.19	0.13	0.48	0.23	0.16
Turkey (5)	0.56	0.32	0.22	0.55	0.30	0.22
United Arab Emirates	0.32	0.10	0.05	0.30	0.09	0.04
United States	-	-	-	-	-	-
International Median	**0.37**	**0.14**	**0.09**	**0.39**	**0.15**	**0.10**
Benchmarking Participants						
Ontario, Canada	0.26	0.07	0.05	0.30	0.09	0.06
Quebec, Canada	0.33	0.11	0.07	0.37	0.14	0.09
Moscow City, Russian Fed.	0.31	0.10	0.06	0.34	0.12	0.07
Madrid, Spain	0.39	0.15	0.11	0.38	0.15	0.10
Abu Dhabi, UAE	0.35	0.12	0.07	0.33	0.11	0.06
Dubai, UAE	0.30	0.09	0.05	0.30	0.09	0.04

A dash (–) indicates comparable data not available.

Instructional Clarity in Mathematics Lessons – Grade 4

About the Scale

The *Instructional Clarity in Mathematics Lessons* scale was created based on students' responses to six items listed below.

How much do you agree with these statements about your mathematics lessons?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
ASBM03A	1) I know what my teacher expects me to do	○	○	○	○
ASBM03B	2) My teacher is easy to understand	○	○	○	○
ASBM03C	3) My teacher has clear answers to my questions	○	○	○	○
ASBM03D	4) My teacher is good at explaining mathematics	○	○	○	○
ASBM03E	5) My teacher does a variety of things to help us learn	○	○	○	○
ASBM03F	6) My teacher explains a topic again when we don't understand	○	○	○	○

High Clarity | Moderate Clarity | Low Clarity

Scale Cut Scores 8.7 6.8

Item Parameters for the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* Scale – Grade 4

Item	delta	tau_1	tau_2	tau_3	Infit
ASBM03A	0.55669	-0.41814	-0.67074	1.08888	1.35
ASBM03B	0.16551	-0.64001	-0.56532	1.20533	0.99
ASBM03C	-0.01951	-0.59199	-0.40578	0.99777	0.90
ASBM03D	-0.32365	-0.23962	-0.42717	0.66679	0.84
ASBM03E	-0.25206	-0.14221	-0.49347	0.63568	0.95
ASBM03F	-0.12698	-0.03204	-0.45776	0.48980	1.02

Scale Transformation Constants for the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* Scale – Grade 4

Scale Transformation Constants	
A = 6.877313	Transformed Scale Score = 6.877313 + 1.528419 • Logit Scale Score
B = 1.528419	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	2.55279	
1	3.94739	
2	4.59152	
3	5.03338	
4	5.37899	
5	5.67743	
6	5.94784	
7	6.20271	
8	6.45082	
9	6.70099	6.8
10	6.95523	
11	7.22643	
12	7.52472	
13	7.86085	
14	8.25838	
15	8.75219	8.7
16	9.40649	
17	10.36648	
18	12.25706	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item					
			ASBM03A	ASBM03B	ASBM03C	ASBM03D	ASBM03E	ASBM03F
Albania	0.46	40	0.21	0.75	0.65	0.80	0.54	0.65
Armenia	0.77	53	0.36	0.73	0.75	0.84	0.85	0.73
Australia	0.82	53	0.61	0.78	0.79	0.80	0.75	0.62
Austria	0.75	48	0.45	0.64	0.77	0.79	0.76	0.69
Azerbaijan	0.67	43	0.38	0.59	0.69	0.78	0.70	0.73
Bahrain	0.75	47	0.47	0.72	0.77	0.78	0.74	0.58
Belgium (Flemish)	0.76	47	0.52	0.69	0.74	0.77	0.67	0.68
Bosnia and Herzegovina	0.80	52	0.56	0.72	0.78	0.81	0.70	0.74
Bulgaria	0.75	47	0.48	0.64	0.77	0.74	0.72	0.72
Canada	0.78	49	0.54	0.76	0.77	0.78	0.72	0.60
Chile	0.77	49	0.49	0.71	0.77	0.79	0.67	0.72
Chinese Taipei	0.86	60	0.72	0.77	0.78	0.80	0.80	0.76
Croatia	0.82	55	0.53	0.74	0.82	0.81	0.74	0.76
Cyprus	0.80	53	0.56	0.63	0.77	0.81	0.79	0.76
Czech Republic	0.85	59	0.59	0.82	0.80	0.82	0.80	0.75
Denmark	0.79	50	0.48	0.76	0.74	0.82	0.75	0.66
England	0.77	48	0.53	0.76	0.78	0.79	0.70	0.56
Finland	0.82	55	0.44	0.81	0.75	0.84	0.81	0.72
France	0.77	47	0.51	0.73	0.75	0.75	0.68	0.68
Georgia	0.66	47	0.34	0.75	0.75	0.80	0.70	0.66
Germany	0.80	52	0.49	0.80	0.76	0.80	0.71	0.72
Hong Kong SAR	0.89	65	0.72	0.81	0.85	0.84	0.84	0.77
Hungary	0.81	52	0.58	0.69	0.79	0.81	0.75	0.70
Iran, Islamic Rep. of	0.80	53	0.51	0.67	0.78	0.83	0.76	0.78
Ireland	0.77	48	0.47	0.75	0.78	0.80	0.72	0.59
Italy	0.64	39	0.34	0.60	0.72	0.76	0.62	0.62
Japan	0.84	58	0.48	0.80	0.82	0.84	0.82	0.76
Kazakhstan	0.69	47	0.39	0.62	0.75	0.80	0.75	0.72
Korea, Rep. of	0.83	57	0.46	0.76	0.80	0.86	0.82	0.76
Kosovo	0.51	42	0.22	0.73	0.69	0.73	0.73	0.64
Kuwait	0.80	51	0.52	0.72	0.79	0.78	0.76	0.69
Latvia	0.77	50	0.46	0.71	0.75	0.80	0.75	0.70
Lithuania	0.72	44	0.51	0.68	0.73	0.74	0.65	0.63
Malta	0.73	44	0.49	0.70	0.75	0.75	0.69	0.59
Montenegro	0.71	43	0.50	0.69	0.72	0.76	0.58	0.66
Morocco	0.62	40	0.32	0.52	0.70	0.74	0.73	0.67
Netherlands	0.79	50	0.57	0.76	0.78	0.75	0.69	0.66
New Zealand	0.81	52	0.57	0.76	0.79	0.80	0.73	0.64
North Macedonia	0.65	41	0.41	0.65	0.74	0.70	0.64	0.64

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item					
			ASBM03A	ASBM03B	ASBM03C	ASBM03D	ASBM03E	ASBM03F
Northern Ireland	0.77	48	0.53	0.78	0.78	0.79	0.68	0.57
Norway (5)	0.80	52	0.48	0.75	0.78	0.81	0.74	0.71
Oman	0.72	45	0.43	0.67	0.72	0.75	0.69	0.70
Pakistan	0.82	54	0.57	0.77	0.77	0.78	0.70	0.78
Philippines	0.70	41	0.42	0.67	0.65	0.71	0.68	0.66
Poland	0.83	55	0.59	0.68	0.79	0.82	0.77	0.78
Portugal	0.68	41	0.45	0.65	0.68	0.75	0.64	0.65
Qatar	0.81	51	0.51	0.74	0.79	0.79	0.76	0.68
Russian Federation	0.76	50	0.49	0.68	0.77	0.79	0.75	0.69
Saudi Arabia	0.75	47	0.46	0.69	0.77	0.77	0.74	0.63
Serbia	0.81	52	0.60	0.73	0.78	0.80	0.67	0.73
Singapore	0.83	54	0.59	0.79	0.80	0.81	0.74	0.66
Slovak Republic	0.77	48	0.52	0.75	0.72	0.78	0.76	0.61
South Africa (5)	0.78	48	0.58	0.70	0.70	0.76	0.68	0.72
Spain	0.73	47	0.34	0.66	0.77	0.79	0.73	0.71
Sweden	0.81	54	0.45	0.77	0.81	0.82	0.79	0.68
Turkey (5)	0.70	44	0.37	0.63	0.70	0.76	0.73	0.72
United Arab Emirates	0.81	53	0.55	0.75	0.79	0.80	0.77	0.69
United States	0.79	50	0.58	0.75	0.77	0.78	0.72	0.61
Benchmarking Participants								
Ontario, Canada	0.79	49	0.55	0.76	0.78	0.78	0.72	0.58
Quebec, Canada	0.78	50	0.52	0.76	0.76	0.77	0.72	0.66
Moscow City, Russian Fed.	0.78	52	0.46	0.70	0.81	0.80	0.77	0.72
Madrid, Spain	0.67	42	0.31	0.67	0.74	0.75	0.67	0.66
Abu Dhabi, UAE	0.84	55	0.58	0.76	0.80	0.80	0.79	0.72
Dubai, UAE	0.76	47	0.50	0.71	0.75	0.78	0.71	0.61

Relationship Between the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Albania	0.10	0.01	0.01
Armenia	0.12	0.01	0.02
Australia	0.13	0.02	0.02
Austria	0.10	0.01	0.01
Azerbaijan	0.17	0.03	0.03
Bahrain	0.12	0.01	0.02
Belgium (Flemish)	0.07	0.01	0.01
Bosnia and Herzegovina	0.16	0.03	0.04
Bulgaria	0.08	0.01	0.01
Canada	0.09	0.01	0.01
Chile	0.15	0.02	0.04
Chinese Taipei	0.21	0.04	0.04
Croatia	0.10	0.01	0.00
Cyprus	0.10	0.01	0.01
Czech Republic	0.06	0.00	0.01
Denmark	0.16	0.03	0.02
England	0.11	0.01	0.01
Finland	0.10	0.01	0.01
France	0.02	0.00	0.00
Georgia	0.09	0.01	0.01
Germany	0.10	0.01	0.01
Hong Kong SAR	0.20	0.04	0.04
Hungary	0.17	0.03	0.02
Iran, Islamic Rep. of	0.13	0.02	0.03
Ireland	0.07	0.00	0.00
Italy	0.09	0.01	0.02
Japan	0.02	0.00	0.00
Kazakhstan	0.11	0.01	0.01
Korea, Rep. of	0.19	0.04	0.04
Kosovo	0.09	0.01	0.03
Kuwait	0.17	0.03	0.04
Latvia	0.12	0.01	0.01
Lithuania	0.03	0.00	0.00
Malta	0.14	0.02	0.02
Montenegro	0.12	0.02	0.02
Morocco	0.15	0.02	0.03
Netherlands	0.04	0.00	0.01
New Zealand	0.03	0.00	0.00

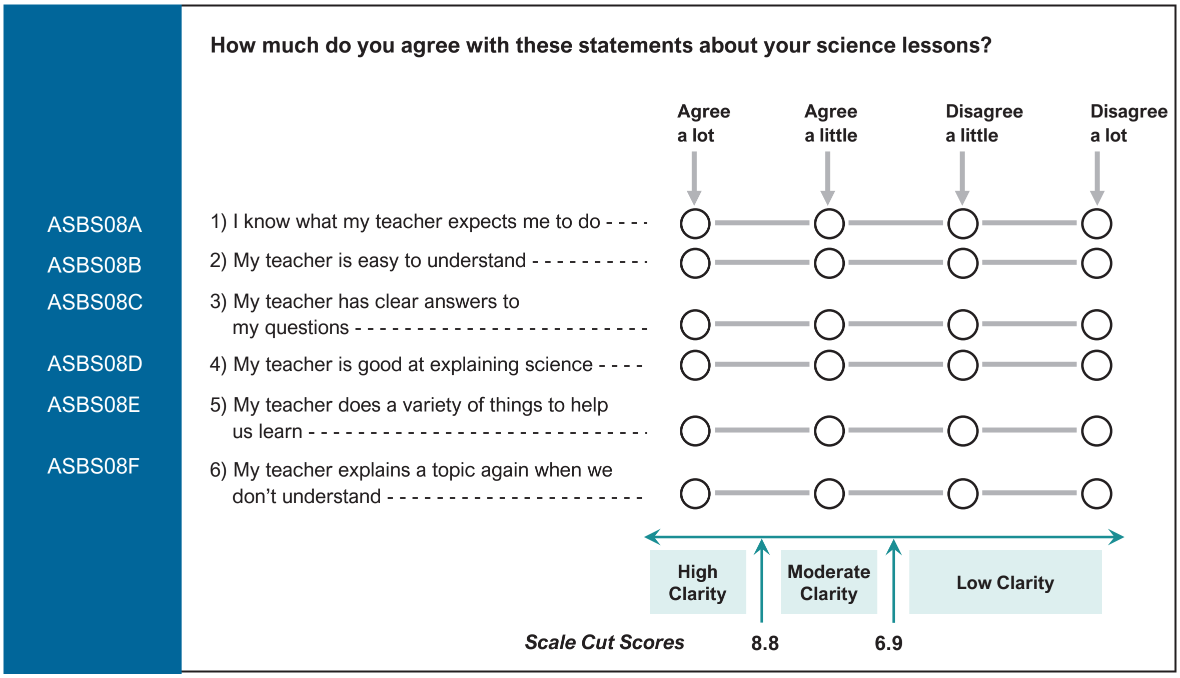
Relationship Between the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
North Macedonia	0.20	0.04	0.03
Northern Ireland	0.10	0.01	0.01
Norway (5)	0.07	0.01	0.01
Oman	0.16	0.03	0.04
Pakistan	0.18	0.03	0.05
Philippines	0.30	0.09	0.09
Poland	0.10	0.01	0.01
Portugal	0.06	0.00	0.01
Qatar	0.18	0.03	0.05
Russian Federation	0.06	0.00	0.01
Saudi Arabia	0.17	0.03	0.04
Serbia	0.10	0.01	0.01
Singapore	0.24	0.06	0.05
Slovak Republic	-0.02	0.00	0.00
South Africa (5)	0.28	0.08	0.09
Spain	0.09	0.01	0.02
Sweden	0.03	0.00	0.01
Turkey (5)	0.29	0.08	0.07
United Arab Emirates	0.18	0.03	0.04
United States	0.14	0.02	0.02
International Median	**0.11**	**0.01**	**0.02**
Benchmarking Participants			
Ontario, Canada	0.11	0.01	0.01
Quebec, Canada	0.09	0.01	0.01
Moscow City, Russian Fed.	0.10	0.01	0.01
Madrid, Spain	0.05	0.00	0.01
Abu Dhabi, UAE	0.18	0.03	0.04
Dubai, UAE	0.09	0.01	0.01

Instructional Clarity in Science Lessons – Grade 4

About the Scale

The *Instructional Clarity in Science Lessons* scale was created based on students' responses to six items listed below.



Item Parameters for the TIMSS 2019 *Instructional Clarity in Science Lessons* Scale – Grade 4

Item	delta	tau_1	tau_2	tau_3	Infit
ASBS08A	0.64445	-0.65910	-0.54914	1.20824	1.43
ASBS08B	0.05025	-0.67012	-0.52376	1.19388	0.96
ASBS08C	-0.14200	-0.73473	-0.33617	1.07090	0.86
ASBS08D	-0.32015	-0.44293	-0.45339	0.89632	0.86
ASBS08E	-0.21665	-0.52139	-0.40110	0.92249	0.93
ASBS08F	-0.01590	-0.31434	-0.42998	0.74432	1.05

Scale Transformation Constants for the TIMSS 2019 *Instructional Clarity in Science Lessons* Scale – Grade 4

Scale Transformation Constants	
A = 6.991185	Transformed Scale Score = 6.991185 + 1.413250 • Logit Scale Score
B = 1.413250	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Instructional Clarity in Science Lessons* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	2.65952	
1	4.04481	
2	4.69156	
3	5.13557	
4	5.48896	
5	5.79290	
6	6.06900	
7	6.32969	
8	6.58700	
9	6.83933	6.9
10	7.10005	
11	7.37696	
12	7.67947	
13	8.01729	
14	8.41140	
15	8.89130	8.8
16	9.51083	
17	10.40151	
18	12.14438	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instructional Clarity in Science Lessons* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item					
			ASBS08A	ASBS08B	ASBS08C	ASBS08D	ASBS08E	ASBS08F
Albania	0.62	48	0.29	0.81	0.71	0.86	0.69	0.67
Armenia	0.82	59	0.47	0.79	0.80	0.83	0.85	0.80
Australia	0.86	59	0.62	0.83	0.85	0.82	0.79	0.68
Austria	0.78	51	0.50	0.63	0.81	0.80	0.78	0.74
Azerbaijan	0.75	50	0.40	0.65	0.79	0.79	0.76	0.77
Bahrain	0.76	49	0.47	0.74	0.81	0.78	0.76	0.59
Belgium (Flemish)	0.80	51	0.59	0.74	0.78	0.75	0.68	0.72
Bosnia and Herzegovina	0.78	51	0.54	0.74	0.79	0.78	0.69	0.71
Bulgaria	0.77	49	0.50	0.70	0.81	0.74	0.70	0.71
Canada	0.83	55	0.58	0.80	0.82	0.81	0.76	0.64
Chile	0.80	52	0.51	0.69	0.79	0.81	0.73	0.74
Chinese Taipei	0.89	65	0.76	0.83	0.81	0.82	0.83	0.79
Croatia	0.86	60	0.62	0.79	0.85	0.84	0.76	0.76
Cyprus	0.87	61	0.63	0.71	0.84	0.85	0.84	0.81
Czech Republic	0.88	64	0.61	0.87	0.82	0.86	0.83	0.79
Denmark	0.84	58	0.52	0.81	0.81	0.85	0.81	0.72
England	0.83	56	0.57	0.82	0.82	0.82	0.77	0.65
Finland	0.85	61	0.49	0.85	0.80	0.87	0.83	0.77
France	0.82	53	0.57	0.78	0.79	0.78	0.72	0.72
Georgia	0.72	50	0.34	0.77	0.79	0.80	0.74	0.67
Germany	0.84	57	0.51	0.84	0.82	0.83	0.75	0.76
Hong Kong SAR	0.94	76	0.80	0.89	0.90	0.89	0.90	0.84
Hungary	0.84	57	0.58	0.75	0.84	0.84	0.76	0.73
Iran, Islamic Rep. of	0.80	53	0.46	0.74	0.78	0.82	0.76	0.77
Ireland	0.83	55	0.55	0.82	0.81	0.83	0.76	0.66
Italy	0.70	43	0.37	0.63	0.77	0.77	0.67	0.63
Japan	0.87	63	0.56	0.82	0.84	0.85	0.85	0.79
Kazakhstan	0.75	51	0.44	0.67	0.76	0.81	0.81	0.72
Korea, Rep. of	0.87	63	0.52	0.80	0.84	0.88	0.85	0.82
Kosovo	0.62	49	0.25	0.79	0.80	0.79	0.72	0.69
Kuwait	0.80	52	0.50	0.72	0.79	0.80	0.76	0.72
Latvia	0.81	54	0.53	0.77	0.79	0.76	0.78	0.73
Lithuania	0.76	47	0.57	0.70	0.74	0.73	0.71	0.68
Malta	0.82	54	0.58	0.78	0.80	0.78	0.74	0.68
Montenegro	0.80	52	0.55	0.75	0.79	0.78	0.68	0.75
Morocco	0.68	42	0.37	0.55	0.74	0.74	0.74	0.68
Netherlands	0.83	55	0.64	0.81	0.83	0.72	0.73	0.70
New Zealand	0.87	61	0.68	0.82	0.84	0.80	0.80	0.73
North Macedonia	0.71	47	0.39	0.69	0.78	0.81	0.69	0.67

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instructional Clarity in Science Lessons* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item					
			ASBS08A	ASBS08B	ASBS08C	ASBS08D	ASBS08E	ASBS08F
Northern Ireland	0.82	54	0.59	0.80	0.82	0.77	0.76	0.64
Norway (5)	0.81	54	0.56	0.78	0.76	0.80	0.76	0.71
Oman	0.76	48	0.37	0.73	0.77	0.78	0.73	0.72
Pakistan	0.81	53	0.51	0.74	0.80	0.77	0.71	0.78
Philippines	0.72	42	0.44	0.68	0.67	0.71	0.70	0.66
Poland	0.89	64	0.70	0.78	0.84	0.86	0.81	0.81
Portugal	0.78	51	0.53	0.67	0.77	0.78	0.74	0.76
Qatar	0.82	54	0.51	0.77	0.82	0.81	0.78	0.67
Russian Federation	0.79	54	0.50	0.71	0.81	0.82	0.79	0.72
Saudi Arabia	0.76	49	0.43	0.74	0.77	0.77	0.77	0.65
Serbia	0.84	56	0.65	0.77	0.80	0.80	0.72	0.74
Singapore	0.87	61	0.65	0.82	0.83	0.84	0.80	0.73
Slovak Republic	0.81	53	0.55	0.79	0.74	0.81	0.77	0.69
South Africa (5)	0.80	50	0.59	0.74	0.74	0.76	0.70	0.72
Spain	0.76	50	0.34	0.70	0.80	0.80	0.76	0.76
Sweden	0.83	58	0.50	0.80	0.84	0.83	0.82	0.72
Turkey (5)	0.76	50	0.43	0.68	0.74	0.80	0.77	0.75
United Arab Emirates	0.84	57	0.58	0.78	0.82	0.82	0.79	0.71
United States	0.83	56	0.62	0.80	0.81	0.81	0.76	0.66
Benchmarking Participants								
Ontario, Canada	0.82	54	0.59	0.79	0.82	0.79	0.75	0.65
Quebec, Canada	0.84	58	0.56	0.82	0.83	0.83	0.78	0.71
Moscow City, Russian Fed.	0.81	56	0.51	0.74	0.83	0.83	0.81	0.71
Madrid, Spain	0.75	49	0.38	0.70	0.78	0.78	0.74	0.72
Abu Dhabi, UAE	0.86	60	0.62	0.80	0.83	0.82	0.80	0.73
Dubai, UAE	0.82	54	0.56	0.76	0.81	0.81	0.77	0.66

Relationship Between the TIMSS 2019 *Instructional Clarity in Science Lessons* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Albania	0.12	0.02	0.02
Armenia	0.12	0.02	0.02
Australia	-0.01	0.00	0.00
Austria	0.06	0.00	0.00
Azerbaijan	0.12	0.01	0.02
Bahrain	0.19	0.03	0.04
Belgium (Flemish)	0.04	0.00	0.00
Bosnia and Herzegovina	0.14	0.02	0.03
Bulgaria	0.08	0.01	0.01
Canada	0.04	0.00	0.00
Chile	0.13	0.02	0.03
Chinese Taipei	0.13	0.02	0.02
Croatia	0.07	0.00	0.01
Cyprus	0.07	0.01	0.01
Czech Republic	0.03	0.00	0.01
Denmark	0.04	0.00	0.00
England	-0.02	0.00	0.00
Finland	0.07	0.00	0.01
France	-0.01	0.00	0.00
Georgia	0.07	0.00	0.01
Germany	0.08	0.01	0.01
Hong Kong SAR	0.19	0.04	0.04
Hungary	0.07	0.01	0.01
Iran, Islamic Rep. of	0.17	0.03	0.03
Ireland	0.01	0.00	0.00
Italy	0.07	0.01	0.01
Japan	-0.03	0.00	0.00
Kazakhstan	0.16	0.03	0.02
Korea, Rep. of	0.10	0.01	0.01
Kosovo	0.11	0.01	0.04
Kuwait	0.18	0.03	0.05
Latvia	0.01	0.00	0.00
Lithuania	0.03	0.00	0.00
Malta	0.05	0.00	0.00
Montenegro	0.10	0.01	0.01
Morocco	0.16	0.03	0.04
Netherlands	-0.01	0.00	0.01
New Zealand	-0.06	0.00	0.00

Relationship Between the TIMSS 2019 *Instructional Clarity in Science Lessons* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
North Macedonia	0.17	0.03	0.04
Northern Ireland	-0.02	0.00	0.00
Norway (5)	-0.04	0.00	0.00
Oman	0.20	0.04	0.05
Pakistan	0.13	0.02	0.02
Philippines	0.28	0.08	0.09
Poland	0.04	0.00	0.01
Portugal	0.08	0.01	0.01
Qatar	0.23	0.05	0.07
Russian Federation	-0.01	0.00	0.00
Saudi Arabia	0.22	0.05	0.06
Serbia	0.06	0.00	0.00
Singapore	0.12	0.01	0.01
Slovak Republic	-0.01	0.00	0.00
South Africa (5)	0.28	0.08	0.08
Spain	0.11	0.01	0.02
Sweden	-0.02	0.00	0.01
Turkey (5)	0.27	0.07	0.07
United Arab Emirates	0.22	0.05	0.05
United States	0.11	0.01	0.02
International Median	**0.08**	**0.01**	**0.01**
Benchmarking Participants			
Ontario, Canada	0.03	0.00	0.00
Quebec, Canada	0.03	0.00	0.00
Moscow City, Russian Fed.	0.05	0.00	0.01
Madrid, Spain	0.06	0.00	0.01
Abu Dhabi, UAE	0.22	0.05	0.05
Dubai, UAE	0.09	0.01	0.01

Parents' Perceptions of Their Child's School – Grade 4

About the Scale

The *Parents' Perceptions of Their Child's School* scale was created based on students' responses to eight items listed below.[1]

What do you think of your child's school?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
ASBH09A T	1) My child's school does a good job including me in my child's education	○	○	○	○
ASBH09B T	2) My child's school provides a safe environment	○	○	○	○
ASBH09C T	3) My child's school cares about my child's progress in school	○	○	○	○
ASBH09D T	4) My child's school does a good job informing me of his/her progress	○	○	○	○
ASBH09E T	5) My child's school promotes high academic standards	○	○	○	○
ASBH09F T	6) My child's school does a good job in helping him/her become better in reading	○	○	○	○
ASBH09G T	7) My child's school does a good job in helping him/her become better in mathematics	○	○	○	○
ASBH09H T	8) My child's school does a good job in helping him/her become better in science	○	○	○	○

Very Satisfied | Somewhat Satisfied | Less than Satisfied

Scale Cut Scores 9.7 6.7

[1] For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Disagree a little" and "Disagree a lot" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

T Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Parents' Perceptions of Their Child's School* Scale – Grade 4

Item	delta	tau_1	tau_2	Infit
ASBH09A	0.07406	-1.70344	1.70344	1.10
ASBH09B	-0.56525	-1.69657	1.69657	1.36
ASBH09C	-0.49961	-1.72317	1.72317	0.87
ASBH09D	-0.09816	-1.42520	1.42520	1.00
ASBH09E	1.03037	-1.68923	1.68923	1.19
ASBH09F	-0.18620	-1.52570	1.52570	0.88
ASBH09G	-0.01740	-1.58204	1.58204	0.85
ASBH09H	0.26219	-1.62188	1.62188	0.93

Scale Transformation Constants for the TIMSS 2019 *Parents' Perceptions of Their Child's School* Scale – Grade 4

Scale Transformation Constants	
A = 8.205877	Transformed Scale Score = 8.205877 + 0.941262 • Logit Scale Score
B = 0.941262	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Parents' Perceptions of Their Child's School* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	3.91475	
1	5.08120	
2	5.70687	
3	6.18579	
4	6.60143	6.7
5	6.99714	
6	7.39016	
7	7.79429	
8	8.20790	
9	8.61865	
10	9.01615	
11	9.40453	
12	9.79727	9.7
13	10.21729	
14	10.70311	
15	11.34573	
16	12.53666	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Parents' Perceptions of Their Child's School* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			ASBH09A	ASBH09B	ASBH09C	ASBH09D	ASBH09E	ASBH09F	ASBH09G	ASBH09H
Albania	0.86	54	0.75	0.64	0.78	0.74	0.63	0.75	0.81	0.78
Armenia	0.85	50	0.59	0.64	0.78	0.55	0.74	0.80	0.78	0.75
Australia	-	-	-	-	-	-	-	-	-	-
Austria	0.90	59	0.70	0.58	0.82	0.80	0.70	0.81	0.84	0.83
Azerbaijan	0.87	54	0.74	0.67	0.79	0.72	0.69	0.77	0.76	0.73
Bahrain	0.91	61	0.78	0.68	0.81	0.79	0.79	0.78	0.80	0.80
Belgium (Flemish)	0.89	57	0.78	0.64	0.80	0.81	0.61	0.77	0.81	0.78
Bosnia and Herzegovina	0.91	62	0.77	0.66	0.83	0.78	0.73	0.82	0.84	0.84
Bulgaria	0.92	64	0.73	0.70	0.84	0.79	0.80	0.83	0.85	0.85
Canada	0.91	62	0.77	0.57	0.82	0.79	0.79	0.83	0.85	0.82
Chile	0.92	63	0.77	0.69	0.83	0.81	0.77	0.84	0.82	0.81
Chinese Taipei	0.91	63	0.79	0.69	0.84	0.84	0.71	0.78	0.85	0.82
Croatia	0.92	64	0.75	0.68	0.84	0.76	0.79	0.86	0.85	0.86
Cyprus	0.89	57	0.70	0.63	0.80	0.74	0.77	0.80	0.83	0.77
Czech Republic	0.91	63	0.72	0.58	0.82	0.79	0.76	0.85	0.89	0.88
Denmark	0.92	64	0.77	0.69	0.82	0.81	0.82	0.81	0.85	0.81
England	-	-	-	-	-	-	-	-	-	-
Finland	0.90	58	0.74	0.60	0.80	0.75	0.70	0.83	0.84	0.83
France	0.91	60	0.77	0.65	0.82	0.79	0.75	0.79	0.84	0.78
Georgia	0.88	55	0.62	0.66	0.79	0.74	0.72	0.76	0.82	0.81
Germany	0.89	56	0.78	0.59	0.79	0.80	0.57	0.81	0.83	0.80
Hong Kong SAR	0.89	57	0.77	0.53	0.80	0.82	0.74	0.79	0.79	0.77
Hungary	0.91	62	0.77	0.64	0.81	0.79	0.76	0.83	0.84	0.84
Iran, Islamic Rep. of	0.87	53	0.67	0.62	0.77	0.72	0.65	0.79	0.79	0.78
Ireland	0.88	57	0.79	0.56	0.80	0.80	0.77	0.76	0.81	0.71
Italy	0.89	58	0.76	0.52	0.79	0.78	0.73	0.81	0.84	0.82
Japan	0.88	54	0.65	0.61	0.73	0.73	0.75	0.82	0.81	0.78
Kazakhstan	0.90	59	0.73	0.70	0.81	0.72	0.79	0.83	0.80	0.76
Korea, Rep. of	0.89	57	0.65	0.65	0.78	0.77	0.68	0.80	0.83	0.83
Kosovo	0.86	53	0.72	0.65	0.79	0.71	0.61	0.74	0.79	0.79
Kuwait	0.92	64	0.77	0.69	0.84	0.78	0.83	0.83	0.82	0.81
Latvia	0.90	58	0.66	0.61	0.82	0.70	0.79	0.82	0.84	0.82
Lithuania	0.89	59	0.64	0.66	0.80	0.76	0.70	0.84	0.85	0.86
Malta	0.85	50	0.69	0.50	0.77	0.74	0.69	0.74	0.78	0.72
Montenegro	0.91	63	0.78	0.68	0.84	0.78	0.75	0.83	0.84	0.85
Morocco	0.86	53	0.67	0.61	0.76	0.72	0.59	0.81	0.82	0.79
Netherlands	-	-	-	-	-	-	-	-	-	-
New Zealand	0.90	60	0.78	0.54	0.83	0.81	0.78	0.85	0.83	0.75
North Macedonia	0.87	55	0.62	0.68	0.79	0.74	0.65	0.78	0.83	0.83

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Parents' Perceptions of Their Child's School* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			ASBH09A	ASBH09B	ASBH09C	ASBH09D	ASBH09E	ASBH09F	ASBH09G	ASBH09H
Northern Ireland	0.88	56	0.80	0.49	0.82	0.80	0.75	0.76	0.84	0.68
Norway (5)	-	-	-	-	-	-	-	-	-	-
Oman	0.89	57	0.76	0.66	0.80	0.74	0.75	0.76	0.79	0.79
Pakistan	0.87	52	0.66	0.68	0.69	0.73	0.73	0.75	0.75	0.77
Philippines	0.89	56	0.69	0.65	0.74	0.74	0.75	0.79	0.81	0.80
Poland	0.89	56	0.67	0.66	0.84	0.70	0.73	0.81	0.78	0.79
Portugal	0.90	59	0.77	0.60	0.82	0.77	0.71	0.82	0.85	0.80
Qatar	0.90	60	0.75	0.65	0.82	0.77	0.80	0.79	0.80	0.81
Russian Federation	0.90	59	0.61	0.68	0.80	0.73	0.69	0.85	0.87	0.88
Saudi Arabia	0.91	62	0.75	0.66	0.83	0.80	0.77	0.82	0.83	0.82
Serbia	0.91	62	0.77	0.66	0.83	0.78	0.72	0.82	0.84	0.85
Singapore	0.89	58	0.75	0.60	0.82	0.79	0.71	0.77	0.81	0.80
Slovak Republic	0.91	60	0.68	0.61	0.82	0.75	0.78	0.82	0.86	0.85
South Africa (5)	0.86	51	0.69	0.63	0.74	0.72	0.68	0.75	0.77	0.75
Spain	0.90	59	0.75	0.60	0.81	0.78	0.72	0.79	0.83	0.84
Sweden	0.92	63	0.76	0.67	0.82	0.80	0.81	0.82	0.85	0.84
Turkey (5)	0.91	61	0.78	0.71	0.80	0.78	0.73	0.82	0.82	0.81
United Arab Emirates	0.92	63	0.79	0.63	0.84	0.80	0.81	0.81	0.82	0.82
United States	-	-	-	-	-	-	-	-	-	-
Benchmarking Participants										
Ontario, Canada	0.91	63	0.80	0.55	0.83	0.80	0.82	0.83	0.85	0.82
Quebec, Canada	0.90	58	0.72	0.57	0.81	0.77	0.73	0.82	0.85	0.79
Moscow City, Russian Fed.	0.90	60	0.50	0.68	0.81	0.70	0.79	0.87	0.88	0.88
Madrid, Spain	0.90	60	0.78	0.59	0.82	0.81	0.72	0.79	0.83	0.82
Abu Dhabi, UAE	0.92	65	0.78	0.69	0.85	0.81	0.82	0.83	0.82	0.81
Dubai, UAE	0.91	61	0.78	0.57	0.83	0.79	0.79	0.80	0.82	0.81

A dash (–) indicates comparable data not available.

Relationship Between the TIMSS 2019 *Parents' Perceptions of Their Child's School* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Albania	0.06	0.00	0.00	0.07	0.00	0.00
Armenia	0.06	0.00	0.00	0.04	0.00	0.00
Australia	-	-	-	-	-	-
Austria	0.08	0.01	0.01	0.01	0.00	0.00
Azerbaijan	0.05	0.00	0.00	0.03	0.00	0.00
Bahrain	0.06	0.00	0.01	0.14	0.02	0.02
Belgium (Flemish)	0.02	0.00	0.00	-0.03	0.00	0.00
Bosnia and Herzegovina	-0.01	0.00	0.00	-0.02	0.00	0.00
Bulgaria	-0.09	0.01	0.01	-0.11	0.01	0.01
Canada	0.03	0.00	0.00	-0.01	0.00	0.00
Chile	0.04	0.00	0.00	0.00	0.00	0.00
Chinese Taipei	-0.04	0.00	0.00	-0.09	0.01	0.00
Croatia	0.03	0.00	0.00	0.02	0.00	0.00
Cyprus	0.03	0.00	0.00	0.02	0.00	0.00
Czech Republic	-0.02	0.00	0.00	-0.03	0.00	0.00
Denmark	0.11	0.01	0.01	0.07	0.00	0.00
England	-	-	-	-	-	-
Finland	0.06	0.00	0.01	0.04	0.00	0.00
France	0.09	0.01	0.01	0.03	0.00	0.00
Georgia	-0.03	0.00	0.00	-0.05	0.00	0.00
Germany	0.10	0.01	0.01	0.04	0.00	0.00
Hong Kong SAR	0.11	0.01	0.01	0.12	0.01	0.02
Hungary	-0.02	0.00	0.00	-0.04	0.00	0.00
Iran, Islamic Rep. of	0.07	0.01	0.00	0.04	0.00	0.00
Ireland	0.02	0.00	0.00	-0.01	0.00	0.00
Italy	0.05	0.00	0.00	0.06	0.00	0.00
Japan	0.05	0.00	0.00	0.07	0.00	0.00
Kazakhstan	0.02	0.00	0.00	-0.03	0.00	0.00
Korea, Rep. of	0.02	0.00	0.00	-0.02	0.00	0.00
Kosovo	0.02	0.00	0.00	0.01	0.00	0.00
Kuwait	0.13	0.02	0.02	0.14	0.02	0.02
Latvia	0.01	0.00	0.00	0.01	0.00	0.00
Lithuania	0.04	0.00	0.00	0.04	0.00	0.00
Malta	0.01	0.00	0.00	-0.05	0.00	0.00
Montenegro	-0.01	0.00	0.00	-0.03	0.00	0.00
Morocco	0.14	0.02	0.01	0.13	0.02	0.02
Netherlands	-	-	-	-	-	-

Relationship Between the TIMSS 2019 *Parents' Perceptions of Their Child's School* Scale and TIMSS 2019 Achievement – Grade 4

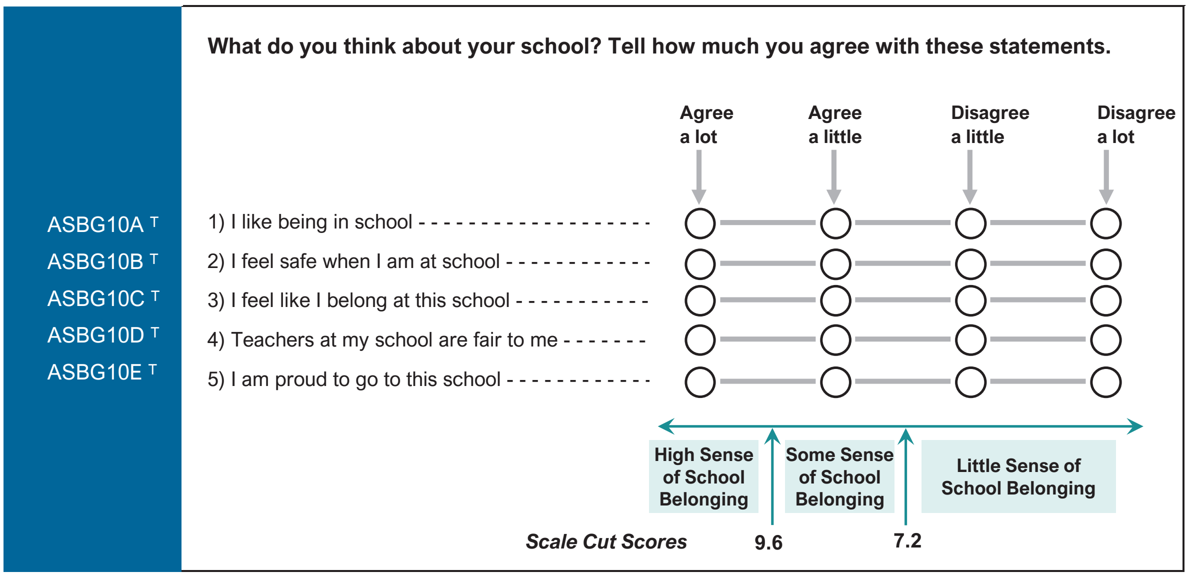
Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		*r*	r^2	
New Zealand	0.03	0.00	0.00	0.01	0.00	0.00
North Macedonia	-0.03	0.00	0.00	-0.03	0.00	0.00
Northern Ireland	0.12	0.01	0.02	0.09	0.01	0.02
Norway (5)	-	-	-	-	-	-
Oman	0.19	0.04	0.03	0.18	0.03	0.03
Pakistan	0.21	0.04	0.03	0.22	0.05	0.03
Philippines	0.19	0.04	0.02	0.20	0.04	0.03
Poland	-0.03	0.00	0.00	-0.06	0.00	0.00
Portugal	0.05	0.00	0.01	0.03	0.00	0.00
Qatar	0.08	0.01	0.01	0.12	0.02	0.02
Russian Federation	-0.02	0.00	0.00	-0.03	0.00	0.00
Saudi Arabia	0.04	0.00	0.00	0.07	0.01	0.01
Serbia	-0.07	0.01	0.01	-0.08	0.01	0.01
Singapore	0.09	0.01	0.01	0.07	0.00	0.01
Slovak Republic	-0.06	0.00	0.00	-0.13	0.02	0.01
South Africa (5)	0.18	0.03	0.02	0.16	0.03	0.02
Spain	0.05	0.00	0.01	0.03	0.00	0.00
Sweden	0.05	0.00	0.00	0.02	0.00	0.00
Turkey (5)	-0.03	0.00	0.00	-0.04	0.00	0.00
United Arab Emirates	0.10	0.01	0.01	0.11	0.01	0.01
United States	-	-	-	-	-	-
International Median	**0.04**	**0.00**	**0.00**	**0.02**	**0.00**	**0.00**
Benchmarking Participants						
Ontario, Canada	0.07	0.01	0.01	-0.02	0.00	0.00
Quebec, Canada	0.03	0.00	0.00	-0.02	0.00	0.00
Moscow City, Russian Fed.	-0.04	0.00	0.00	-0.07	0.00	0.01
Madrid, Spain	0.04	0.00	0.00	0.04	0.00	0.00
Abu Dhabi, UAE	0.12	0.01	0.02	0.12	0.01	0.01
Dubai, UAE	0.13	0.02	0.01	0.13	0.02	0.02

A dash (–) indicates comparable data not available.

Sense of School Belonging – Grade 4

About the Scale

The *Sense of School Belonging* scale was created based on students' responses to five items listed below.



[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Sense of School Belonging* Scale – Grade 4

Item	delta	tau_1	tau_2	tau_3	Infit
ASBG10A	0.38967	-0.43337	-0.67867	1.11204	1.06
ASBG10B	-0.10231	-0.48418	-0.45813	0.94231	0.99
ASBG10C	0.06274	-0.30275	-0.38949	0.69224	1.00
ASBG10D	-0.21351	-0.22387	-0.45259	0.67646	1.11
ASBG10E	-0.13659	-0.09719	-0.47752	0.57471	0.92

Scale Transformation Constants for the TIMSS 2019 *Sense of School Belonging* Scale – Grade 4

Scale Transformation Constants	
A = 7.434179	Transformed Scale Score = 7.434179 + 1.647177 • Logit Scale Score
B = 1.647177	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Sense of School Belonging* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	3.14477	
1	4.60013	
2	5.28784	
3	5.76505	
4	6.15081	
5	6.49005	
6	6.80548	
7	7.11672	7.2
8	7.42069	
9	7.74564	
10	8.10425	
11	8.51746	
12	9.02423	
13	9.69880	9.6
14	10.71074	
15	12.75064	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Sense of School Belonging* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item				
			ASBG10A	ASBG10B	ASBG10C	ASBG10D	ASBG10E
Albania	0.62	40	0.53	0.63	0.63	0.66	0.68
Armenia	0.77	54	0.77	0.73	0.64	0.73	0.78
Australia	0.81	57	0.74	0.73	0.78	0.70	0.83
Austria	0.76	52	0.73	0.78	0.72	0.58	0.77
Azerbaijan	0.72	48	0.70	0.67	0.70	0.63	0.75
Bahrain	0.78	53	0.73	0.73	0.77	0.60	0.79
Belgium (Flemish)	0.76	51	0.70	0.70	0.74	0.67	0.76
Bosnia and Herzegovina	0.76	52	0.71	0.70	0.73	0.65	0.80
Bulgaria	0.77	52	0.69	0.69	0.79	0.63	0.81
Canada	0.80	56	0.73	0.73	0.79	0.68	0.81
Chile	0.75	51	0.67	0.75	0.77	0.62	0.77
Chinese Taipei	0.76	52	0.73	0.74	0.75	0.67	0.70
Croatia	0.77	52	0.70	0.78	0.71	0.64	0.78
Cyprus	0.81	57	0.76	0.72	0.75	0.70	0.83
Czech Republic	0.76	51	0.70	0.76	0.64	0.71	0.77
Denmark	0.80	56	0.74	0.76	0.79	0.64	0.79
England	0.81	57	0.75	0.75	0.77	0.68	0.82
Finland	0.84	61	0.79	0.80	0.78	0.75	0.78
France	0.68	46	0.68	0.69	0.69	0.49	0.79
Georgia	0.63	42	0.66	0.64	0.58	0.61	0.74
Germany	0.78	53	0.70	0.77	0.74	0.65	0.78
Hong Kong SAR	0.84	61	0.80	0.81	0.84	0.69	0.76
Hungary	0.77	53	0.67	0.70	0.75	0.70	0.82
Iran, Islamic Rep. of	0.64	44	0.72	0.70	0.75	0.65	0.46
Ireland	0.79	56	0.70	0.77	0.76	0.69	0.80
Italy	0.75	51	0.69	0.71	0.71	0.66	0.78
Japan	0.80	56	0.78	0.75	0.72	0.68	0.82
Kazakhstan	0.76	52	0.73	0.65	0.74	0.71	0.76
Korea, Rep. of	0.79	55	0.76	0.76	0.72	0.68	0.79
Kosovo	0.59	40	0.58	0.60	0.64	0.61	0.73
Kuwait	0.79	54	0.74	0.73	0.79	0.64	0.77
Latvia	0.76	51	0.72	0.72	0.73	0.63	0.77
Lithuania	0.71	48	0.74	0.66	0.63	0.63	0.78
Malta	0.78	53	0.75	0.73	0.75	0.62	0.79
Montenegro	0.72	48	0.71	0.67	0.68	0.61	0.78
Morocco	0.67	44	0.66	0.65	0.70	0.57	0.73
Netherlands	0.78	54	0.74	0.72	0.76	0.68	0.76
New Zealand	0.80	57	0.73	0.73	0.79	0.67	0.83
North Macedonia	0.62	41	0.58	0.67	0.66	0.60	0.69

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Sense of School Belonging* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item				
			ASBG10A	ASBG10B	ASBG10C	ASBG10D	ASBG10E
Northern Ireland	0.78	55	0.71	0.75	0.77	0.67	0.80
Norway (5)	0.77	52	0.68	0.70	0.78	0.66	0.79
Oman	0.76	51	0.72	0.73	0.73	0.62	0.76
Pakistan	0.79	54	0.65	0.77	0.75	0.77	0.74
Philippines	0.66	43	0.69	0.66	0.65	0.53	0.73
Poland	0.78	54	0.75	0.71	0.73	0.65	0.81
Portugal	0.75	51	0.74	0.72	0.69	0.58	0.81
Qatar	0.79	54	0.72	0.76	0.78	0.62	0.78
Russian Federation	0.78	54	0.75	0.71	0.73	0.69	0.78
Saudi Arabia	0.75	50	0.68	0.70	0.76	0.63	0.75
Serbia	0.79	55	0.74	0.72	0.78	0.66	0.81
Singapore	0.80	55	0.76	0.74	0.78	0.66	0.79
Slovak Republic	0.76	52	0.70	0.73	0.72	0.66	0.78
South Africa (5)	0.70	46	0.70	0.69	0.71	0.54	0.74
Spain	0.75	51	0.69	0.72	0.71	0.65	0.78
Sweden	0.82	58	0.72	0.74	0.80	0.71	0.83
Turkey (5)	0.66	44	0.64	0.71	0.72	0.47	0.74
United Arab Emirates	0.80	56	0.75	0.75	0.77	0.65	0.80
United States	0.82	58	0.73	0.74	0.81	0.70	0.84
Benchmarking Participants							
Ontario, Canada	0.80	56	0.71	0.73	0.79	0.69	0.81
Quebec, Canada	0.81	56	0.77	0.72	0.78	0.67	0.81
Moscow City, Russian Fed.	0.80	56	0.79	0.73	0.75	0.67	0.79
Madrid, Spain	0.75	52	0.70	0.72	0.71	0.69	0.79
Abu Dhabi, UAE	0.80	55	0.75	0.75	0.77	0.64	0.81
Dubai, UAE	0.78	53	0.74	0.74	0.76	0.64	0.77

Relationship Between the TIMSS 2019 *Sense of School Belonging* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Albania	0.03	0.00	0.00	0.04	0.00	0.00
Armenia	0.05	0.00	0.00	0.06	0.00	0.01
Australia	0.15	0.02	0.02	0.13	0.02	0.01
Austria	0.11	0.01	0.01	0.09	0.01	0.01
Azerbaijan	0.08	0.01	0.01	0.03	0.00	0.00
Bahrain	0.04	0.00	0.00	0.11	0.01	0.01
Belgium (Flemish)	0.11	0.01	0.02	0.13	0.02	0.02
Bosnia and Herzegovina	0.03	0.00	0.00	0.03	0.00	0.00
Bulgaria	-0.14	0.02	0.01	-0.15	0.02	0.01
Canada	0.05	0.00	0.01	0.06	0.00	0.01
Chile	0.12	0.02	0.02	0.11	0.01	0.02
Chinese Taipei	0.07	0.01	0.01	0.07	0.00	0.01
Croatia	0.04	0.00	0.00	0.07	0.00	0.01
Cyprus	0.03	0.00	0.00	0.05	0.00	0.00
Czech Republic	0.02	0.00	0.01	-0.01	0.00	0.01
Denmark	0.14	0.02	0.02	0.14	0.02	0.02
England	0.12	0.01	0.02	0.10	0.01	0.01
Finland	0.08	0.01	0.01	0.04	0.00	0.01
France	0.09	0.01	0.01	0.09	0.01	0.01
Georgia	-0.01	0.00	0.00	-0.04	0.00	0.00
Germany	0.15	0.02	0.02	0.12	0.01	0.02
Hong Kong SAR	0.15	0.02	0.02	0.15	0.02	0.02
Hungary	0.03	0.00	0.00	0.01	0.00	0.00
Iran, Islamic Rep. of	-0.07	0.01	0.01	-0.09	0.01	0.01
Ireland	0.12	0.01	0.02	0.11	0.01	0.02
Italy	0.10	0.01	0.01	0.08	0.01	0.01
Japan	0.12	0.01	0.01	0.08	0.01	0.01
Kazakhstan	0.08	0.01	0.01	0.11	0.01	0.01
Korea, Rep. of	0.14	0.02	0.02	0.12	0.01	0.01
Kosovo	0.06	0.00	0.01	0.10	0.01	0.02
Kuwait	0.06	0.00	0.01	0.11	0.01	0.02
Latvia	0.09	0.01	0.01	0.08	0.01	0.01
Lithuania	0.02	0.00	0.00	0.02	0.00	0.00
Malta	0.06	0.00	0.01	0.02	0.00	0.00
Montenegro	0.02	0.00	0.00	-0.03	0.00	0.00
Morocco	0.12	0.01	0.02	0.15	0.02	0.02
Netherlands	0.10	0.01	0.01	0.10	0.01	0.02

Relationship Between the TIMSS 2019 *Sense of School Belonging* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
New Zealand	0.10	0.01	0.01	0.10	0.01	0.01
North Macedonia	0.03	0.00	0.01	0.05	0.00	0.01
Northern Ireland	0.16	0.03	0.03	0.15	0.02	0.02
Norway (5)	0.09	0.01	0.01	0.07	0.00	0.01
Oman	0.11	0.01	0.02	0.11	0.01	0.02
Pakistan	0.07	0.00	0.01	0.02	0.00	0.00
Philippines	0.16	0.03	0.04	0.14	0.02	0.03
Poland	-0.02	0.00	0.00	-0.02	0.00	0.00
Portugal	0.02	0.00	0.00	0.01	0.00	0.00
Qatar	0.10	0.01	0.01	0.15	0.02	0.03
Russian Federation	0.05	0.00	0.00	-0.01	0.00	0.00
Saudi Arabia	0.08	0.01	0.01	0.09	0.01	0.01
Serbia	-0.06	0.00	0.00	-0.07	0.01	0.00
Singapore	0.12	0.01	0.01	0.08	0.01	0.01
Slovak Republic	-0.07	0.00	0.00	-0.13	0.02	0.01
South Africa (5)	0.12	0.01	0.02	0.12	0.01	0.02
Spain	0.08	0.01	0.01	0.09	0.01	0.01
Sweden	0.12	0.02	0.01	0.09	0.01	0.01
Turkey (5)	0.13	0.02	0.02	0.11	0.01	0.02
United Arab Emirates	0.16	0.03	0.02	0.18	0.03	0.03
United States	0.18	0.03	0.04	0.16	0.02	0.03
International Median	**0.08**	**0.01**	**0.01**	**0.09**	**0.01**	**0.01**
Benchmarking Participants						
Ontario, Canada	0.08	0.01	0.01	0.06	0.00	0.00
Quebec, Canada	0.06	0.00	0.01	0.05	0.00	0.01
Moscow City, Russian Fed.	0.07	0.01	0.01	0.03	0.00	0.00
Madrid, Spain	-0.01	0.00	0.00	0.01	0.00	0.00
Abu Dhabi, UAE	0.09	0.01	0.01	0.12	0.01	0.02
Dubai, UAE	0.08	0.01	0.01	0.08	0.01	0.01

Student Bullying – Grade 4

About the Scale

The *Student Bullying* scale was created based on students' responses to eleven items listed below.

During this school year, how often have other students from your school done any of the following things to you, including through texting or the Internet?

		Never	A few times a year	Once or twice a month	At least once a week
ASBG11A	1) Made fun of me or called me names	○	○	○	○
ASBG11B	2) Left me out of their games or activities	○	○	○	○
ASBG11C	3) Spread lies about me	○	○	○	○
ASBG11D	4) Stole something from me	○	○	○	○
ASBG11E	5) Damaged something of mine on purpose	○	○	○	○
ASBG11F	6) Hit or hurt me (*e.g., shoving, hitting, kicking*)	○	○	○	○
ASBG11G	7) Made me do things I didn't want to do	○	○	○	○
ASBG11H	8) Sent me nasty or hurtful messages online	○	○	○	○
ASBG11I	9) Shared nasty or hurtful messages about me online	○	○	○	○
ASBG11J	10) Shared embarrassing photos of me online	○	○	○	○
ASBG11K	11) Threatened me	○	○	○	○

Never or Almost Never	About Monthly	About Weekly

Scale Cut Scores 9.2 7.4

Item Parameters for the TIMSS 2019 *Student Bullying* Scale – Grade 4

Item	delta	tau_1	tau_2	tau_3	Infit
ASBG11A	0.70887	0.25598	-0.31167	0.05569	1.10
ASBG11B	0.49313	0.00135	-0.01632	0.01497	1.19
ASBG11C	0.43075	0.05672	-0.08605	0.02933	1.03
ASBG11D	0.00016	0.34024	-0.14813	-0.19211	1.07
ASBG11E	-0.08448	0.35351	-0.03527	-0.31824	0.96
ASBG11F	0.38543	0.03273	-0.18101	0.14828	1.03
ASBG11G	-0.03872	0.28678	0.08095	-0.36773	0.98
ASBG11H	-0.44333	0.51278	0.34983	-0.86261	0.97
ASBG11I	-0.60515	0.67976	0.53743	-1.21719	0.87
ASBG11J	-0.73072	0.82778	0.68417	-1.51195	0.93
ASBG11K	-0.11594	0.45350	-0.08114	-0.37236	0.96

Scale Transformation Constants for the TIMSS 2019 *Student Bullying* Scale – Grade 4

Scale Transformation Constants	
A = 7.495914	Transformed Scale Score = 7.495914 + 1.758100 • Logit Scale Score
B = 1.758100	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Student Bullying* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	2.86807	
1	4.32091	
2	4.88934	
3	5.25172	
4	5.52624	
5	5.75084	
6	5.94126	
7	6.11702	
8	6.27822	
9	6.42850	
10	6.57077	
11	6.70718	
12	6.83947	
13	6.96901	
14	7.09688	
15	7.22397	
16	7.35113	7.4
17	7.47839	
18	7.60884	
19	7.74099	
20	7.87654	
21	8.01658	
22	8.16227	
23	8.31512	
24	8.47681	
25	8.64712	
26	8.83393	
27	9.04042	
28	9.27434	9.2
29	9.54836	
30	9.88387	
31	10.33461	
32	11.04030	
33	12.71988	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Student Bullying* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item										
			ASBG11A	ASBG11B	ASBG11C	ASBG11D	ASBG11E	ASBG11F	ASBG11G	ASBG11H	ASBG11I	ASBG11J	ASBG11K
Albania	0.85	45	0.59	0.60	0.55	0.63	0.68	0.63	0.69	0.77	0.76	0.77	0.69
Armenia	0.90	55	0.63	0.56	0.66	0.75	0.76	0.74	0.80	0.81	0.82	0.77	0.77
Australia	0.85	41	0.68	0.63	0.70	0.62	0.70	0.69	0.66	0.57	0.57	0.47	0.71
Austria	0.85	41	0.67	0.65	0.71	0.58	0.66	0.67	0.68	0.64	0.62	0.54	0.64
Azerbaijan	0.82	40	0.55	0.52	0.62	0.65	0.65	0.61	0.63	0.68	0.68	0.68	0.66
Bahrain	0.88	45	0.61	0.65	0.68	0.63	0.70	0.67	0.70	0.69	0.70	0.66	0.68
Belgium (Flemish)	0.82	36	0.69	0.66	0.69	0.54	0.62	0.67	0.61	0.49	0.50	0.44	0.65
Bosnia and Herzegovina	0.86	44	0.62	0.60	0.69	0.64	0.73	0.69	0.68	0.71	0.71	0.67	0.58
Bulgaria	0.83	41	0.63	0.44	0.63	0.57	0.67	0.71	0.70	0.68	0.66	0.59	0.70
Canada	0.86	42	0.67	0.62	0.70	0.64	0.71	0.68	0.65	0.63	0.63	0.52	0.70
Chile	0.88	47	0.63	0.66	0.67	0.64	0.71	0.69	0.71	0.71	0.72	0.69	0.72
Chinese Taipei	0.83	39	0.66	0.61	0.66	0.61	0.67	0.66	0.70	0.55	0.54	0.52	0.70
Croatia	0.86	43	0.67	0.66	0.69	0.57	0.71	0.70	0.67	0.68	0.64	0.55	0.66
Cyprus	0.86	42	0.68	0.61	0.70	0.58	0.71	0.69	0.67	0.58	0.65	0.55	0.71
Czech Republic	0.86	44	0.67	0.53	0.70	0.63	0.72	0.66	0.69	0.67	0.69	0.63	0.70
Denmark	0.85	41	0.65	0.64	0.72	0.51	0.65	0.66	0.67	0.65	0.65	0.45	0.71
England	0.86	43	0.70	0.62	0.72	0.63	0.68	0.67	0.66	0.65	0.64	0.48	0.72
Finland	0.85	43	0.70	0.64	0.72	0.59	0.70	0.70	0.66	0.67	0.62	0.49	0.70
France	0.82	37	0.65	0.57	0.68	0.56	0.67	0.66	0.64	0.54	0.58	0.47	0.68
Georgia	0.84	46	0.55	0.45	0.65	0.63	0.69	0.59	0.72	0.77	0.79	0.76	0.73
Germany	0.85	41	0.63	0.60	0.68	0.63	0.68	0.66	0.66	0.64	0.67	0.56	0.66
Hong Kong SAR	0.86	43	0.55	0.62	0.65	0.62	0.73	0.66	0.72	0.68	0.70	0.64	0.69
Hungary	0.82	38	0.63	0.60	0.70	0.59	0.66	0.67	0.66	0.58	0.53	0.45	0.67
Iran, Islamic Rep. of	0.83	39	0.55	0.48	0.62	0.61	0.66	0.63	0.67	0.66	0.67	0.63	0.68
Ireland	0.86	43	0.69	0.65	0.72	0.60	0.69	0.69	0.67	0.63	0.63	0.53	0.67
Italy	0.82	37	0.66	0.60	0.66	0.58	0.65	0.64	0.63	0.54	0.55	0.50	0.69
Japan	0.80	35	0.69	0.65	0.67	0.58	0.60	0.67	0.71	0.43	0.41	0.42	0.60
Kazakhstan	0.83	41	0.57	0.48	0.65	0.59	0.69	0.65	0.68	0.69	0.68	0.65	0.66
Korea, Rep. of	0.81	38	0.64	0.59	0.66	0.51	0.66	0.70	0.68	0.63	0.62	0.39	0.61
Kosovo	0.91	56	0.64	0.67	0.65	0.73	0.79	0.73	0.79	0.79	0.83	0.80	0.79
Kuwait	0.89	49	0.62	0.63	0.67	0.66	0.69	0.69	0.75	0.74	0.75	0.74	0.74
Latvia	0.86	42	0.65	0.57	0.69	0.57	0.72	0.69	0.67	0.65	0.65	0.57	0.69
Lithuania	0.81	36	0.66	0.56	0.67	0.58	0.60	0.68	0.62	0.57	0.52	0.47	0.65
Malta	0.85	41	0.64	0.59	0.69	0.60	0.67	0.67	0.63	0.63	0.63	0.56	0.68
Montenegro	0.86	46	0.60	0.62	0.68	0.69	0.71	0.66	0.71	0.71	0.74	0.70	0.65
Morocco	0.83	38	0.59	0.55	0.60	0.54	0.63	0.64	0.65	0.62	0.65	0.66	0.67
Netherlands	0.85	43	0.68	0.56	0.69	0.63	0.71	0.69	0.68	0.63	0.66	0.56	0.69
New Zealand	0.87	44	0.67	0.61	0.72	0.68	0.73	0.70	0.70	0.63	0.62	0.53	0.71
North Macedonia	0.89	50	0.65	0.62	0.68	0.69	0.74	0.69	0.72	0.74	0.80	0.70	0.75
Northern Ireland	0.85	41	0.71	0.63	0.72	0.57	0.65	0.65	0.64	0.63	0.64	0.48	0.68

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Student Bullying* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item										
			ASBG11A	ASBG11B	ASBG11C	ASBG11D	ASBG11E	ASBG11F	ASBG11G	ASBG11H	ASBG11I	ASBG11J	ASBG11K
Norway (5)	0.85	41	0.67	0.67	0.72	0.58	0.63	0.67	0.65	0.64	0.63	0.49	0.70
Oman	0.85	41	0.59	0.55	0.64	0.57	0.64	0.62	0.67	0.69	0.70	0.69	0.68
Pakistan	0.88	46	0.60	0.57	0.68	0.66	0.68	0.72	0.72	0.69	0.71	0.68	0.72
Philippines	0.80	33	0.41	0.50	0.55	0.57	0.57	0.47	0.64	0.67	0.67	0.62	0.58
Poland	0.87	46	0.67	0.62	0.70	0.66	0.72	0.70	0.69	0.69	0.69	0.61	0.72
Portugal	0.86	45	0.62	0.60	0.65	0.66	0.70	0.68	0.70	0.67	0.66	0.67	0.72
Qatar	0.89	48	0.60	0.63	0.67	0.65	0.73	0.69	0.70	0.74	0.75	0.72	0.71
Russian Federation	0.83	39	0.67	0.45	0.67	0.59	0.65	0.70	0.64	0.61	0.65	0.53	0.66
Saudi Arabia	0.88	47	0.63	0.60	0.65	0.65	0.72	0.69	0.72	0.72	0.73	0.71	0.70
Serbia	0.85	43	0.60	0.62	0.68	0.58	0.69	0.64	0.67	0.68	0.71	0.62	0.69
Singapore	0.84	40	0.60	0.59	0.71	0.62	0.69	0.66	0.67	0.61	0.62	0.49	0.67
Slovak Republic	0.84	41	0.64	0.60	0.69	0.64	0.68	0.65	0.65	0.59	0.62	0.51	0.70
South Africa (5)	0.83	37	0.56	0.60	0.59	0.46	0.60	0.61	0.66	0.66	0.66	0.66	0.63
Spain	0.86	43	0.64	0.62	0.69	0.65	0.69	0.70	0.69	0.62	0.64	0.55	0.71
Sweden	0.85	42	0.68	0.61	0.73	0.59	0.68	0.67	0.64	0.67	0.66	0.49	0.66
Turkey (5)	0.83	40	0.56	0.59	0.65	0.50	0.66	0.64	0.63	0.68	0.70	0.63	0.68
United Arab Emirates	0.89	48	0.60	0.64	0.69	0.66	0.72	0.69	0.71	0.73	0.73	0.68	0.71
United States	0.86	44	0.67	0.64	0.69	0.64	0.71	0.69	0.66	0.64	0.65	0.55	0.70
Benchmarking Participants													
Ontario, Canada	0.86	43	0.68	0.62	0.70	0.65	0.72	0.67	0.63	0.64	0.63	0.52	0.68
Quebec, Canada	0.84	41	0.67	0.59	0.69	0.59	0.67	0.68	0.63	0.62	0.63	0.49	0.71
Moscow City, Russian Fed.	0.85	41	0.68	0.52	0.69	0.61	0.68	0.73	0.65	0.62	0.64	0.50	0.68
Madrid, Spain	0.84	40	0.62	0.65	0.68	0.59	0.69	0.69	0.65	0.58	0.59	0.52	0.72
Abu Dhabi, UAE	0.89	47	0.60	0.61	0.68	0.65	0.73	0.69	0.71	0.73	0.74	0.69	0.71
Dubai, UAE	0.87	45	0.61	0.63	0.68	0.66	0.70	0.69	0.69	0.71	0.70	0.66	0.67

Relationship Between the TIMSS 2019 *Student Bullying* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Albania	0.12	0.01	0.04	0.11	0.01	0.04
Armenia	0.10	0.01	0.02	0.14	0.02	0.03
Australia	0.13	0.02	0.03	0.12	0.01	0.03
Austria	0.17	0.03	0.04	0.21	0.04	0.05
Azerbaijan	0.15	0.02	0.03	0.14	0.02	0.03
Bahrain	0.12	0.01	0.02	0.25	0.06	0.09
Belgium (Flemish)	0.09	0.01	0.01	0.11	0.01	0.02
Bosnia and Herzegovina	0.12	0.01	0.03	0.10	0.01	0.03
Bulgaria	0.16	0.03	0.05	0.16	0.03	0.05
Canada	0.12	0.01	0.02	0.13	0.02	0.03
Chile	0.19	0.04	0.07	0.18	0.03	0.07
Chinese Taipei	0.08	0.01	0.01	0.11	0.01	0.02
Croatia	0.05	0.00	0.02	0.03	0.00	0.01
Cyprus	0.14	0.02	0.03	0.14	0.02	0.03
Czech Republic	0.16	0.02	0.05	0.15	0.02	0.05
Denmark	0.12	0.01	0.02	0.10	0.01	0.02
England	0.13	0.02	0.03	0.12	0.01	0.02
Finland	0.12	0.02	0.03	0.12	0.01	0.03
France	0.15	0.02	0.04	0.13	0.02	0.03
Georgia	0.15	0.02	0.03	0.15	0.02	0.04
Germany	0.19	0.04	0.06	0.18	0.03	0.06
Hong Kong SAR	0.09	0.01	0.02	0.06	0.00	0.01
Hungary	0.09	0.01	0.01	0.08	0.01	0.01
Iran, Islamic Rep. of	0.07	0.00	0.01	0.04	0.00	0.01
Ireland	0.12	0.02	0.03	0.13	0.02	0.03
Italy	0.11	0.01	0.02	0.10	0.01	0.03
Japan	0.06	0.00	0.00	0.06	0.00	0.00
Kazakhstan	0.11	0.01	0.01	0.08	0.01	0.01
Korea, Rep. of	0.03	0.00	0.00	0.02	0.00	0.00
Kosovo	0.18	0.03	0.06	0.22	0.05	0.07
Kuwait	0.18	0.03	0.05	0.21	0.04	0.05
Latvia	0.16	0.03	0.04	0.21	0.04	0.05
Lithuania	0.09	0.01	0.03	0.10	0.01	0.03
Malta	0.13	0.02	0.03	0.11	0.01	0.03
Montenegro	0.17	0.03	0.03	0.15	0.02	0.03
Morocco	0.13	0.02	0.03	0.12	0.01	0.03
Netherlands	0.06	0.00	0.01	0.08	0.01	0.02

Relationship Between the TIMSS 2019 *Student Bullying* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
New Zealand	0.17	0.03	0.05	0.20	0.04	0.06
North Macedonia	0.23	0.05	0.07	0.24	0.06	0.07
Northern Ireland	0.12	0.02	0.03	0.14	0.02	0.04
Norway (5)	0.08	0.01	0.01	0.09	0.01	0.01
Oman	0.20	0.04	0.05	0.19	0.04	0.05
Pakistan	0.12	0.02	0.02	0.15	0.02	0.03
Philippines	0.27	0.07	0.07	0.29	0.09	0.08
Poland	0.12	0.01	0.03	0.13	0.02	0.03
Portugal	0.11	0.01	0.03	0.13	0.02	0.04
Qatar	0.26	0.07	0.10	0.29	0.08	0.12
Russian Federation	0.15	0.02	0.02	0.14	0.02	0.02
Saudi Arabia	0.22	0.05	0.07	0.26	0.07	0.10
Serbia	0.08	0.01	0.02	0.11	0.01	0.02
Singapore	0.18	0.03	0.05	0.17	0.03	0.05
Slovak Republic	0.15	0.02	0.04	0.16	0.02	0.04
South Africa (5)	0.31	0.09	0.11	0.31	0.10	0.12
Spain	0.18	0.03	0.04	0.19	0.04	0.05
Sweden	0.16	0.03	0.03	0.15	0.02	0.03
Turkey (5)	0.20	0.04	0.07	0.19	0.03	0.06
United Arab Emirates	0.21	0.04	0.06	0.23	0.05	0.07
United States	0.15	0.02	0.05	0.15	0.02	0.05
International Median	**0.13**	**0.02**	**0.03**	**0.14**	**0.02**	**0.03**
Benchmarking Participants						
Ontario, Canada	0.13	0.02	0.03	0.13	0.02	0.03
Quebec, Canada	0.14	0.02	0.03	0.12	0.02	0.03
Moscow City, Russian Fed.	0.17	0.03	0.04	0.17	0.03	0.04
Madrid, Spain	0.07	0.00	0.02	0.10	0.01	0.03
Abu Dhabi, UAE	0.21	0.04	0.05	0.23	0.05	0.06
Dubai, UAE	0.11	0.01	0.02	0.12	0.02	0.03

Students Confident in Mathematics – Grade 4

About the Scale

The *Students Confident in Mathematics* scale was created based on students' responses to nine items listed below.

How much do you agree with these statements about mathematics?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
ASBM05A [T]	1) I usually do well in mathematics				
ASBM05B [T]	2) Mathematics is harder for me than for many of my classmates [R]				
ASBM05C [T]	3) I am just not good at mathematics [R]				
ASBM05D [T]	4) I learn things quickly in mathematics				
ASBM05E [T]	5) Mathematics makes me nervous [R]				
ASBM05F [T]	6) I am good at working out difficult mathematics problems				
ASBM05G [T]	7) My teacher tells me I am good at mathematics				
ASBM05H [T]	8) Mathematics is harder for me than any other subject [R]				
ASBM05I [T]	9) Mathematics makes me confused [R]				

Very Confident | Somewhat Confident | Not Confident

Scale Cut Scores **10.7** **8.5**

[R] Reverse coded

[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Students Confident in Mathematics* Scale – Grade 4

Item	delta	tau_1	tau_2	tau_3	Infit
ASBM05A	-0.56785	-0.51008	-0.52132	1.03140	0.91
ASBM05B [R]	0.26875	-0.54868	0.23636	0.31232	1.00
ASBM05C [R]	0.04819	-0.38143	0.20974	0.17169	0.91
ASBM05D	-0.23147	-0.61721	-0.24060	0.85781	0.98
ASBM05E [R]	0.01818	-0.26390	0.17416	0.08974	1.12
ASBM05F	0.24431	-0.77636	-0.21888	0.99524	1.06
ASBM05G	-0.04911	-0.65149	-0.29073	0.94222	1.18
ASBM05H [R]	0.15465	-0.20402	0.18886	0.01516	0.91
ASBM05I [R]	0.11435	-0.32921	0.17352	0.15569	0.98

[R] Reverse coded

Scale Transformation Constants for the TIMSS 2019 *Students Confident in Mathematics* Scale – Grade 4

Scale Transformation Constants	
A = 8.556200	Transformed Scale Score = 8.556200 + 1.689256 • Logit Scale Score
B = 1.689256	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Students Confident in Mathematics* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	2.79993	
1	4.55349	
2	5.36447	
3	5.91005	
4	6.32925	
5	6.67668	
6	6.97600	
7	7.24155	
8	7.48123	
9	7.70454	
10	7.91361	
11	8.11184	
12	8.30222	
13	8.48743	8.5
14	8.66992	
15	8.85216	
16	9.03664	
17	9.22621	
18	9.42421	
19	9.63460	
20	9.86233	
21	10.11082	
22	10.39472	
23	10.72977	10.7
24	11.14456	
25	11.70061	
26	12.55225	
27	14.40638	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Confident in Mathematics* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			ASBM05A	ASBM05B R	ASBM05C R	ASBM05D	ASBM05E R	ASBM05F	ASBM05G	ASBM05H R	ASBM05I R
Albania	0.82	42	0.65	0.68	0.74	0.61	0.58	0.65	0.65	0.69	0.55
Armenia	0.81	40	0.65	0.60	0.73	0.60	0.57	0.60	0.63	0.69	0.63
Australia	0.87	51	0.76	0.73	0.79	0.72	0.62	0.76	0.39	0.78	0.77
Austria	0.89	54	0.77	0.75	0.77	0.76	0.61	0.70	0.69	0.80	0.76
Azerbaijan	0.81	40	0.57	0.64	0.74	0.59	0.60	0.58	0.63	0.69	0.65
Bahrain	0.81	39	0.39	0.68	0.71	0.55	0.67	0.54	0.50	0.76	0.70
Belgium (Flemish)	0.88	52	0.79	0.77	0.82	0.68	0.55	0.73	0.61	0.79	0.70
Bosnia and Herzegovina	0.86	47	0.72	0.69	0.72	0.67	0.63	0.65	0.68	0.73	0.71
Bulgaria	0.88	51	0.75	0.72	0.75	0.75	0.61	0.71	0.67	0.76	0.72
Canada	0.87	50	0.76	0.75	0.77	0.71	0.64	0.71	0.39	0.77	0.77
Chile	0.82	41	0.63	0.60	0.72	0.63	0.61	0.60	0.47	0.73	0.74
Chinese Taipei	0.85	46	0.73	0.67	0.64	0.74	0.56	0.69	0.64	0.73	0.72
Croatia	0.89	52	0.75	0.73	0.72	0.77	0.66	0.69	0.61	0.79	0.77
Cyprus	0.87	49	0.70	0.74	0.77	0.67	0.65	0.67	0.57	0.74	0.75
Czech Republic	0.86	48	0.73	0.70	0.78	0.73	0.64	0.57	0.57	0.76	0.72
Denmark	0.88	50	0.76	0.73	0.81	0.76	0.55	0.66	0.51	0.75	0.75
England	0.88	52	0.77	0.74	0.81	0.74	0.64	0.76	0.35	0.80	0.76
Finland	0.86	48	0.76	0.71	0.80	0.75	0.64	0.70	0.48	0.76	0.57
France	0.88	52	0.77	0.73	0.79	0.74	0.58	0.70	0.55	0.75	0.82
Georgia	0.81	40	0.63	0.69	0.75	0.58	0.60	0.46	0.49	0.73	0.68
Germany	0.90	54	0.76	0.77	0.80	0.74	0.64	0.70	0.65	0.78	0.75
Hong Kong SAR	0.84	45	0.70	0.65	0.78	0.67	0.59	0.70	0.56	0.73	0.63
Hungary	0.88	51	0.78	0.73	0.78	0.73	0.55	0.70	0.72	0.74	0.69
Iran, Islamic Rep. of	0.79	38	0.62	0.52	0.52	0.63	0.64	0.63	0.58	0.69	0.68
Ireland	0.86	50	0.75	0.75	0.78	0.74	0.60	0.75	0.31	0.80	0.73
Italy	0.85	46	0.71	0.70	0.64	0.72	0.65	0.59	0.55	0.75	0.73
Japan	0.87	49	0.72	0.67	0.83	0.70	0.65	0.76	0.43	0.72	0.75
Kazakhstan	0.86	47	0.68	0.71	0.79	0.73	0.58	0.69	0.58	0.73	0.67
Korea, Rep. of	0.88	53	0.84	0.76	0.76	0.75	0.46	0.80	0.53	0.80	0.75
Kosovo	0.74	33	0.53	0.69	0.71	0.41	0.64	0.34	0.49	0.73	0.51
Kuwait	0.76	34	0.29	0.70	0.72	0.41	0.71	0.35	0.38	0.74	0.69
Latvia	0.89	53	0.79	0.73	0.79	0.77	0.61	0.71	0.66	0.77	0.68
Lithuania	0.87	50	0.78	0.70	0.70	0.73	0.62	0.73	0.66	0.76	0.64
Malta	0.87	49	0.71	0.73	0.78	0.69	0.67	0.64	0.51	0.77	0.75
Montenegro	0.84	44	0.66	0.69	0.70	0.61	0.66	0.57	0.60	0.74	0.72
Morocco	0.74	32	0.39	0.67	0.69	0.39	0.67	0.33	0.37	0.72	0.68
Netherlands	0.88	53	0.78	0.74	0.84	0.71	0.49	0.75	0.51	0.79	0.83
New Zealand	0.83	43	0.71	0.64	0.75	0.66	0.55	0.70	0.38	0.74	0.69
North Macedonia	0.82	41	0.59	0.72	0.72	0.55	0.65	0.50	0.58	0.74	0.70
Northern Ireland	0.87	51	0.77	0.73	0.80	0.74	0.63	0.74	0.34	0.78	0.77

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Confident in Mathematics* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			ASBM05A	ASBM05B R	ASBM05C R	ASBM05D	ASBM05E R	ASBM05F	ASBM05G	ASBM05H R	ASBM05I R
Norway (5)	0.87	50	0.75	0.73	0.79	0.74	0.54	0.76	0.43	0.76	0.74
Oman	0.71	31	0.16	0.73	0.65	0.34	0.73	0.19	0.25	0.76	0.72
Pakistan	0.70	30	0.21	0.66	0.71	0.19	0.72	0.21	0.20	0.73	0.74
Philippines	0.54	26	-0.31	0.67	0.60	-0.25	0.57	-0.32	-0.34	0.67	0.61
Poland	0.87	49	0.75	0.71	0.71	0.75	0.58	0.72	0.57	0.77	0.71
Portugal	0.85	46	0.70	0.62	0.77	0.65	0.62	0.68	0.60	0.71	0.73
Qatar	0.79	37	0.37	0.72	0.75	0.43	0.71	0.43	0.39	0.77	0.74
Russian Federation	0.88	52	0.75	0.71	0.79	0.79	0.61	0.71	0.62	0.75	0.73
Saudi Arabia	0.76	36	0.12	0.75	0.74	0.37	0.76	0.24	0.33	0.79	0.76
Serbia	0.90	55	0.77	0.76	0.81	0.72	0.66	0.71	0.68	0.77	0.76
Singapore	0.87	51	0.77	0.75	0.79	0.70	0.63	0.70	0.47	0.78	0.75
Slovak Republic	0.88	50	0.76	0.69	0.80	0.68	0.65	0.65	0.58	0.76	0.75
South Africa (5)	0.68	28	0.20	0.69	0.70	0.21	0.66	0.16	0.12	0.74	0.69
Spain	0.84	44	0.66	0.69	0.73	0.61	0.64	0.61	0.54	0.72	0.75
Sweden	0.87	50	0.76	0.70	0.77	0.76	0.58	0.73	0.43	0.80	0.74
Turkey (5)	0.84	45	0.71	0.69	0.76	0.64	0.50	0.67	0.61	0.75	0.70
United Arab Emirates	0.80	39	0.50	0.70	0.73	0.51	0.66	0.48	0.46	0.75	0.71
United States	0.86	48	0.69	0.75	0.77	0.68	0.63	0.65	0.37	0.80	0.76
Benchmarking Participants											
Ontario, Canada	0.87	50	0.76	0.77	0.79	0.69	0.63	0.71	0.39	0.77	0.78
Quebec, Canada	0.87	50	0.78	0.73	0.71	0.74	0.66	0.74	0.39	0.75	0.78
Moscow City, Russian Fed.	0.90	56	0.76	0.77	0.82	0.81	0.65	0.74	0.60	0.78	0.77
Madrid, Spain	0.84	44	0.68	0.66	0.71	0.62	0.65	0.61	0.54	0.70	0.78
Abu Dhabi, UAE	0.78	36	0.42	0.71	0.74	0.45	0.66	0.42	0.39	0.75	0.70
Dubai, UAE	0.82	42	0.62	0.68	0.75	0.59	0.62	0.58	0.51	0.73	0.70

R Reverse coded

Relationship Between the TIMSS 2019 *Students Confident in Mathematics* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Albania	0.39	0.15	0.16
Armenia	0.37	0.14	0.14
Australia	0.44	0.20	0.19
Austria	0.48	0.23	0.22
Azerbaijan	0.38	0.15	0.15
Bahrain	0.26	0.07	0.06
Belgium (Flemish)	0.46	0.21	0.21
Bosnia and Herzegovina	0.40	0.16	0.16
Bulgaria	0.42	0.18	0.19
Canada	0.44	0.20	0.20
Chile	0.41	0.17	0.17
Chinese Taipei	0.44	0.19	0.18
Croatia	0.44	0.20	0.19
Cyprus	0.48	0.24	0.22
Czech Republic	0.41	0.17	0.17
Denmark	0.47	0.22	0.20
England	0.45	0.20	0.19
Finland	0.42	0.18	0.18
France	0.43	0.18	0.18
Georgia	0.34	0.12	0.12
Germany	0.47	0.22	0.22
Hong Kong SAR	0.39	0.15	0.17
Hungary	0.50	0.25	0.25
Iran, Islamic Rep. of	0.30	0.09	0.09
Ireland	0.41	0.17	0.15
Italy	0.29	0.09	0.08
Japan	0.44	0.19	0.19
Kazakhstan	0.26	0.07	0.06
Korea, Rep. of	0.48	0.23	0.22
Kosovo	0.37	0.13	0.15
Kuwait	0.28	0.08	0.09
Latvia	0.51	0.26	0.25
Lithuania	0.48	0.23	0.21
Malta	0.39	0.16	0.16
Montenegro	0.42	0.17	0.17
Morocco	0.33	0.11	0.10
Netherlands	0.54	0.29	0.28
New Zealand	0.40	0.16	0.15

Relationship Between the TIMSS 2019 *Students Confident in Mathematics* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
North Macedonia	0.43	0.19	0.18
Northern Ireland	0.43	0.19	0.19
Norway (5)	0.41	0.17	0.18
Oman	0.37	0.13	0.14
Pakistan	0.24	0.06	0.05
Philippines	0.33	0.11	0.10
Poland	0.45	0.20	0.20
Portugal	0.48	0.23	0.22
Qatar	0.30	0.09	0.09
Russian Federation	0.38	0.15	0.14
Saudi Arabia	0.35	0.12	0.14
Serbia	0.47	0.22	0.21
Singapore	0.49	0.24	0.25
Slovak Republic	0.43	0.18	0.17
South Africa (5)	0.41	0.17	0.15
Spain	0.46	0.21	0.21
Sweden	0.37	0.13	0.12
Turkey (5)	0.41	0.17	0.17
United Arab Emirates	0.24	0.06	0.06
United States	0.46	0.21	0.23
International Median	**0.42**	**0.17**	**0.17**
Benchmarking Participants			
Ontario, Canada	0.43	0.19	0.18
Quebec, Canada	0.48	0.23	0.24
Moscow City, Russian Fed.	0.48	0.23	0.23
Madrid, Spain	0.48	0.23	0.23
Abu Dhabi, UAE	0.27	0.08	0.08
Dubai, UAE	0.19	0.04	0.04

Students Confident in Science – Grade 4

About the Scale

The *Students Confident in Science* scale was created based on students' responses to seven items listed below.

How much do you agree with these statements about science?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
ASBS09A [T]	1) I usually do well in science	○	○	○	○
ASBS09B [T]	2) Science is harder for me than for many of my classmates [R]	○	○	○	○
ASBS09C [T]	3) I am just not good at science [R]	○	○	○	○
ASBS09D [T]	4) I learn things quickly in science	○	○	○	○
ASBS09E [T]	5) My teacher tells me I am good at science	○	○	○	○
ASBS09F [T]	6) Science is harder for me than any other subject [R]	○	○	○	○
ASBS09G [T]	7) Science makes me confused [R]	○	○	○	○

Very Confident | Somewhat Confident | Not Confident

Scale Cut Scores **10.2** **8.2**

[R] Reverse coded

[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Students Confident in Science* Scale – Grade 4

Item	delta	tau_1	tau_2	tau_3	Infit
ASBS09A	-0.44367	-0.49754	-0.58589	1.08343	1.00
ASBS09B [R]	0.20303	-0.59182	0.16393	0.42789	0.98
ASBS09C [R]	0.09612	-0.41547	0.10720	0.30827	0.92
ASBS09D	-0.22028	-0.69208	-0.23794	0.93002	1.01
ASBS09E	0.18927	-0.75182	-0.22051	0.97233	1.25
ASBS09F [R]	0.09940	-0.27149	0.06208	0.20941	0.93
ASBS09G [R]	0.07613	-0.18729	0.10026	0.08703	1.00

[R] Reverse coded

Scale Transformation Constants for the TIMSS 2019 *Students Confident in Science* Scale – Grade 4

Scale Transformation Constants	
A = 8.285794	Transformed Scale Score = 8.285794 + 1.534620 • Logit Scale Score
B = 1.534620	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Students Confident in Science* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	3.43187	
1	4.99918	
2	5.73357	
3	6.23436	
4	6.62722	
5	6.95609	
6	7.24267	
7	7.50349	
8	7.74367	
9	7.96952	
10	8.18638	8.2
11	8.39905	
12	8.61225	
13	8.83106	
14	9.06148	
15	9.31112	
16	9.58967	
17	9.90692	
18	10.29731	10.2
19	10.81276	
20	11.59621	
21	13.29214	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Confident in Science* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item						
			ASBS09A	ASBS09B R	ASBS09C R	ASBS09D	ASBS09E	ASBS09F R	ASBS09G R
Albania	0.78	44	0.61	0.73	0.76	0.63	0.60	0.72	0.54
Armenia	0.79	44	0.63	0.67	0.72	0.61	0.62	0.72	0.68
Australia	0.83	51	0.72	0.73	0.79	0.72	0.49	0.76	0.74
Austria	0.86	54	0.73	0.75	0.76	0.72	0.65	0.79	0.74
Azerbaijan	0.77	42	0.44	0.73	0.75	0.51	0.49	0.78	0.72
Bahrain	0.77	43	0.27	0.78	0.78	0.52	0.45	0.80	0.75
Belgium (Flemish)	0.85	54	0.78	0.76	0.82	0.69	0.58	0.78	0.70
Bosnia and Herzegovina	0.82	47	0.65	0.72	0.73	0.63	0.60	0.75	0.72
Bulgaria	0.83	50	0.65	0.73	0.73	0.70	0.57	0.78	0.74
Canada	0.84	51	0.74	0.76	0.75	0.74	0.46	0.76	0.76
Chile	0.75	41	0.47	0.72	0.76	0.53	0.32	0.77	0.76
Chinese Taipei	0.82	48	0.62	0.73	0.70	0.66	0.49	0.80	0.78
Croatia	0.85	53	0.71	0.76	0.76	0.70	0.60	0.80	0.75
Cyprus	0.86	55	0.75	0.74	0.78	0.74	0.65	0.75	0.77
Czech Republic	0.84	51	0.70	0.73	0.80	0.70	0.55	0.79	0.71
Denmark	0.84	51	0.73	0.71	0.80	0.75	0.54	0.73	0.71
England	0.84	53	0.73	0.74	0.81	0.74	0.44	0.80	0.76
Finland	0.80	47	0.72	0.73	0.80	0.74	0.50	0.73	0.49
France	0.86	55	0.76	0.75	0.78	0.75	0.53	0.76	0.82
Georgia	0.80	46	0.62	0.73	0.78	0.62	0.49	0.75	0.69
Germany	0.85	52	0.72	0.76	0.76	0.71	0.59	0.78	0.75
Hong Kong SAR	0.76	42	0.49	0.76	0.81	0.47	0.29	0.81	0.71
Hungary	0.86	55	0.75	0.75	0.80	0.73	0.67	0.79	0.68
Iran, Islamic Rep. of	0.78	43	0.51	0.65	0.77	0.57	0.47	0.77	0.78
Ireland	0.83	50	0.73	0.76	0.79	0.71	0.43	0.75	0.71
Italy	0.80	46	0.68	0.68	0.61	0.72	0.60	0.76	0.70
Japan	0.83	50	0.69	0.70	0.83	0.71	0.49	0.76	0.73
Kazakhstan	0.84	52	0.71	0.75	0.79	0.74	0.60	0.75	0.68
Korea, Rep. of	0.85	53	0.81	0.74	0.76	0.71	0.56	0.75	0.73
Kosovo	0.73	39	0.53	0.75	0.74	0.50	0.52	0.73	0.52
Kuwait	0.73	39	0.19	0.80	0.77	0.41	0.37	0.80	0.73
Latvia	0.85	53	0.76	0.72	0.81	0.72	0.61	0.77	0.69
Lithuania	0.83	50	0.77	0.75	0.70	0.72	0.65	0.78	0.56
Malta	0.84	52	0.70	0.73	0.79	0.72	0.55	0.79	0.73
Montenegro	0.78	43	0.45	0.76	0.77	0.49	0.43	0.79	0.77
Morocco	0.73	38	0.39	0.72	0.75	0.46	0.37	0.76	0.70
Netherlands	0.82	49	0.65	0.72	0.77	0.68	0.45	0.76	0.79
New Zealand	0.75	40	0.55	0.70	0.76	0.56	0.31	0.75	0.70
North Macedonia	0.78	44	0.56	0.74	0.70	0.58	0.55	0.76	0.69
Northern Ireland	0.84	52	0.74	0.74	0.82	0.74	0.42	0.77	0.74

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Confident in Science* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item						
			ASBS09A	ASBS09B R	ASBS09C R	ASBS09D	ASBS09E	ASBS09F R	ASBS09G R
Norway (5)	0.80	47	0.70	0.74	0.79	0.70	0.40	0.73	0.70
Oman	0.72	38	0.19	0.78	0.69	0.41	0.35	0.82	0.75
Pakistan	0.70	36	0.33	0.73	0.75	0.33	0.30	0.73	0.75
Philippines	0.48	31	-0.33	0.70	0.66	-0.19	-0.37	0.73	0.64
Poland	0.84	51	0.74	0.76	0.69	0.74	0.54	0.79	0.73
Portugal	0.75	41	0.60	0.66	0.71	0.60	0.48	0.64	0.73
Qatar	0.77	42	0.28	0.81	0.79	0.41	0.35	0.83	0.78
Russian Federation	0.86	55	0.71	0.78	0.81	0.76	0.61	0.80	0.72
Saudi Arabia	0.73	40	0.14	0.78	0.79	0.40	0.36	0.81	0.77
Serbia	0.84	52	0.63	0.74	0.80	0.68	0.60	0.78	0.77
Singapore	0.85	53	0.74	0.77	0.80	0.67	0.51	0.80	0.75
Slovak Republic	0.84	51	0.73	0.70	0.80	0.69	0.56	0.75	0.74
South Africa (5)	0.66	34	0.21	0.77	0.74	0.29	0.14	0.79	0.69
Spain	0.80	46	0.62	0.75	0.72	0.61	0.48	0.74	0.77
Sweden	0.85	53	0.74	0.74	0.80	0.76	0.50	0.79	0.73
Turkey (5)	0.81	47	0.63	0.71	0.78	0.60	0.51	0.77	0.74
United Arab Emirates	0.78	43	0.46	0.77	0.76	0.54	0.46	0.79	0.73
United States	0.82	49	0.66	0.76	0.77	0.66	0.39	0.80	0.76
Benchmarking Participants									
Ontario, Canada	0.84	52	0.73	0.78	0.79	0.72	0.43	0.76	0.76
Quebec, Canada	0.83	51	0.75	0.73	0.65	0.78	0.51	0.77	0.78
Moscow City, Russian Fed.	0.87	58	0.73	0.79	0.83	0.78	0.60	0.81	0.74
Madrid, Spain	0.82	49	0.68	0.73	0.73	0.65	0.52	0.75	0.79
Abu Dhabi, UAE	0.77	42	0.39	0.78	0.77	0.48	0.40	0.80	0.73
Dubai, UAE	0.79	45	0.57	0.73	0.74	0.64	0.53	0.75	0.70

R Reverse coded

Relationship Between the TIMSS 2019 *Students Confident in Science* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Albania	0.32	0.10	0.09
Armenia	0.27	0.07	0.07
Australia	0.19	0.04	0.04
Austria	0.31	0.10	0.11
Azerbaijan	0.28	0.08	0.07
Bahrain	0.36	0.13	0.13
Belgium (Flemish)	0.27	0.07	0.08
Bosnia and Herzegovina	0.27	0.08	0.08
Bulgaria	0.38	0.14	0.15
Canada	0.19	0.04	0.04
Chile	0.31	0.09	0.11
Chinese Taipei	0.31	0.10	0.10
Croatia	0.26	0.07	0.06
Cyprus	0.26	0.07	0.07
Czech Republic	0.25	0.06	0.07
Denmark	0.26	0.07	0.07
England	0.21	0.04	0.05
Finland	0.20	0.04	0.05
France	0.30	0.09	0.09
Georgia	0.20	0.04	0.05
Germany	0.35	0.12	0.13
Hong Kong SAR	0.27	0.07	0.08
Hungary	0.31	0.10	0.10
Iran, Islamic Rep. of	0.36	0.13	0.14
Ireland	0.17	0.03	0.03
Italy	0.17	0.03	0.04
Japan	0.20	0.04	0.04
Kazakhstan	0.20	0.04	0.03
Korea, Rep. of	0.28	0.08	0.07
Kosovo	0.33	0.11	0.12
Kuwait	0.30	0.09	0.09
Latvia	0.20	0.04	0.04
Lithuania	0.24	0.06	0.06
Malta	0.27	0.07	0.09
Montenegro	0.34	0.11	0.12
Morocco	0.33	0.11	0.12
Netherlands	0.27	0.08	0.09
New Zealand	0.22	0.05	0.05

Relationship Between the TIMSS 2019 *Students Confident in Science* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
North Macedonia	0.42	0.18	0.18
Northern Ireland	0.23	0.05	0.06
Norway (5)	0.22	0.05	0.05
Oman	0.38	0.14	0.15
Pakistan	0.21	0.04	0.04
Philippines	0.39	0.16	0.14
Poland	0.23	0.05	0.06
Portugal	0.28	0.08	0.09
Qatar	0.33	0.11	0.12
Russian Federation	0.19	0.03	0.04
Saudi Arabia	0.35	0.12	0.13
Serbia	0.25	0.06	0.07
Singapore	0.26	0.07	0.08
Slovak Republic	0.26	0.07	0.07
South Africa (5)	0.41	0.17	0.17
Spain	0.27	0.07	0.08
Sweden	0.19	0.03	0.04
Turkey (5)	0.33	0.11	0.12
United Arab Emirates	0.27	0.07	0.08
United States	0.25	0.06	0.08
International Median	**0.27**	**0.07**	**0.08**
Benchmarking Participants			
Ontario, Canada	0.18	0.03	0.04
Quebec, Canada	0.21	0.04	0.05
Moscow City, Russian Fed.	0.19	0.03	0.04
Madrid, Spain	0.28	0.08	0.10
Abu Dhabi, UAE	0.30	0.09	0.10
Dubai, UAE	0.17	0.03	0.03

Students Like Learning Mathematics – Grade 4

About the Scale

The *Students Like Learning Mathematics* scale was created based on students' responses to nine items listed below.

How much do you agree with these statements about learning mathematics?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
ASBM02A [T]	1) I enjoy learning mathematics	◯	◯	◯	◯
ASBM02B [T]	2) I wish I did not have to study mathematics [R]	◯	◯	◯	◯
ASBM02C [T]	3) Mathematics is boring [R]	◯	◯	◯	◯
ASBM02D [T]	4) I learn many interesting things in mathematics	◯	◯	◯	◯
ASBM02E [T]	5) I like mathematics	◯	◯	◯	◯
ASBM02F [T]	6) I like any schoolwork that involves numbers	◯	◯	◯	◯
ASBM02G [T]	7) I like to solve mathematics problems	◯	◯	◯	◯
ASBM02H [T]	8) I look forward to mathematics lessons	◯	◯	◯	◯
ASBM02I [T]	9) Mathematics is one of my favorite subjects	◯	◯	◯	◯

Very Much Like		Somewhat Like		Do Not Like

Scale Cut Scores 10.2 8.4

[R] Reverse coded

[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Students Like Learning Mathematics* Scale – Grade 4

Item	delta	tau_1	tau_2	tau_3	Infit
ASBM02A	-0.36071	-0.72190	-0.46691	1.18881	0.78
ASBM02B [R]	0.25692	-0.42734	0.08319	0.34415	1.77
ASBM02C [R]	0.13685	-0.71463	0.14695	0.56768	1.33
ASBM02D	-0.64160	-0.71305	-0.29401	1.00706	1.14
ASBM02E	-0.20358	-0.49481	-0.33283	0.82764	0.64
ASBM02F	0.01126	-1.11828	-0.10444	1.22272	1.01
ASBM02G	0.09150	-0.86159	-0.20061	1.06220	1.00
ASBM02H	0.34970	-0.96071	-0.17174	1.13245	0.83
ASBM02I	0.35966	-0.48032	-0.05428	0.53460	0.88

[R] Reverse coded

Scale Transformation Constants for the TIMSS 2019 *Students Like Learning Mathematics* Scale – Grade 4

Scale Transformation Constants	
A = 8.402636	Transformed Scale Score = 8.402636 + 1.234546 • Logit Scale Score
B = 1.234546	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Students Like Learning Mathematics* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	3.85307	
1	5.17652	
2	5.79688	
3	6.21498	
4	6.54051	
5	6.81155	
6	7.04782	
7	7.26046	
8	7.45571	
9	7.64004	
10	7.81505	
11	7.98325	
12	8.14669	
13	8.30721	8.4
14	8.46650	
15	8.62627	
16	8.78843	
17	8.95513	
18	9.12899	
19	9.31319	
20	9.51007	
21	9.72723	
22	9.97246	
23	10.25879	10.2
24	10.60755	
25	11.06393	
26	11.73885	
27	13.14434	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Like Learning Mathematics* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			ASBM02A	ASBM02B R	ASBM02C R	ASBM02D	ASBM02E	ASBM02F	ASBM02G	ASBM02H	ASBM02I
Albania	0.81	46	0.78	0.33	0.55	0.53	0.82	0.69	0.75	0.76	0.77
Armenia	0.86	52	0.71	0.39	0.39	0.74	0.86	0.75	0.79	0.82	0.84
Australia	0.94	68	0.89	0.73	0.80	0.70	0.91	0.78	0.84	0.90	0.86
Austria	0.94	68	0.89	0.70	0.74	0.73	0.91	0.79	0.86	0.90	0.86
Azerbaijan	0.78	50	0.79	0.03	0.38	0.72	0.85	0.70	0.79	0.80	0.84
Bahrain	0.89	56	0.78	0.52	0.57	0.75	0.88	0.70	0.81	0.81	0.82
Belgium (Flemish)	0.93	64	0.88	0.76	0.81	0.64	0.92	0.76	0.69	0.84	0.84
Bosnia and Herzegovina	0.91	60	0.83	0.55	0.67	0.68	0.90	0.79	0.77	0.89	0.84
Bulgaria	0.93	65	0.85	0.56	0.73	0.75	0.90	0.83	0.88	0.84	0.84
Canada	0.93	64	0.87	0.65	0.77	0.69	0.90	0.76	0.81	0.87	0.84
Chile	0.90	57	0.81	0.51	0.65	0.66	0.89	0.70	0.82	0.85	0.83
Chinese Taipei	0.94	67	0.87	0.70	0.69	0.76	0.91	0.83	0.85	0.87	0.85
Croatia	0.94	67	0.89	0.68	0.82	0.78	0.93	0.84	0.60	0.92	0.86
Cyprus	0.92	62	0.85	0.65	0.70	0.73	0.89	0.76	0.79	0.86	0.84
Czech Republic	0.93	65	0.89	0.63	0.72	0.73	0.92	0.79	0.73	0.91	0.88
Denmark	0.93	65	0.86	0.69	0.79	0.79	0.89	0.76	0.75	0.88	0.84
England	0.93	66	0.88	0.71	0.78	0.69	0.91	0.76	0.82	0.89	0.85
Finland	0.93	65	0.86	0.67	0.82	0.74	0.91	0.80	0.70	0.87	0.85
France	0.91	60	0.86	0.67	0.74	0.71	0.89	0.71	0.69	0.82	0.84
Georgia	0.86	50	0.78	0.47	0.53	0.61	0.85	0.70	0.74	0.82	0.78
Germany	0.94	67	0.87	0.71	0.74	0.74	0.91	0.81	0.78	0.90	0.87
Hong Kong SAR	0.93	66	0.86	0.56	0.68	0.79	0.91	0.85	0.86	0.87	0.87
Hungary	0.92	62	0.88	0.62	0.70	0.67	0.91	0.79	0.74	0.87	0.85
Iran, Islamic Rep. of	0.87	52	0.80	0.53	0.53	0.68	0.87	0.59	0.82	0.76	0.81
Ireland	0.93	65	0.87	0.68	0.77	0.71	0.91	0.74	0.81	0.88	0.83
Italy	0.93	64	0.86	0.72	0.80	0.67	0.89	0.78	0.71	0.87	0.85
Japan	0.93	66	0.88	0.60	0.71	0.75	0.91	0.81	0.86	0.88	0.84
Kazakhstan	0.88	54	0.76	0.39	0.69	0.65	0.85	0.73	0.78	0.83	0.82
Korea, Rep. of	0.95	71	0.88	0.69	0.81	0.75	0.92	0.87	0.91	0.84	0.87
Kosovo	0.63	37	0.70	0.08	0.26	0.56	0.85	0.75	0.51	0.63	0.72
Kuwait	0.87	53	0.78	0.26	0.37	0.75	0.88	0.76	0.83	0.83	0.82
Latvia	0.93	64	0.87	0.65	0.78	0.70	0.89	0.77	0.83	0.86	0.84
Lithuania	0.92	61	0.86	0.71	0.73	0.59	0.89	0.78	0.85	0.79	0.81
Malta	0.91	59	0.85	0.61	0.75	0.66	0.88	0.70	0.71	0.85	0.84
Montenegro	0.90	57	0.83	0.53	0.66	0.68	0.88	0.76	0.72	0.88	0.79
Morocco	0.78	42	0.58	0.32	0.37	0.67	0.80	0.69	0.72	0.78	0.69
Netherlands	0.92	63	0.83	0.68	0.77	0.64	0.90	0.76	0.85	0.84	0.82
New Zealand	0.94	66	0.87	0.71	0.76	0.71	0.91	0.77	0.84	0.88	0.85
North Macedonia	0.81	46	0.80	0.35	0.43	0.59	0.86	0.66	0.63	0.77	0.80

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Like Learning Mathematics* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			ASBM02A	ASBM02B R	ASBM02C R	ASBM02D	ASBM02E	ASBM02F	ASBM02G	ASBM02H	ASBM02I
Northern Ireland	0.93	64	0.86	0.66	0.77	0.66	0.90	0.79	0.80	0.88	0.83
Norway (5)	0.94	67	0.87	0.66	0.79	0.75	0.90	0.79	0.86	0.87	0.84
Oman	0.80	44	0.72	0.27	0.40	0.73	0.84	0.66	0.78	0.59	0.77
Pakistan	0.75	42	0.65	-0.03	-0.05	0.73	0.80	0.68	0.76	0.74	0.76
Philippines	0.76	36	0.69	0.31	0.40	0.65	0.74	0.63	0.66	0.46	0.72
Poland	0.93	64	0.87	0.67	0.75	0.76	0.90	0.77	0.86	0.72	0.85
Portugal	0.92	62	0.86	0.66	0.71	0.68	0.89	0.76	0.80	0.85	0.81
Qatar	0.90	56	0.78	0.49	0.56	0.73	0.88	0.75	0.82	0.83	0.83
Russian Federation	0.90	58	0.84	0.53	0.72	0.66	0.87	0.76	0.73	0.84	0.83
Saudi Arabia	0.86	52	0.72	0.36	0.49	0.75	0.86	0.72	0.82	0.83	0.78
Serbia	0.93	66	0.85	0.66	0.79	0.76	0.91	0.80	0.78	0.90	0.84
Singapore	0.93	65	0.86	0.74	0.77	0.72	0.90	0.73	0.81	0.85	0.86
Slovak Republic	0.93	63	0.88	0.65	0.78	0.56	0.90	0.74	0.82	0.90	0.85
South Africa (5)	0.80	42	0.72	0.29	0.35	0.62	0.79	0.71	0.69	0.70	0.73
Spain	0.90	58	0.81	0.51	0.72	0.66	0.88	0.73	0.78	0.84	0.84
Sweden	0.94	70	0.87	0.71	0.85	0.77	0.93	0.84	0.82	0.88	0.84
Turkey (5)	0.88	54	0.77	0.61	0.71	0.47	0.87	0.65	0.80	0.79	0.83
United Arab Emirates	0.89	56	0.80	0.51	0.57	0.71	0.87	0.73	0.79	0.82	0.83
United States	0.93	65	0.87	0.64	0.74	0.71	0.91	0.77	0.87	0.88	0.85
Benchmarking Participants											
Ontario, Canada	0.93	64	0.88	0.65	0.77	0.68	0.91	0.74	0.83	0.86	0.85
Quebec, Canada	0.92	62	0.85	0.62	0.76	0.71	0.89	0.76	0.76	0.87	0.82
Moscow City, Russian Fed.	0.92	61	0.87	0.53	0.75	0.70	0.89	0.75	0.77	0.85	0.86
Madrid, Spain	0.90	56	0.81	0.54	0.73	0.65	0.87	0.71	0.74	0.83	0.82
Abu Dhabi, UAE	0.89	55	0.80	0.47	0.53	0.73	0.87	0.73	0.79	0.82	0.83
Dubai, UAE	0.91	59	0.81	0.59	0.71	0.64	0.87	0.73	0.82	0.83	0.83

R Reverse coded

Relationship Between the TIMSS 2019 *Students Like Learning Mathematics* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Albania	0.30	0.09	0.06
Armenia	0.25	0.06	0.05
Australia	0.24	0.06	0.05
Austria	0.19	0.04	0.03
Azerbaijan	0.30	0.09	0.08
Bahrain	0.19	0.03	0.03
Belgium (Flemish)	0.18	0.03	0.03
Bosnia and Herzegovina	0.16	0.03	0.03
Bulgaria	0.15	0.02	0.02
Canada	0.23	0.05	0.05
Chile	0.21	0.04	0.04
Chinese Taipei	0.26	0.07	0.06
Croatia	0.21	0.04	0.05
Cyprus	0.24	0.06	0.06
Czech Republic	0.21	0.05	0.04
Denmark	0.20	0.04	0.03
England	0.22	0.05	0.05
Finland	0.16	0.03	0.02
France	0.21	0.05	0.04
Georgia	0.18	0.03	0.03
Germany	0.23	0.05	0.05
Hong Kong SAR	0.23	0.05	0.06
Hungary	0.21	0.04	0.04
Iran, Islamic Rep. of	0.17	0.03	0.03
Ireland	0.20	0.04	0.04
Italy	0.13	0.02	0.02
Japan	0.31	0.09	0.08
Kazakhstan	0.15	0.02	0.01
Korea, Rep. of	0.31	0.10	0.09
Kosovo	0.36	0.13	0.10
Kuwait	0.24	0.06	0.06
Latvia	0.26	0.07	0.06
Lithuania	0.15	0.02	0.02
Malta	0.19	0.04	0.03
Montenegro	0.21	0.04	0.04
Morocco	0.29	0.08	0.08
Netherlands	0.18	0.03	0.03
New Zealand	0.15	0.02	0.02

Relationship Between the TIMSS 2019 *Students Like Learning Mathematics* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
North Macedonia	0.29	0.08	0.08
Northern Ireland	0.25	0.06	0.06
Norway (5)	0.12	0.02	0.02
Oman	0.28	0.08	0.07
Pakistan	0.12	0.01	0.03
Philippines	0.41	0.16	0.17
Poland	0.20	0.04	0.04
Portugal	0.25	0.06	0.05
Qatar	0.18	0.03	0.05
Russian Federation	0.18	0.03	0.03
Saudi Arabia	0.26	0.07	0.07
Serbia	0.17	0.03	0.03
Singapore	0.30	0.09	0.09
Slovak Republic	0.09	0.01	0.01
South Africa (5)	0.34	0.12	0.13
Spain	0.19	0.04	0.04
Sweden	0.11	0.01	0.01
Turkey (5)	0.23	0.05	0.05
United Arab Emirates	0.21	0.04	0.05
United States	0.19	0.04	0.05
International Median	**0.21**	**0.04**	**0.04**
Benchmarking Participants			
Ontario, Canada	0.21	0.04	0.05
Quebec, Canada	0.26	0.07	0.06
Moscow City, Russian Fed.	0.25	0.06	0.06
Madrid, Spain	0.17	0.03	0.03
Abu Dhabi, UAE	0.16	0.03	0.03
Dubai, UAE	0.12	0.02	0.02

Students Like Learning Science – Grade 4

About the Scale

The *Students Like Learning Science* scale was created based on students' responses to nine items listed below.

How much do you agree with these statements about learning science?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
ASBS07A T	1) I enjoy learning science	○	○	○	○
ASBS07B T	2) I wish I did not have to study science R	○	○	○	○
ASBS07C T	3) Science is boring R	○	○	○	○
ASBS07D T	4) I learn many interesting things in science	○	○	○	○
ASBS07E T	5) I like science	○	○	○	○
ASBS07F T	6) I look forward to learning science in school	○	○	○	○
ASBS07G T	7) Science teaches me how things in the world work	○	○	○	○
ASBS07H T	8) I like to do science experiments	○	○	○	○
ASBS07I T	9) Science is one of my favorite subjects	○	○	○	○

	Very Much Like		Somewhat Like		Do Not Like
Scale Cut Scores		9.7		7.6	

R Reverse coded

T Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Students Like Learning Science* Scale – Grade 4

Item	delta	tau_1	tau_2	tau_3	Infit
ASBS07A	-0.12179	-0.53434	-0.49509	1.02943	0.81
ASBS07B R	0.51587	-0.34763	-0.09494	0.44257	1.55
ASBS07C R	0.30062	-0.44642	-0.04532	0.49174	1.25
ASBS07D	-0.46068	-0.25588	-0.41313	0.66901	0.94
ASBS07E	-0.06693	-0.45384	-0.32221	0.77605	0.66
ASBS07F	0.26162	-0.76615	-0.14592	0.91207	0.82
ASBS07G	-0.54440	-0.39669	-0.49739	0.89408	1.13
ASBS07H	-0.38772	-0.17842	-0.29340	0.47182	1.26
ASBS07I	0.50341	-0.65684	-0.14297	0.79981	0.90

R Reverse coded

Scale Transformation Constants for the TIMSS 2019 *Students Like Learning Science* Scale – Grade 4

Scale Transformation Constants	
A = 7.692952	Transformed Scale Score = 7.692952 + 1.486733 • Logit Scale Score
B = 1.486733	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Students Like Learning Science* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	2.69347	
1	4.16658	
2	4.83600	
3	5.28195	
4	5.62666	
5	5.91275	
6	6.16040	
7	6.38769	
8	6.59779	
9	6.79572	
10	6.98534	
11	7.16962	
12	7.35091	
13	7.53117	7.6
14	7.71287	
15	7.89534	
16	8.08274	
17	8.27639	
18	8.47890	
19	8.69346	
20	8.92203	
21	9.17441	
22	9.45888	
23	9.79069	9.7
24	10.19505	
25	10.72673	
26	11.51982	
27	13.19236	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Like Learning Science* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			ASBS07A	ASBS07B R	ASBS07C R	ASBS07D	ASBS07E	ASBS07F	ASBS07G	ASBS07H	ASBS07I
Albania	0.75	41	0.76	0.33	0.54	0.60	0.84	0.76	0.52	0.54	0.70
Armenia	0.86	50	0.74	0.39	0.37	0.76	0.87	0.83	0.72	0.66	0.84
Australia	0.93	64	0.88	0.72	0.78	0.78	0.91	0.90	0.67	0.68	0.84
Austria	0.88	54	0.83	0.64	0.73	0.74	0.91	0.90	0.55	0.28	0.84
Azerbaijan	0.76	45	0.76	0.04	0.36	0.75	0.85	0.80	0.66	0.59	0.80
Bahrain	0.86	51	0.74	0.51	0.52	0.79	0.86	0.77	0.71	0.66	0.81
Belgium (Flemish)	0.91	58	0.86	0.76	0.82	0.70	0.91	0.84	0.54	0.43	0.84
Bosnia and Herzegovina	0.87	52	0.82	0.48	0.61	0.72	0.89	0.87	0.55	0.62	0.81
Bulgaria	0.87	52	0.82	0.45	0.63	0.76	0.86	0.79	0.67	0.62	0.81
Canada	0.91	59	0.86	0.64	0.76	0.75	0.91	0.89	0.60	0.60	0.85
Chile	0.85	49	0.75	0.43	0.56	0.73	0.86	0.81	0.66	0.60	0.78
Chinese Taipei	0.92	62	0.86	0.61	0.58	0.83	0.91	0.88	0.74	0.74	0.85
Croatia	0.89	56	0.85	0.61	0.73	0.76	0.91	0.89	0.54	0.44	0.83
Cyprus	0.92	61	0.87	0.68	0.74	0.78	0.91	0.87	0.68	0.60	0.86
Czech Republic	0.91	59	0.86	0.62	0.71	0.74	0.91	0.90	0.65	0.57	0.88
Denmark	0.92	61	0.84	0.71	0.79	0.83	0.90	0.89	0.63	0.52	0.82
England	0.92	63	0.87	0.73	0.78	0.75	0.91	0.90	0.66	0.62	0.84
Finland	0.92	63	0.85	0.70	0.78	0.79	0.91	0.87	0.67	0.68	0.84
France	0.92	61	0.87	0.73	0.77	0.76	0.90	0.85	0.63	0.65	0.83
Georgia	0.85	48	0.78	0.54	0.56	0.63	0.87	0.82	0.55	0.62	0.75
Germany	0.92	61	0.83	0.75	0.74	0.78	0.89	0.88	0.65	0.62	0.84
Hong Kong SAR	0.91	60	0.83	0.42	0.48	0.83	0.92	0.89	0.79	0.78	0.86
Hungary	0.89	55	0.85	0.62	0.70	0.73	0.90	0.87	0.58	0.47	0.82
Iran, Islamic Rep. of	0.85	48	0.76	0.50	0.52	0.69	0.84	0.74	0.66	0.68	0.81
Ireland	0.91	60	0.84	0.66	0.74	0.74	0.90	0.89	0.63	0.67	0.82
Italy	0.89	53	0.83	0.69	0.73	0.71	0.87	0.85	0.46	0.47	0.83
Japan	0.92	61	0.86	0.60	0.68	0.77	0.90	0.90	0.69	0.74	0.83
Kazakhstan	0.88	52	0.78	0.47	0.65	0.73	0.84	0.82	0.70	0.66	0.80
Korea, Rep. of	0.93	64	0.87	0.71	0.79	0.78	0.90	0.85	0.70	0.72	0.86
Kosovo	0.71	41	0.77	0.11	0.37	0.61	0.86	0.73	0.66	0.59	0.72
Kuwait	0.82	47	0.67	0.27	0.37	0.75	0.83	0.81	0.74	0.73	0.78
Latvia	0.88	52	0.84	0.64	0.75	0.76	0.87	0.65	0.57	0.52	0.82
Lithuania	0.88	54	0.85	0.71	0.71	0.67	0.88	0.79	0.61	0.50	0.80
Malta	0.91	59	0.85	0.60	0.74	0.75	0.90	0.87	0.63	0.65	0.84
Montenegro	0.82	47	0.76	0.34	0.48	0.71	0.86	0.84	0.62	0.67	0.71
Morocco	0.80	43	0.64	0.40	0.42	0.71	0.80	0.76	0.68	0.72	0.68
Netherlands	0.90	56	0.84	0.57	0.72	0.73	0.91	0.84	0.62	0.65	0.81
New Zealand	0.92	61	0.86	0.53	0.63	0.79	0.90	0.89	0.72	0.80	0.83
North Macedonia	0.81	46	0.79	0.27	0.29	0.75	0.86	0.78	0.69	0.63	0.77

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Like Learning Science* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			ASBS07A	ASBS07B R	ASBS07C R	ASBS07D	ASBS07E	ASBS07F	ASBS07G	ASBS07H	ASBS07I
Northern Ireland	0.91	58	0.84	0.65	0.71	0.74	0.89	0.87	0.63	0.69	0.80
Norway (5)	0.91	60	0.87	0.71	0.79	0.78	0.91	0.89	0.55	0.56	0.82
Oman	0.79	43	0.66	0.36	0.45	0.68	0.80	0.68	0.71	0.69	0.77
Pakistan	0.78	41	0.58	0.22	0.20	0.75	0.76	0.68	0.74	0.75	0.77
Philippines	0.76	37	0.69	0.29	0.36	0.68	0.74	0.54	0.64	0.67	0.68
Poland	0.90	56	0.85	0.61	0.64	0.82	0.88	0.71	0.70	0.67	0.83
Portugal	0.88	53	0.82	0.55	0.57	0.73	0.87	0.82	0.67	0.69	0.76
Qatar	0.86	51	0.73	0.42	0.47	0.75	0.86	0.81	0.73	0.70	0.81
Russian Federation	0.88	54	0.82	0.57	0.70	0.73	0.87	0.83	0.64	0.56	0.82
Saudi Arabia	0.85	50	0.70	0.40	0.48	0.79	0.85	0.81	0.71	0.73	0.79
Serbia	0.90	56	0.84	0.55	0.69	0.75	0.89	0.86	0.63	0.63	0.82
Singapore	0.91	58	0.85	0.70	0.73	0.76	0.89	0.86	0.64	0.55	0.83
Slovak Republic	0.90	56	0.86	0.66	0.76	0.61	0.89	0.88	0.60	0.57	0.83
South Africa (5)	0.79	41	0.69	0.29	0.35	0.66	0.78	0.71	0.68	0.69	0.70
Spain	0.85	49	0.81	0.43	0.63	0.73	0.88	0.84	0.57	0.42	0.80
Sweden	0.93	64	0.86	0.73	0.82	0.80	0.91	0.89	0.66	0.62	0.83
Turkey (5)	0.86	48	0.75	0.55	0.65	0.60	0.84	0.76	0.60	0.63	0.80
United Arab Emirates	0.88	54	0.77	0.51	0.56	0.76	0.88	0.82	0.72	0.69	0.82
United States	0.91	59	0.86	0.62	0.69	0.76	0.91	0.88	0.64	0.64	0.84
Benchmarking Participants											
Ontario, Canada	0.91	60	0.85	0.66	0.77	0.75	0.90	0.89	0.64	0.57	0.84
Quebec, Canada	0.91	59	0.87	0.60	0.75	0.77	0.91	0.90	0.51	0.63	0.87
Moscow City, Russian Fed.	0.89	56	0.85	0.61	0.74	0.74	0.89	0.84	0.60	0.48	0.85
Madrid, Spain	0.86	49	0.83	0.47	0.64	0.72	0.89	0.84	0.55	0.36	0.81
Abu Dhabi, UAE	0.88	54	0.77	0.46	0.50	0.79	0.88	0.84	0.71	0.71	0.83
Dubai, UAE	0.88	53	0.79	0.57	0.69	0.68	0.88	0.81	0.68	0.61	0.81

R Reverse coded

Relationship Between the TIMSS 2019 *Students Like Learning Science* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Albania	0.20	0.04	0.03
Armenia	0.22	0.05	0.05
Australia	0.10	0.01	0.01
Austria	0.17	0.03	0.02
Azerbaijan	0.22	0.05	0.04
Bahrain	0.33	0.11	0.10
Belgium (Flemish)	0.12	0.02	0.01
Bosnia and Herzegovina	0.08	0.01	0.01
Bulgaria	0.14	0.02	0.02
Canada	0.08	0.01	0.01
Chile	0.20	0.04	0.03
Chinese Taipei	0.17	0.03	0.02
Croatia	0.11	0.01	0.01
Cyprus	0.18	0.03	0.03
Czech Republic	0.09	0.01	0.01
Denmark	0.13	0.02	0.02
England	0.07	0.01	0.00
Finland	0.01	0.00	0.00
France	0.14	0.02	0.02
Georgia	0.08	0.01	0.01
Germany	0.18	0.03	0.03
Hong Kong SAR	0.23	0.05	0.05
Hungary	0.09	0.01	0.01
Iran, Islamic Rep. of	0.26	0.07	0.07
Ireland	0.15	0.02	0.02
Italy	0.09	0.01	0.01
Japan	0.13	0.02	0.01
Kazakhstan	0.14	0.02	0.02
Korea, Rep. of	0.20	0.04	0.03
Kosovo	0.28	0.08	0.07
Kuwait	0.33	0.11	0.10
Latvia	0.05	0.00	0.00
Lithuania	0.07	0.01	0.00
Malta	0.21	0.04	0.04
Montenegro	0.13	0.02	0.03
Morocco	0.35	0.12	0.10
Netherlands	0.16	0.03	0.02
New Zealand	0.15	0.02	0.02

Relationship Between the TIMSS 2019 *Students Like Learning Science* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
North Macedonia	0.31	0.10	0.09
Northern Ireland	0.21	0.04	0.04
Norway (5)	0.09	0.01	0.01
Oman	0.34	0.11	0.11
Pakistan	0.19	0.04	0.05
Philippines	0.50	0.25	0.23
Poland	0.08	0.01	0.01
Portugal	0.20	0.04	0.05
Qatar	0.34	0.12	0.13
Russian Federation	0.04	0.00	0.00
Saudi Arabia	0.33	0.11	0.12
Serbia	0.03	0.00	0.01
Singapore	0.15	0.02	0.02
Slovak Republic	0.06	0.00	0.00
South Africa (5)	0.42	0.18	0.17
Spain	0.10	0.01	0.01
Sweden	0.05	0.00	0.01
Turkey (5)	0.29	0.09	0.08
United Arab Emirates	0.32	0.10	0.09
United States	0.13	0.02	0.02
International Median	**0.15**	**0.02**	**0.02**
Benchmarking Participants			
Ontario, Canada	0.07	0.00	0.01
Quebec, Canada	0.10	0.01	0.01
Moscow City, Russian Fed.	0.03	0.00	0.00
Madrid, Spain	0.09	0.01	0.01
Abu Dhabi, UAE	0.32	0.10	0.10
Dubai, UAE	0.15	0.02	0.02

Scales Based on Principals' Reports

Instruction Affected by Mathematics Resource Shortages – Grade 4

About the Scale

The *Instruction Affected by Mathematics Resource Shortages* scale was created based on principals' responses to thirteen items listed below.

How much is your school's capacity to provide instruction affected by a shortage or inadequacy of the following?

		Not at all	A little	Some	A lot
	A. General School Resources				
ACBG13AA T	1) Instructional materials (e.g., textbooks)	○	○	○	○
ACBG13AB T	2) Supplies (e.g., papers, pencils, materials)	○	○	○	○
ACBG13AC T	3) School buildings and grounds	○	○	○	○
ACBG13AD T	4) Heating/cooling and lighting systems	○	○	○	○
ACBG13AE T	5) Instructional space (e.g., classrooms)	○	○	○	○
ACBG13AF T	6) Technologically competent staff	○	○	○	○
ACBG13AG T	7) Audio-visual resources for delivery of instruction (e.g., interactive white boards, digital projectors)	○	○	○	○
ACBG13AH T	8) Computer technology for teaching and learning (e.g., computers or tablets for student use)	○	○	○	○
	B. Resources for Mathematics Instruction				
ACBG13BA T	1) Teachers with a specialization in mathematics	○	○	○	○
ACBG13BB T	2) Computer software/applications for mathematics instruction	○	○	○	○
ACBG13BC T	3) Library resources relevant to mathematics instruction	○	○	○	○
ACBG13BD T	4) Calculators for mathematics instruction	○	○	○	○
ACBG13BE T	5) Concrete objects or materials to help students understand quantities or procedures	○	○	○	○

Not Affected		Somewhat Affected		Affected A Lot
Scale Cut Scores	11.3		6.7	

T Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Instruction Affected by Mathematics Resource Shortages* Scale – Grade 4

Item	delta	tau_1	tau_2	tau_3	Infit
ACBG13AA	-0.14948	-0.24821	0.06659	0.18162	0.94
ACBG13AB	-0.38241	-0.22495	0.19483	0.03012	0.86
ACBG13AC	0.12747	-0.72711	0.13093	0.59618	1.11
ACBG13AD	-0.13054	-0.44482	0.17299	0.27183	0.97
ACBG13AE	0.17053	-0.44105	-0.02256	0.46361	1.12
ACBG13AF	0.18053	-1.21486	-0.06356	1.27842	1.00
ACBG13AG	0.09526	-0.75238	0.02378	0.72860	0.95
ACBG13AH	0.32224	-0.95290	0.03310	0.91980	1.10
ACBG13BA	0.00799	-0.52502	0.00573	0.51929	1.10
ACBG13BB	0.16554	-1.32115	0.08121	1.23994	1.07
ACBG13BC	0.07814	-1.30714	0.04775	1.25939	1.15
ACBG13BD	-0.52163	-0.74124	0.19132	0.54992	1.35
ACBG13BE	0.03636	-1.10053	-0.05826	1.15879	0.86

Scale Transformation Constants for the TIMSS 2019 *Instruction Affected by Mathematics Resource Shortages* Scale – Grade 4

Scale Transformation Constants	
A = 8.983616	Transformed Scale Score = 8.983616 + 1.642771 • Logit Scale Score
B = 1.642771	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Instruction Affected by Mathematics Resource Shortages* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	2.26751	
1	4.10533	
2	4.97458	
3	5.55459	
4	5.99316	
5	6.34731	
6	6.64531	6.7
7	6.90436	
8	7.13441	
9	7.34236	
10	7.53320	
11	7.70956	
12	7.87690	
13	8.03580	
14	8.18794	
15	8.33477	
16	8.47754	
17	8.61741	
18	8.75540	
19	8.89246	
20	9.03124	
21	9.16743	
22	9.30709	
23	9.44946	
24	9.59555	
25	9.74642	
26	9.90334	
27	10.06763	
28	10.23978	
29	10.42410	
30	10.62249	
31	10.83857	
32	11.07716	
33	11.34502	11.3
34	11.65202	
35	12.01510	
36	12.46230	
37	13.05030	
38	13.92655	
39	15.77025	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instruction Affected by Mathematics Resource Shortages* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item												
			ACBG13AA	ACBG13AB	ACBG13AC	ACBG13AD	ACBG13AE	ACBG13AF	ACBG13AG	ACBG13AH	ACBG13BA	ACBG13BB	ACBG13BC	ACBG13BD	ACBG13BE
Albania	0.90	46	0.73	0.81	0.81	0.77	0.75	0.73	0.62	0.63	0.63	0.46	0.48	0.47	0.78
Armenia	0.88	42	0.65	0.53	0.61	0.68	0.54	0.68	0.64	0.64	0.69	0.68	0.71	0.63	0.71
Australia	0.92	51	0.66	0.72	0.71	0.67	0.74	0.68	0.78	0.80	0.56	0.74	0.72	0.72	0.75
Austria	0.82	33	0.54	0.57	0.60	0.50	0.61	0.61	0.58	0.60	0.63	0.61	0.61	0.50	0.53
Azerbaijan	0.86	38	0.53	0.58	0.60	0.66	0.61	0.65	0.52	0.61	0.61	0.60	0.71	0.59	0.65
Bahrain	0.97	74	0.92	0.91	0.88	0.90	0.94	0.89	0.88	0.83	0.90	0.79	0.71	0.69	0.88
Belgium (Flemish)	0.89	46	0.76	0.76	0.62	0.71	0.62	0.61	0.78	0.71	0.52	0.66	0.48	0.71	0.78
Bosnia and Herzegovina	0.89	43	0.74	0.69	0.72	0.67	0.68	0.60	0.69	0.68	0.54	0.56	0.62	0.58	0.72
Bulgaria	0.86	41	0.76	0.78	0.75	0.80	0.65	0.31	0.75	0.66	0.64	0.47	0.53	0.37	0.63
Canada	0.90	47	0.75	0.77	0.73	0.69	0.64	0.66	0.76	0.74	0.51	0.64	0.67	0.65	0.67
Chile	0.92	52	0.68	0.70	0.67	0.59	0.69	0.79	0.83	0.80	0.54	0.69	0.78	0.72	0.82
Chinese Taipei	0.92	51	0.72	0.74	0.70	0.80	0.69	0.76	0.80	0.76	0.66	0.72	0.65	0.60	0.68
Croatia	0.85	38	0.65	0.71	0.67	0.71	0.69	0.65	0.54	0.38	0.56	0.62	0.54	0.41	0.72
Cyprus	0.92	51	0.83	0.81	0.62	0.73	0.78	0.77	0.73	0.54	0.67	0.77	0.69	0.64	0.69
Czech Republic	0.77	29	0.61	0.57	0.28	0.45	0.40	0.47	0.67	0.58	0.37	0.73	0.54	0.45	0.72
Denmark	0.91	49	0.77	0.78	0.62	0.58	0.59	0.68	0.80	0.78	0.59	0.75	0.62	0.68	0.81
England	0.84	37	0.56	0.44	0.60	0.58	0.52	0.58	0.65	0.66	0.49	0.78	0.55	0.67	0.75
Finland	0.88	43	0.69	0.60	0.59	0.57	0.71	0.73	0.72	0.68	0.59	0.74	0.55	0.54	0.73
France	0.82	32	0.57	0.52	0.59	0.43	0.54	0.56	0.69	0.52	0.57	0.63	0.60	0.49	0.62
Georgia	0.89	45	0.75	0.67	0.59	0.66	0.61	0.67	0.71	0.76	0.67	0.74	0.73	0.18	0.72
Germany	0.83	35	0.61	0.67	0.47	0.36	0.52	0.57	0.66	0.75	0.36	0.71	0.63	0.51	0.65
Hong Kong SAR	0.95	61	0.82	0.80	0.71	0.84	0.77	0.74	0.85	0.83	0.76	0.79	0.76	0.70	0.77
Hungary	0.91	48	0.72	0.58	0.66	0.75	0.73	0.62	0.76	0.71	0.53	0.66	0.67	0.69	0.84
Iran, Islamic Rep. of	0.91	47	0.71	0.76	0.68	0.77	0.79	0.74	0.69	0.43	0.79	0.65	0.57	0.54	0.72
Ireland	0.84	36	0.64	0.66	0.54	0.66	0.50	0.64	0.76	0.57	0.37	0.56	0.54	0.60	0.62
Italy	0.89	44	0.67	0.62	0.62	0.58	0.67	0.61	0.70	0.69	0.57	0.71	0.71	0.73	0.73
Japan	0.89	46	0.75	0.75	0.77	0.60	0.77	0.59	0.66	0.66	0.54	0.56	0.55	0.70	0.83
Kazakhstan	0.94	58	0.77	0.73	0.66	0.78	0.76	0.74	0.81	0.78	0.78	0.81	0.82	0.63	0.76
Korea, Rep. of	0.95	63	0.85	0.88	0.74	0.89	0.87	0.79	0.81	0.84	0.69	0.70	0.71	0.64	0.85
Kosovo	0.83	35	0.53	0.64	0.57	0.53	0.45	0.62	0.69	0.63	0.46	0.61	0.55	0.64	0.68
Kuwait	0.94	60	0.84	0.84	0.78	0.83	0.87	0.77	0.82	0.78	0.81	0.63	0.60	0.58	0.84
Latvia	0.89	44	0.74	0.71	0.63	0.71	0.65	0.70	0.62	0.57	0.73	0.56	0.77	0.53	0.67
Lithuania	0.87	40	0.64	0.69	0.36	0.53	0.41	0.67	0.66	0.66	0.62	0.73	0.72	0.61	0.76
Malta	0.89	44	0.79	0.76	0.63	0.54	0.53	0.75	0.83	0.84	0.49	0.62	0.41	0.37	0.79
Montenegro	0.89	44	0.69	0.71	0.70	0.69	0.68	0.71	0.68	0.70	0.51	0.66	0.65	0.55	0.65
Morocco	0.80	38	-0.17	0.10	0.35	0.55	-0.13	0.75	0.72	0.68	0.65	0.83	0.84	0.80	0.68
Netherlands	0.84	35	0.58	0.46	0.62	0.49	0.52	0.68	0.60	0.59	0.45	0.79	0.68	0.50	0.67
New Zealand	0.90	47	0.74	0.78	0.68	0.76	0.69	0.72	0.82	0.71	0.49	0.60	0.63	0.61	0.64
North Macedonia	0.91	47	0.71	0.76	0.70	0.77	0.71	0.79	0.72	0.67	0.59	0.62	0.55	0.60	0.68
Northern Ireland	0.86	40	0.76	0.67	0.65	0.54	0.58	0.59	0.69	0.73	0.51	0.66	0.52	0.56	0.66

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instruction Affected by Mathematics Resource Shortages* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item												
			ACBG13AA	ACBG13AB	ACBG13AC	ACBG13AD	ACBG13AE	ACBG13AF	ACBG13AG	ACBG13AH	ACBG13BA	ACBG13BB	ACBG13BC	ACBG13BD	ACBG13BE
Norway (5)	0.85	37	0.67	0.61	0.63	0.64	0.60	0.60	0.67	0.53	0.58	0.67	0.45	0.54	0.64
Oman	0.94	57	0.86	0.82	0.70	0.85	0.84	0.85	0.70	0.72	0.80	0.75	0.62	0.42	0.79
Pakistan	0.77	31	0.60	0.69	0.80	0.79	0.83	0.75	0.27	-0.01	0.73	-0.13	0.01	0.05	0.27
Philippines	0.85	36	0.57	0.63	0.53	0.59	0.54	0.61	0.70	0.64	0.54	0.54	0.64	0.68	0.57
Poland	0.90	45	0.62	0.61	0.68	0.64	0.73	0.75	0.75	0.67	0.62	0.69	0.55	0.60	0.74
Portugal	0.91	48	0.64	0.64	0.76	0.71	0.76	0.66	0.70	0.68	0.58	0.71	0.72	0.72	0.73
Qatar	0.98	85	0.93	0.95	0.91	0.96	0.94	0.95	0.95	0.93	0.92	0.91	0.91	0.79	0.93
Russian Federation	0.92	52	0.77	0.68	0.66	0.69	0.70	0.76	0.79	0.75	0.66	0.74	0.79	0.56	0.79
Saudi Arabia	0.90	46	0.75	0.77	0.63	0.78	0.74	0.70	0.73	0.47	0.73	0.62	0.42	0.61	0.75
Serbia	0.87	40	0.77	0.73	0.73	0.71	0.75	0.67	0.64	0.59	0.46	0.49	0.47	0.45	0.65
Singapore	0.98	80	0.94	0.93	0.90	0.93	0.89	0.85	0.94	0.94	0.85	0.86	0.78	0.90	0.92
Slovak Republic	0.92	52	0.72	0.82	0.63	0.84	0.65	0.61	0.81	0.72	0.79	0.72	0.72	0.55	0.71
South Africa (5)	0.84	36	0.16	0.11	0.48	0.64	0.41	0.69	0.76	0.70	0.49	0.74	0.76	0.66	0.70
Spain	0.86	39	0.57	0.67	0.61	0.60	0.68	0.65	0.72	0.59	0.59	0.63	0.60	0.53	0.66
Sweden	0.86	40	0.66	0.73	0.58	0.63	0.61	0.62	0.71	0.64	0.58	0.67	0.35	0.63	0.71
Turkey (5)	0.92	52	0.82	0.80	0.67	0.78	0.78	0.59	0.79	0.67	0.76	0.69	0.62	0.52	0.78
United Arab Emirates	0.97	74	0.89	0.88	0.89	0.90	0.89	0.91	0.91	0.86	0.84	0.79	0.78	0.76	0.88
United States	0.93	55	0.77	0.81	0.77	0.78	0.70	0.78	0.79	0.77	0.57	0.67	0.69	0.74	0.75
Benchmarking Participants															
Ontario, Canada	0.89	44	0.68	0.72	0.77	0.63	0.53	0.68	0.76	0.76	0.50	0.64	0.65	0.58	0.62
Quebec, Canada	0.93	54	0.86	0.82	0.68	0.75	0.76	0.69	0.78	0.73	0.63	0.67	0.74	0.71	0.70
Moscow City, Russian Fed.	0.95	65	0.86	0.82	0.77	0.79	0.85	0.83	0.83	0.81	0.80	0.83	0.83	0.59	0.81
Madrid, Spain	0.88	41	0.66	0.68	0.72	0.69	0.70	0.46	0.62	0.61	0.65	0.57	0.66	0.59	0.71
Abu Dhabi, UAE	0.96	70	0.90	0.86	0.86	0.87	0.89	0.91	0.89	0.82	0.80	0.68	0.70	0.75	0.87
Dubai, UAE	0.98	80	0.91	0.94	0.89	0.92	0.85	0.93	0.94	0.88	0.88	0.88	0.86	0.81	0.89

Relationship Between the TIMSS 2019 *Instruction Affected by Mathematics Resource Shortages* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Albania	0.14	0.02	0.03
Armenia	0.05	0.00	0.01
Australia	0.09	0.01	0.01
Austria	0.08	0.01	0.01
Azerbaijan	0.01	0.00	0.00
Bahrain	0.13	0.02	0.01
Belgium (Flemish)	0.02	0.00	0.00
Bosnia and Herzegovina	0.04	0.00	0.00
Bulgaria	0.14	0.02	0.01
Canada	0.08	0.01	0.01
Chile	0.15	0.02	0.04
Chinese Taipei	-0.01	0.00	0.00
Croatia	0.06	0.00	0.00
Cyprus	0.04	0.00	0.01
Czech Republic	-0.03	0.00	0.00
Denmark	0.03	0.00	0.00
England	0.15	0.02	0.00
Finland	0.04	0.00	0.00
France	0.13	0.02	0.01
Georgia	0.03	0.00	0.00
Germany	0.08	0.01	0.00
Hong Kong SAR	-0.06	0.00	0.01
Hungary	-0.04	0.00	0.00
Iran, Islamic Rep. of	0.12	0.01	0.01
Ireland	0.11	0.01	0.01
Italy	0.08	0.01	0.01
Japan	0.01	0.00	0.00
Kazakhstan	-0.01	0.00	0.00
Korea, Rep. of	0.02	0.00	0.00
Kosovo	0.01	0.00	0.00
Kuwait	0.04	0.00	0.01
Latvia	-0.10	0.01	0.01
Lithuania	-0.02	0.00	0.00
Malta	0.08	0.01	0.01
Montenegro	0.02	0.00	0.01
Morocco	-0.25	0.06	0.07
Netherlands	0.04	0.00	0.00
New Zealand	0.09	0.01	0.01

Relationship Between the TIMSS 2019 *Instruction Affected by Mathematics Resource Shortages* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
North Macedonia	0.07	0.01	0.00
Northern Ireland	0.04	0.00	0.00
Norway (5)	0.02	0.00	0.00
Oman	0.05	0.00	0.01
Pakistan	-0.16	0.03	0.02
Philippines	0.00	0.00	0.01
Poland	0.04	0.00	0.00
Portugal	0.06	0.00	0.00
Qatar	0.24	0.06	0.06
Russian Federation	0.11	0.01	0.01
Saudi Arabia	-0.04	0.00	0.01
Serbia	0.11	0.01	0.01
Singapore	0.02	0.00	0.01
Slovak Republic	0.08	0.01	0.01
South Africa (5)	0.10	0.01	0.05
Spain	0.13	0.02	0.02
Sweden	0.05	0.00	0.01
Turkey (5)	0.09	0.01	0.01
United Arab Emirates	0.18	0.03	0.06
United States	0.04	0.00	0.00
International Median	**0.04**	**0.00**	**0.01**
Benchmarking Participants			
Ontario, Canada	0.13	0.02	0.03
Quebec, Canada	0.05	0.00	0.00
Moscow City, Russian Fed.	-0.04	0.00	0.00
Madrid, Spain	0.05	0.00	0.00
Abu Dhabi, UAE	0.22	0.05	0.07
Dubai, UAE	0.05	0.00	0.04

Instruction Affected by Science Resource Shortages – Grade 4

About the Scale

The *Instruction Affected by Science Resource Shortages* scale was created based on principals' responses to twelve items listed below.

How much is your school's capacity to provide instruction affected by a shortage or inadequacy of the following?

		Not at all	A little	Some	A lot
	A. General School Resources				
ACBG13AA [T]	1) Instructional materials (e.g., textbooks)	○	○	○	○
ACBG13AB [T]	2) Supplies (e.g., papers, pencils, materials)	○	○	○	○
ACBG13AC [T]	3) School buildings and grounds	○	○	○	○
ACBG13AD [T]	4) Heating/cooling and lighting systems	○	○	○	○
ACBG13AE [T]	5) Instructional space (e.g., classrooms)	○	○	○	○
ACBG13AF [T]	6) Technologically competent staff	○	○	○	○
ACBG13AG [T]	7) Audio-visual resources for delivery of instruction (e.g., interactive white boards, digital projectors)	○	○	○	○
ACBG13AH [T]	8) Computer technology for teaching and learning (e.g., computers or tablets for student use)	○	○	○	○
	B. Resources for Science Instruction				
ACBG13CA [T]	1) Teachers with a specialization in science	○	○	○	○
ACBG13CB [T]	2) Computer software/applications for science instruction	○	○	○	○
ACBG13CC [T]	3) Library resources relevant to science instruction	○	○	○	○
ACBG13CD [T]	4) Science equipment and materials for experiments	○	○	○	○

Not Affected | **Somewhat Affected** | **Affected A Lot**

Scale Cut Scores **11.4** **7.0**

[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Instruction Affected by Science Resource Shortages* Scale – Grade 4

Item	delta	tau_1	tau_2	tau_3	Infit
ACBG13AA	-0.26317	-0.24640	0.07411	0.17229	0.96
ACBG13AB	-0.49732	-0.22859	0.20397	0.02462	0.89
ACBG13AC	0.00857	-0.72314	0.13827	0.58487	1.07
ACBG13AD	-0.24911	-0.44652	0.18290	0.26362	0.95
ACBG13AE	0.04975	-0.43752	-0.01409	0.45161	1.08
ACBG13AF	0.05434	-1.21339	-0.05364	1.26703	0.97
ACBG13AG	-0.02743	-0.75072	0.03353	0.71719	0.97
ACBG13AH	0.19884	-0.94645	0.03998	0.90647	1.10
ACBG13CA	0.05539	-0.62018	0.03427	0.58591	1.20
ACBG13CB	0.19116	-1.27970	0.01680	1.26290	1.11
ACBG13CC	-0.02535	-1.27782	-0.03564	1.31346	1.02
ACBG13CD	0.50433	-1.02898	-0.04452	1.07350	1.08

Scale Transformation Constants for the TIMSS 2019 *Instruction Affected by Science Resource Shortages* Scale – Grade 4

Scale Transformation Constants	
A = 9.169706	Transformed Scale Score = 9.169706 + 1.659666 • Logit Scale Score
B = 1.659666	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Instruction Affected by Science Resource Shortages* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	2.54711	
1	4.40071	
2	5.27500	
3	5.85711	
4	6.29613	
5	6.64979	
6	6.94814	7.0
7	7.20753	
8	7.43843	
9	7.64797	
10	7.84045	
11	8.02213	
12	8.19407	
13	8.35844	
14	8.51709	
15	8.67164	
16	8.82352	
17	8.97398	
18	9.12232	
19	9.27534	
20	9.42850	
21	9.58484	
22	9.74563	
23	9.91222	
24	10.08616	
25	10.26917	
26	10.46207	
27	10.67024	
28	10.89636	
29	11.14546	
30	11.42448	11.4
31	11.74332	
32	12.11928	
33	12.58077	
34	13.18547	
35	14.08130	
36	15.95552	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instruction Affected by Science Resource Shortages* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item											
			ACBG13AA	ACBG13AB	ACBG13AC	ACBG13AD	ACBG13AE	ACBG13AF	ACBG13AG	ACBG13AH	ACBG13CA	ACBG13CB	ACBG13CC	ACBG13CD
Albania	0.91	50	0.67	0.76	0.79	0.76	0.71	0.71	0.67	0.68	0.68	0.61	0.66	0.71
Armenia	0.87	42	0.60	0.54	0.65	0.69	0.55	0.68	0.68	0.67	0.69	0.69	0.70	0.64
Australia	0.89	46	0.65	0.68	0.71	0.63	0.73	0.72	0.75	0.80	0.47	0.57	0.70	0.68
Austria	0.83	35	0.49	0.55	0.64	0.55	0.63	0.60	0.59	0.64	0.53	0.60	0.67	0.56
Azerbaijan	0.85	38	0.52	0.53	0.56	0.65	0.56	0.73	0.56	0.62	0.67	0.65	0.75	0.54
Bahrain	0.97	76	0.92	0.91	0.88	0.91	0.95	0.88	0.89	0.83	0.90	0.76	0.76	0.87
Belgium (Flemish)	0.89	46	0.66	0.66	0.66	0.72	0.66	0.65	0.77	0.72	0.51	0.76	0.68	0.66
Bosnia and Herzegovina	0.89	44	0.70	0.69	0.73	0.69	0.71	0.62	0.72	0.69	0.58	0.61	0.66	0.53
Bulgaria	0.83	41	0.79	0.81	0.75	0.82	0.65	0.29	0.74	0.64	0.65	0.43	0.50	0.29
Canada	0.89	45	0.74	0.75	0.74	0.68	0.65	0.69	0.75	0.77	0.47	0.60	0.68	0.48
Chile	0.91	51	0.66	0.68	0.70	0.62	0.74	0.77	0.80	0.77	0.64	0.69	0.77	0.69
Chinese Taipei	0.92	55	0.74	0.73	0.70	0.79	0.72	0.75	0.82	0.79	0.70	0.69	0.70	0.78
Croatia	0.87	42	0.67	0.73	0.69	0.74	0.66	0.68	0.57	0.35	0.60	0.63	0.57	0.75
Cyprus	0.91	52	0.84	0.80	0.64	0.75	0.80	0.77	0.71	0.59	0.59	0.78	0.65	0.67
Czech Republic	0.77	29	0.57	0.55	0.37	0.42	0.42	0.46	0.65	0.63	0.40	0.74	0.57	0.60
Denmark	0.90	48	0.79	0.75	0.63	0.57	0.58	0.68	0.81	0.77	0.59	0.70	0.68	0.70
England	0.84	38	0.48	0.26	0.67	0.56	0.49	0.63	0.61	0.65	0.60	0.71	0.76	0.77
Finland	0.88	43	0.69	0.55	0.63	0.57	0.73	0.71	0.71	0.69	0.58	0.61	0.61	0.73
France	0.81	34	0.45	0.37	0.44	0.28	0.39	0.65	0.72	0.61	0.66	0.74	0.74	0.65
Georgia	0.90	49	0.74	0.64	0.60	0.66	0.63	0.69	0.72	0.76	0.66	0.78	0.74	0.73
Germany	0.83	37	0.62	0.66	0.54	0.41	0.56	0.55	0.67	0.74	0.48	0.68	0.65	0.60
Hong Kong SAR	0.93	58	0.83	0.81	0.74	0.85	0.80	0.77	0.84	0.83	0.50	0.62	0.77	0.69
Hungary	0.90	48	0.71	0.61	0.70	0.76	0.75	0.62	0.76	0.71	0.59	0.74	0.67	0.65
Iran, Islamic Rep. of	0.91	51	0.72	0.76	0.71	0.78	0.81	0.74	0.67	0.41	0.79	0.74	0.61	0.71
Ireland	0.82	34	0.60	0.61	0.58	0.64	0.55	0.64	0.79	0.64	0.52	0.48	0.47	0.45
Italy	0.87	42	0.63	0.62	0.64	0.59	0.69	0.65	0.71	0.67	0.58	0.73	0.67	0.53
Japan	0.90	50	0.77	0.79	0.78	0.60	0.79	0.61	0.65	0.68	0.61	0.59	0.68	0.86
Kazakhstan	0.93	58	0.77	0.74	0.67	0.77	0.76	0.75	0.81	0.78	0.80	0.79	0.76	0.73
Korea, Rep. of	0.96	70	0.87	0.90	0.78	0.90	0.89	0.80	0.83	0.84	0.73	0.77	0.81	0.88
Kosovo	0.83	35	0.59	0.68	0.60	0.61	0.53	0.64	0.65	0.61	0.50	0.60	0.54	0.55
Kuwait	0.95	64	0.85	0.84	0.78	0.83	0.88	0.79	0.81	0.78	0.81	0.68	0.66	0.86
Latvia	0.89	46	0.76	0.70	0.66	0.71	0.68	0.68	0.61	0.57	0.71	0.63	0.72	0.69
Lithuania	0.86	40	0.66	0.67	0.40	0.54	0.44	0.70	0.67	0.70	0.55	0.77	0.68	0.72
Malta	0.88	45	0.77	0.72	0.67	0.56	0.58	0.77	0.80	0.83	0.53	0.56	0.44	0.67
Montenegro	0.89	45	0.72	0.72	0.72	0.71	0.67	0.71	0.70	0.71	0.51	0.62	0.70	0.57
Morocco	0.76	37	-0.21	0.07	0.33	0.55	-0.18	0.77	0.75	0.70	0.64	0.84	0.83	0.71
Netherlands	0.83	37	0.60	0.44	0.71	0.53	0.47	0.64	0.65	0.45	0.60	0.67	0.71	0.70
New Zealand	0.88	46	0.75	0.80	0.73	0.77	0.72	0.72	0.83	0.74	0.31	0.48	0.61	0.44
North Macedonia	0.90	48	0.72	0.78	0.70	0.78	0.72	0.78	0.75	0.68	0.60	0.64	0.61	0.51
Northern Ireland	0.86	40	0.69	0.61	0.64	0.48	0.55	0.62	0.69	0.74	0.62	0.58	0.64	0.69

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instruction Affected by Science Resource Shortages* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item											
			ACBG13AA	ACBG13AB	ACBG13AC	ACBG13AD	ACBG13AE	ACBG13AF	ACBG13AG	ACBG13AH	ACBG13CA	ACBG13CB	ACBG13CC	ACBG13CD
Norway (5)	0.83	35	0.65	0.50	0.65	0.63	0.63	0.67	0.67	0.54	0.51	0.63	0.53	0.48
Oman	0.94	62	0.86	0.82	0.71	0.85	0.86	0.85	0.69	0.72	0.83	0.74	0.73	0.73
Pakistan	0.77	33	0.26	0.57	0.41	0.64	0.45	0.73	0.65	0.50	0.67	0.59	0.69	0.59
Philippines	0.83	36	0.59	0.65	0.57	0.62	0.59	0.63	0.70	0.60	0.42	0.50	0.65	0.66
Poland	0.89	46	0.59	0.61	0.68	0.68	0.73	0.76	0.75	0.72	0.67	0.66	0.66	0.59
Portugal	0.90	47	0.62	0.64	0.75	0.70	0.76	0.68	0.71	0.70	0.55	0.69	0.67	0.71
Qatar	0.99	87	0.94	0.95	0.92	0.96	0.94	0.95	0.95	0.93	0.93	0.88	0.92	0.92
Russian Federation	0.92	54	0.76	0.67	0.68	0.67	0.71	0.77	0.78	0.77	0.68	0.78	0.77	0.76
Saudi Arabia	0.90	48	0.72	0.75	0.62	0.76	0.71	0.71	0.77	0.53	0.72	0.70	0.56	0.72
Serbia	0.88	44	0.76	0.72	0.75	0.70	0.76	0.68	0.69	0.59	0.43	0.53	0.62	0.64
Singapore	0.98	80	0.93	0.92	0.90	0.93	0.90	0.86	0.95	0.94	0.81	0.88	0.84	0.90
Slovak Republic	0.91	51	0.71	0.81	0.65	0.84	0.67	0.62	0.82	0.73	0.77	0.71	0.70	0.43
South Africa (5)	0.83	38	0.02	-0.01	0.43	0.63	0.34	0.63	0.80	0.75	0.60	0.79	0.82	0.80
Spain	0.86	40	0.56	0.65	0.65	0.62	0.69	0.68	0.73	0.60	0.59	0.59	0.65	0.56
Sweden	0.85	39	0.67	0.72	0.62	0.65	0.65	0.63	0.71	0.67	0.57	0.59	0.36	0.61
Turkey (5)	0.92	53	0.81	0.79	0.69	0.80	0.80	0.60	0.80	0.68	0.77	0.67	0.60	0.72
United Arab Emirates	0.97	77	0.89	0.88	0.89	0.90	0.90	0.91	0.91	0.86	0.85	0.78	0.83	0.90
United States	0.92	53	0.76	0.77	0.78	0.78	0.72	0.79	0.79	0.76	0.57	0.61	0.69	0.66
Benchmarking Participants														
Ontario, Canada	0.89	45	0.69	0.70	0.77	0.61	0.53	0.69	0.73	0.78	0.62	0.64	0.71	0.56
Quebec, Canada	0.88	47	0.86	0.83	0.71	0.76	0.79	0.71	0.80	0.75	0.25	0.53	0.62	0.28
Moscow City, Russian Fed.	0.95	67	0.86	0.81	0.76	0.78	0.85	0.83	0.83	0.82	0.81	0.79	0.81	0.85
Madrid, Spain	0.87	41	0.66	0.67	0.73	0.67	0.70	0.50	0.60	0.60	0.63	0.63	0.65	0.63
Abu Dhabi, UAE	0.96	72	0.89	0.86	0.87	0.88	0.89	0.91	0.90	0.83	0.82	0.65	0.77	0.89
Dubai, UAE	0.98	81	0.91	0.95	0.88	0.92	0.85	0.94	0.94	0.90	0.88	0.84	0.89	0.88

Relationship Between the TIMSS 2019 *Instruction Affected by Science Resource Shortages* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Albania	0.13	0.02	0.04
Armenia	0.06	0.00	0.01
Australia	0.10	0.01	0.01
Austria	0.13	0.02	0.02
Azerbaijan	0.01	0.00	0.00
Bahrain	0.18	0.03	0.03
Belgium (Flemish)	0.02	0.00	0.00
Bosnia and Herzegovina	0.05	0.00	0.00
Bulgaria	0.16	0.02	0.01
Canada	0.08	0.01	0.00
Chile	0.14	0.02	0.03
Chinese Taipei	-0.01	0.00	0.00
Croatia	0.02	0.00	0.00
Cyprus	0.01	0.00	0.01
Czech Republic	-0.05	0.00	0.00
Denmark	-0.01	0.00	0.00
England	0.17	0.03	0.02
Finland	0.02	0.00	0.00
France	0.14	0.02	0.02
Georgia	0.02	0.00	0.00
Germany	0.07	0.00	0.00
Hong Kong SAR	-0.05	0.00	0.01
Hungary	-0.04	0.00	0.00
Iran, Islamic Rep. of	0.13	0.02	0.02
Ireland	0.05	0.00	0.01
Italy	0.05	0.00	0.00
Japan	0.01	0.00	0.00
Kazakhstan	0.02	0.00	0.00
Korea, Rep. of	0.02	0.00	0.00
Kosovo	0.01	0.00	0.00
Kuwait	-0.01	0.00	0.01
Latvia	-0.06	0.00	0.01
Lithuania	-0.02	0.00	0.00
Malta	0.13	0.02	0.01
Montenegro	0.00	0.00	0.01
Morocco	-0.22	0.05	0.05
Netherlands	0.07	0.00	0.00
New Zealand	0.08	0.01	0.01

Relationship Between the TIMSS 2019 *Instruction Affected by Science Resource Shortages* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
North Macedonia	0.10	0.01	0.01
Northern Ireland	0.02	0.00	0.00
Norway (5)	0.04	0.00	0.00
Oman	0.04	0.00	0.00
Pakistan	-0.16	0.03	0.03
Philippines	-0.02	0.00	0.02
Poland	0.04	0.00	0.00
Portugal	0.06	0.00	0.00
Qatar	0.20	0.04	0.04
Russian Federation	0.08	0.01	0.00
Saudi Arabia	-0.04	0.00	0.01
Serbia	0.12	0.01	0.01
Singapore	0.01	0.00	0.00
Slovak Republic	0.11	0.01	0.01
South Africa (5)	0.10	0.01	0.05
Spain	0.15	0.02	0.02
Sweden	0.06	0.00	0.01
Turkey (5)	0.09	0.01	0.01
United Arab Emirates	0.15	0.02	0.06
United States	0.05	0.00	0.00
International Median	**0.05**	**0.00**	**0.01**
Benchmarking Participants			
Ontario, Canada	0.10	0.01	0.01
Quebec, Canada	0.02	0.00	0.00
Moscow City, Russian Fed.	-0.05	0.00	0.00
Madrid, Spain	0.05	0.00	0.00
Abu Dhabi, UAE	0.21	0.05	0.07
Dubai, UAE	-0.02	0.00	0.03

School Discipline – Grade 4

About the Scale

The *School Discipline* scale was created based on principals' responses to ten items listed below.

To what degree is each of the following a problem among fourth grade students in your school?

		Not a problem	Minor problem	Moderate problem	Severe Problem
ACBG15A T	1) Arriving late at school	○	○	○	○
ACBG15B T	2) Absenteeism (i.e., unjustified absences)	○	○	○	○
ACBG15C T	3) Classroom disturbance	○	○	○	○
ACBG15D T	4) Cheating	○	○	○	○
ACBG15E T	5) Profanity	○	○	○	○
ACBG15F T	6) Vandalism	○	○	○	○
ACBG15G T	7) Theft	○	○	○	○
ACBG15H T	8) Intimidation or verbal abuse among students (including texting, emailing, etc.)	○	○	○	○
ACBG15I T	9) Physical injury to other students	○	○	○	○
ACBG15J T	10) Intimidation or verbal abuse of teachers or staff (including texting, emailing, etc.)	○	○	○	○

Hardly Any Problems	Minor Problems	Moderate to Severe Problems

Scale Cut Scores 9.7 7.6

T Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *School Discipline* Scale – Grade 4

Item	delta	tau_1	tau_2	tau_3	Infit
ACBG15A	-0.04835	-2.12043	-0.42165	2.54208	1.27
ACBG15B	0.28790	-1.45910	-0.51461	1.97371	1.22
ACBG15C	0.78849	-2.13154	-0.25288	2.38442	1.00
ACBG15D	-0.50395	-1.36108	-0.70227	2.06335	1.07
ACBG15E	0.48418	-1.65568	-0.39478	2.05046	0.91
ACBG15F	-0.31766	-0.47898	-0.62657	1.10555	0.86
ACBG15G	-0.53958	0.31575	-1.17406	0.85831	0.77
ACBG15H	0.21488	-1.40063	-0.58577	1.98640	0.94
ACBG15I	0.30591	-1.40756	-0.73711	2.14467	0.85
ACBG15J	-0.67182	0.16233	-0.91357	0.75124	0.87

Scale Transformation Constants for the TIMSS 2019 *School Discipline* Scale – Grade 4

Scale Transformation Constants	
A = 7.809340	Transformed Scale Score = 7.809340 + 0.952078 • Logit Scale Score
B = 0.952078	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *School Discipline* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	3.70440	
1	4.77250	
2	5.27812	
3	5.61661	
4	5.87500	
5	6.08629	
6	6.26778	
7	6.42945	
8	6.57768	
9	6.71549	
10	6.84975	
11	6.98118	
12	7.11207	
13	7.24453	
14	7.38070	
15	7.52341	7.6
16	7.67299	
17	7.83372	
18	8.00732	
19	8.19660	
20	8.40499	
21	8.63392	
22	8.88534	
23	9.16002	
24	9.45841	
25	9.78042	9.7
26	10.13101	
27	10.52330	
28	10.98914	
29	11.61576	
30	12.79234	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *School Discipline* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item									
			ACBG15A	ACBG15B	ACBG15C	ACBG15D	ACBG15E	ACBG15F	ACBG15G	ACBG15H	ACBG15I	ACBG15J
Albania	0.92	61	0.67	0.80	0.63	0.73	0.75	0.86	0.88	0.83	0.80	0.85
Armenia	0.95	67	0.68	0.77	0.83	0.67	0.88	0.87	0.84	0.88	0.87	0.88
Australia	0.89	52	0.61	0.67	0.75	0.65	0.82	0.74	0.63	0.76	0.80	0.74
Austria	0.89	51	0.59	0.68	0.73	0.65	0.82	0.73	0.70	0.76	0.84	0.62
Azerbaijan	0.96	76	0.73	0.85	0.88	0.73	0.88	0.93	0.94	0.92	0.90	0.93
Bahrain	0.97	80	0.83	0.86	0.83	0.93	0.90	0.93	0.91	0.93	0.93	0.88
Belgium (Flemish)	0.85	45	0.48	0.52	0.76	0.52	0.78	0.71	0.58	0.78	0.80	0.69
Bosnia and Herzegovina	0.96	75	0.74	0.83	0.88	0.75	0.85	0.93	0.94	0.88	0.90	0.92
Bulgaria	0.94	67	0.69	0.82	0.80	0.85	0.81	0.89	0.88	0.81	0.81	0.81
Canada	0.88	49	0.61	0.64	0.70	0.62	0.81	0.76	0.68	0.70	0.74	0.72
Chile	0.90	55	0.54	0.63	0.76	0.76	0.82	0.76	0.75	0.80	0.78	0.75
Chinese Taipei	0.90	54	0.69	0.65	0.69	0.73	0.78	0.81	0.82	0.80	0.79	0.56
Croatia	0.86	45	0.63	0.56	0.63	0.71	0.68	0.76	0.60	0.73	0.74	0.66
Cyprus	0.90	55	0.66	0.72	0.79	0.63	0.80	0.73	0.69	0.82	0.81	0.72
Czech Republic	0.87	47	0.52	0.47	0.67	0.77	0.81	0.76	0.72	0.75	0.76	0.58
Denmark	0.80	36	0.61	0.57	0.61	0.23	0.71	0.58	0.49	0.70	0.76	0.62
England	0.63	24	0.28	0.31	0.50	0.38	0.64	0.60	0.24	0.48	0.63	0.64
Finland	0.84	42	0.50	0.57	0.76	0.43	0.77	0.68	0.59	0.80	0.71	0.58
France	0.88	49	0.53	0.64	0.62	0.64	0.77	0.72	0.67	0.77	0.84	0.74
Georgia	0.97	76	0.77	0.77	0.85	0.80	0.92	0.93	0.93	0.93	0.89	0.89
Germany	0.88	49	0.63	0.61	0.75	0.67	0.79	0.77	0.73	0.70	0.75	0.61
Hong Kong SAR	0.88	49	0.70	0.60	0.71	0.71	0.82	0.79	0.73	0.58	0.77	0.56
Hungary	0.88	50	0.61	0.71	0.72	0.74	0.72	0.71	0.74	0.77	0.71	0.65
Iran, Islamic Rep. of	0.91	55	0.69	0.70	0.74	0.76	0.83	0.78	0.79	0.75	0.70	0.62
Ireland	0.82	41	0.58	0.60	0.80	0.64	0.73	0.51	0.64	0.73	0.63	0.47
Italy	0.91	58	0.45	0.73	0.66	0.61	0.68	0.88	0.87	0.83	0.89	0.86
Japan	0.88	54	0.56	0.42	0.82	0.80	0.87	0.78	0.78	0.80	0.60	0.80
Kazakhstan	0.97	82	0.71	0.92	0.90	0.76	0.90	0.96	0.96	0.96	0.95	0.96
Korea, Rep. of	0.96	73	0.70	0.81	0.78	0.90	0.84	0.91	0.92	0.89	0.89	0.87
Kosovo	0.96	75	0.76	0.79	0.85	0.84	0.87	0.92	0.91	0.92	0.88	0.90
Kuwait	0.94	63	0.61	0.67	0.82	0.83	0.85	0.86	0.83	0.84	0.79	0.80
Latvia	0.79	35	0.47	0.48	0.57	0.51	0.76	0.58	0.60	0.68	0.73	0.44
Lithuania	0.82	40	0.68	0.66	0.60	0.72	0.63	0.71	0.51	0.56	0.51	0.67
Malta	0.88	50	0.68	0.65	0.68	0.69	0.72	0.79	0.69	0.75	0.71	0.72
Montenegro	0.96	73	0.69	0.82	0.86	0.67	0.89	0.91	0.91	0.91	0.89	0.92
Morocco	0.96	76	0.63	0.76	0.87	0.91	0.92	0.93	0.94	0.89	0.91	0.87
Netherlands	0.72	31	0.28	0.41	0.75	0.43	0.59	0.44	0.49	0.68	0.59	0.70
New Zealand	0.89	52	0.65	0.67	0.76	0.55	0.80	0.69	0.74	0.75	0.83	0.76
North Macedonia	0.95	68	0.64	0.73	0.87	0.81	0.84	0.85	0.89	0.86	0.88	0.86
Northern Ireland	0.79	38	0.48	0.53	0.72	0.63	0.71	0.40	0.53	0.64	0.76	0.62

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *School Discipline* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item									
			ACBG15A	ACBG15B	ACBG15C	ACBG15D	ACBG15E	ACBG15F	ACBG15G	ACBG15H	ACBG15I	ACBG15J
Norway (5)	0.88	51	0.64	0.52	0.75	0.68	0.77	0.76	0.69	0.73	0.76	0.79
Oman	0.96	75	0.73	0.82	0.76	0.92	0.89	0.92	0.92	0.91	0.87	0.88
Pakistan	0.91	56	0.65	0.64	0.65	0.79	0.84	0.78	0.86	0.80	0.77	0.69
Philippines	0.87	47	0.55	0.67	0.60	0.76	0.74	0.73	0.73	0.68	0.70	0.65
Poland	0.84	42	0.52	0.64	0.69	0.58	0.74	0.79	0.64	0.66	0.68	0.44
Portugal	0.93	63	0.55	0.83	0.77	0.69	0.75	0.86	0.88	0.85	0.81	0.87
Qatar	0.95	69	0.71	0.77	0.80	0.90	0.82	0.86	0.86	0.84	0.86	0.86
Russian Federation	0.82	40	0.58	0.51	0.58	0.42	0.62	0.73	0.75	0.75	0.70	0.59
Saudi Arabia	0.96	75	0.59	0.73	0.80	0.91	0.91	0.93	0.94	0.93	0.94	0.92
Serbia	0.96	72	0.83	0.80	0.85	0.81	0.81	0.91	0.91	0.87	0.84	0.87
Singapore	0.84	42	0.61	0.61	0.69	0.62	0.73	0.67	0.64	0.66	0.75	0.49
Slovak Republic	0.90	54	0.69	0.71	0.70	0.69	0.82	0.81	0.76	0.74	0.70	0.73
South Africa (5)	0.89	51	0.63	0.65	0.75	0.70	0.73	0.75	0.73	0.76	0.82	0.58
Spain	0.91	57	0.59	0.72	0.66	0.66	0.74	0.86	0.78	0.80	0.83	0.84
Sweden	0.81	38	0.58	0.55	0.73	0.45	0.70	0.63	0.26	0.64	0.73	0.72
Turkey (5)	0.96	73	0.74	0.79	0.85	0.88	0.82	0.88	0.89	0.90	0.84	0.90
United Arab Emirates	0.93	63	0.65	0.70	0.74	0.81	0.81	0.85	0.85	0.85	0.84	0.81
United States	0.89	53	0.64	0.66	0.65	0.68	0.76	0.76	0.79	0.77	0.78	0.75
Benchmarking Participants												
Ontario, Canada	0.88	50	0.62	0.65	0.73	0.55	0.82	0.75	0.71	0.73	0.75	0.72
Quebec, Canada	0.87	48	0.56	0.66	0.61	0.69	0.78	0.78	0.68	0.69	0.69	0.74
Moscow City, Russian Fed.	0.87	48	0.59	0.61	0.64	0.65	0.72	0.78	0.76	0.75	0.70	0.73
Madrid, Spain	0.93	62	0.60	0.79	0.77	0.77	0.81	0.89	0.79	0.81	0.82	0.81
Abu Dhabi, UAE	0.90	53	0.64	0.64	0.71	0.69	0.75	0.76	0.76	0.83	0.83	0.67
Dubai, UAE	0.83	45	0.47	0.57	0.61	0.70	0.75	0.77	0.73	0.62	0.68	0.74

Relationship Between the TIMSS 2019 *School Discipline* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Albania	0.02	0.00	0.00	0.03	0.00	0.00
Armenia	0.05	0.00	0.01	0.06	0.00	0.01
Australia	0.24	0.06	0.05	0.23	0.05	0.05
Austria	0.16	0.02	0.02	0.19	0.04	0.03
Azerbaijan	0.00	0.00	0.01	-0.01	0.00	0.01
Bahrain	0.16	0.03	0.02	0.26	0.07	0.05
Belgium (Flemish)	0.14	0.02	0.02	0.16	0.03	0.02
Bosnia and Herzegovina	0.01	0.00	0.00	0.05	0.00	0.00
Bulgaria	0.21	0.04	0.06	0.21	0.04	0.06
Canada	0.12	0.01	0.02	0.10	0.01	0.01
Chile	0.18	0.03	0.03	0.18	0.03	0.02
Chinese Taipei	0.01	0.00	0.00	0.01	0.00	0.00
Croatia	-0.04	0.00	0.00	-0.03	0.00	0.00
Cyprus	0.13	0.02	0.02	0.12	0.01	0.01
Czech Republic	0.11	0.01	0.01	0.11	0.01	0.01
Denmark	0.10	0.01	0.01	0.09	0.01	0.01
England	0.23	0.05	0.02	0.26	0.07	0.02
Finland	0.08	0.01	0.01	0.08	0.01	0.02
France	0.17	0.03	0.02	0.17	0.03	0.03
Georgia	0.03	0.00	0.00	0.04	0.00	0.00
Germany	0.25	0.06	0.06	0.26	0.07	0.07
Hong Kong SAR	0.15	0.02	0.02	0.18	0.03	0.03
Hungary	0.18	0.03	0.04	0.17	0.03	0.04
Iran, Islamic Rep. of	0.14	0.02	0.01	0.15	0.02	0.01
Ireland	0.15	0.02	0.01	0.15	0.02	0.01
Italy	0.06	0.00	0.00	0.07	0.01	0.00
Japan	0.05	0.00	0.00	0.04	0.00	0.00
Kazakhstan	0.05	0.00	0.00	0.05	0.00	0.00
Korea, Rep. of	0.07	0.00	0.01	0.06	0.00	0.00
Kosovo	-0.07	0.00	0.00	-0.08	0.01	0.01
Kuwait	0.14	0.02	0.02	0.20	0.04	0.04
Latvia	0.00	0.00	0.00	0.01	0.00	0.00
Lithuania	-0.11	0.01	0.01	-0.12	0.01	0.01
Malta	0.17	0.03	0.04	0.17	0.03	0.04
Montenegro	-0.02	0.00	0.00	-0.02	0.00	0.00
Morocco	0.03	0.00	0.02	0.02	0.00	0.02
Netherlands	0.11	0.01	0.01	0.15	0.02	0.01

Relationship Between the TIMSS 2019 *School Discipline* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
New Zealand	0.25	0.06	0.05	0.26	0.07	0.06
North Macedonia	0.08	0.01	0.01	0.09	0.01	0.01
Northern Ireland	0.15	0.02	0.02	0.13	0.02	0.01
Norway (5)	0.06	0.00	0.00	0.06	0.00	0.01
Oman	0.01	0.00	0.00	0.01	0.00	0.00
Pakistan	-0.06	0.00	0.01	-0.06	0.00	0.01
Philippines	0.13	0.02	0.04	0.13	0.02	0.04
Poland	-0.01	0.00	0.00	-0.01	0.00	0.00
Portugal	0.09	0.01	0.01	0.08	0.01	0.01
Qatar	0.05	0.00	0.01	0.06	0.00	0.01
Russian Federation	0.03	0.00	0.00	0.02	0.00	0.00
Saudi Arabia	0.07	0.00	0.00	0.07	0.01	0.01
Serbia	-0.03	0.00	0.00	-0.03	0.00	0.00
Singapore	0.05	0.00	0.00	0.06	0.00	0.00
Slovak Republic	0.17	0.03	0.05	0.20	0.04	0.09
South Africa (5)	0.13	0.02	0.01	0.12	0.02	0.01
Spain	0.15	0.02	0.02	0.16	0.03	0.02
Sweden	0.17	0.03	0.01	0.17	0.03	0.01
Turkey (5)	0.20	0.04	0.05	0.20	0.04	0.05
United Arab Emirates	0.22	0.05	0.05	0.24	0.06	0.06
United States	0.18	0.03	0.03	0.20	0.04	0.03
International Median	**0.11**	**0.01**	**0.01**	**0.10**	**0.01**	**0.01**
Benchmarking Participants						
Ontario, Canada	0.17	0.03	0.04	0.13	0.02	0.02
Quebec, Canada	0.03	0.00	0.01	0.02	0.00	0.00
Moscow City, Russian Fed.	0.00	0.00	0.00	0.00	0.00	0.00
Madrid, Spain	0.10	0.01	0.03	0.10	0.01	0.02
Abu Dhabi, UAE	0.22	0.05	0.04	0.25	0.06	0.05
Dubai, UAE	0.04	0.00	0.00	0.05	0.00	0.00

School Emphasis on Academic Success—Principals' Reports – Grade 4

About the Scale

The *School Emphasis on Academic Success* scale was created based on principals' responses to eleven items listed below.[1]

How would you characterize each of the following within your school?

		Very high	High	Medium	Low	Very low
ACBG14A T	1) Teachers' understanding of the school's curricular goals					
ACBG14B T	2) Teachers' degree of success in implementing the school's curriculum					
ACBG14C T	3) Teachers' expectations for student achievement					
ACBG14D T	4) Teachers' ability to inspire students					
ACBG14E T	5) Parental involvement in school activities					
ACBG14F T	6) Parental commitment to ensure that students are ready to learn					
ACBG14G T	7) Parental expectations for student achievement					
ACBG14H T	8) Parental support for student achievement					
ACBG14I T	9) Students' desire to do well in school					
ACBG14J T	10) Students' ability to reach school's academic goals					
ACBG14K T	11) Students' respect for classmates who excel academically					

Very High Emphasis | High Emphasis | Medium Emphasis

Scale Cut Scores 13.0 9.2

[1] For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Low" and "Very low" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

T Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *School Emphasis on Academic Success—Principals' Reports* Scale – Grade 4

Item	delta	tau_1	tau_2	tau_3	Infit
ACBG14A	-1.51453	-3.71990	0.14195	3.57795	1.12
ACBG14B	-1.06786	-3.81822	0.14006	3.67816	0.98
ACBG14C	-0.77174	-3.36668	-0.07105	3.43773	1.03
ACBG14D	-0.86155	-3.51615	0.09339	3.42276	1.03
ACBG14E	1.29590	-2.49610	0.11848	2.37762	1.15
ACBG14F	1.41750	-2.64527	0.12898	2.51629	0.95
ACBG14G	-0.15848	-2.55875	-0.22955	2.78830	1.14
ACBG14H	1.31862	-2.82962	0.18549	2.64413	0.90
ACBG14I	0.06237	-3.37852	0.05400	3.32452	0.92
ACBG14J	0.48757	-3.88724	0.26174	3.62550	0.88
ACBG14K	-0.20780	-3.11968	-0.13994	3.25962	1.14

Scale Transformation Constants for the TIMSS 2019 *School Emphasis on Academic Success—Principals' Reports* Scale – Grade 4

Scale Transformation Constants	
A = 9.213867	Transformed Scale Score = 9.213867 + 1.105570 • Logit Scale Score
B = 1.105570	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019
***School Emphasis on Academic Success—Principals' Reports* Scale – Grade 4**

Raw Score	Transformed Scale Score	Cutpoint
0	1.33825	
1	2.77163	
2	3.56947	
3	4.18606	
4	4.71768	
5	5.20235	
6	5.65593	
7	6.08580	
8	6.49334	
9	6.87782	
10	7.23985	
11	7.58308	
12	7.91194	
13	8.23156	
14	8.54700	
15	8.86258	
16	9.18132	9.2
17	9.50476	
18	9.83206	
19	10.16093	
20	10.48841	
21	10.81202	
22	11.13070	
23	11.44500	
24	11.75683	
25	12.06931	
26	12.38668	
27	12.71485	
28	13.06080	13.0
29	13.43532	
30	13.85847	
31	14.36918	
32	15.06494	
33	16.40427	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *School Emphasis on Academic Success—Principals' Reports* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item										
			ACBG14A	ACBG14B	ACBG14C	ACBG14D	ACBG14E	ACBG14F	ACBG14G	ACBG14H	ACBG14I	ACBG14J	ACBG14K
Albania	0.86	42	0.62	0.63	0.55	0.62	0.65	0.74	0.63	0.73	0.69	0.73	0.54
Armenia	0.86	43	0.55	0.55	0.63	0.64	0.60	0.81	0.61	0.79	0.73	0.61	0.67
Australia	0.92	55	0.69	0.78	0.74	0.70	0.77	0.75	0.71	0.77	0.78	0.77	0.72
Austria	0.85	41	0.43	0.68	0.68	0.47	0.75	0.80	0.59	0.72	0.60	0.72	0.45
Azerbaijan	0.89	48	0.62	0.73	0.77	0.72	0.65	0.76	0.71	0.70	0.70	0.72	0.55
Bahrain	0.94	63	0.78	0.81	0.83	0.78	0.69	0.79	0.79	0.85	0.85	0.84	0.72
Belgium (Flemish)	0.82	37	0.29	0.48	0.47	0.44	0.74	0.80	0.50	0.74	0.72	0.69	0.60
Bosnia and Herzegovina	0.88	47	0.65	0.65	0.62	0.76	0.68	0.79	0.39	0.67	0.72	0.77	0.73
Bulgaria	0.92	56	0.60	0.69	0.73	0.71	0.77	0.80	0.71	0.82	0.80	0.82	0.72
Canada	0.91	52	0.61	0.68	0.71	0.68	0.76	0.78	0.73	0.81	0.74	0.76	0.68
Chile	0.92	55	0.71	0.78	0.79	0.77	0.70	0.76	0.78	0.79	0.78	0.63	0.69
Chinese Taipei	0.91	53	0.65	0.73	0.66	0.75	0.66	0.77	0.70	0.81	0.78	0.81	0.67
Croatia	0.83	38	0.42	0.64	0.75	0.69	0.62	0.67	0.55	0.65	0.64	0.58	0.52
Cyprus	0.92	55	0.54	0.68	0.59	0.61	0.76	0.88	0.81	0.87	0.82	0.88	0.65
Czech Republic	0.87	44	0.46	0.63	0.58	0.66	0.74	0.73	0.65	0.72	0.75	0.68	0.64
Denmark	0.92	56	0.71	0.74	0.73	0.66	0.73	0.81	0.81	0.85	0.72	0.73	0.69
England	0.88	47	0.70	0.68	0.58	0.63	0.79	0.83	0.81	0.83	0.60	0.63	0.29
Finland	0.88	47	0.57	0.59	0.70	0.61	0.70	0.77	0.71	0.75	0.77	0.73	0.62
France	0.83	40	0.46	0.45	0.43	0.67	0.67	0.79	0.59	0.81	0.76	0.72	0.41
Georgia	0.90	49	0.68	0.68	0.78	0.70	0.74	0.78	0.65	0.78	0.74	0.58	0.58
Germany	0.86	42	0.51	0.56	0.68	0.52	0.73	0.80	0.64	0.78	0.62	0.70	0.54
Hong Kong SAR	0.92	57	0.65	0.78	0.74	0.78	0.66	0.84	0.82	0.81	0.73	0.83	0.65
Hungary	0.89	50	0.37	0.68	0.67	0.75	0.69	0.83	0.59	0.82	0.79	0.75	0.67
Iran, Islamic Rep. of	0.89	49	0.64	0.76	0.55	0.68	0.72	0.76	0.69	0.73	0.69	0.77	0.68
Ireland	0.92	56	0.65	0.66	0.73	0.70	0.77	0.86	0.83	0.80	0.78	0.81	0.58
Italy	0.87	45	0.68	0.73	0.67	0.67	0.58	0.72	0.54	0.77	0.67	0.78	0.53
Japan	0.89	47	0.56	0.58	0.72	0.59	0.68	0.79	0.77	0.82	0.72	0.75	0.47
Kazakhstan	0.90	50	0.63	0.67	0.70	0.74	0.66	0.77	0.70	0.80	0.68	0.71	0.72
Korea, Rep. of	0.92	56	0.64	0.70	0.73	0.70	0.79	0.78	0.72	0.82	0.81	0.78	0.76
Kosovo	0.85	41	0.62	0.61	0.64	0.68	0.65	0.72	0.60	0.63	0.67	0.75	0.43
Kuwait	0.92	57	0.72	0.70	0.78	0.77	0.75	0.80	0.77	0.80	0.77	0.78	0.65
Latvia	0.78	32	0.47	0.46	0.53	0.63	0.62	0.65	0.43	0.72	0.64	0.53	0.41
Lithuania	0.87	43	0.67	0.72	0.71	0.73	0.72	0.71	0.59	0.68	0.55	0.59	0.50
Malta	0.89	48	0.41	0.58	0.71	0.67	0.60	0.84	0.78	0.77	0.75	0.76	0.66
Montenegro	0.82	37	0.69	0.68	0.39	0.53	0.68	0.72	0.34	0.67	0.58	0.68	0.62
Morocco	0.90	51	0.66	0.65	0.74	0.71	0.75	0.81	0.74	0.76	0.67	0.74	0.57
Netherlands	0.81	37	0.55	0.67	0.62	0.60	0.12	0.53	0.65	0.64	0.67	0.77	0.59
New Zealand	0.90	52	0.60	0.71	0.72	0.69	0.63	0.74	0.77	0.79	0.77	0.77	0.67
North Macedonia	0.90	52	0.68	0.70	0.74	0.72	0.71	0.81	0.61	0.72	0.74	0.79	0.68
Northern Ireland	0.90	51	0.43	0.62	0.80	0.60	0.69	0.82	0.84	0.77	0.81	0.72	0.64

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *School Emphasis on Academic Success—Principals' Reports* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item										
			ACBG14A	ACBG14B	ACBG14C	ACBG14D	ACBG14E	ACBG14F	ACBG14G	ACBG14H	ACBG14I	ACBG14J	ACBG14K
Norway (5)	0.85	41	0.61	0.54	0.72	0.64	0.79	0.71	0.67	0.75	0.55	0.43	0.54
Oman	0.90	50	0.62	0.67	0.73	0.70	0.73	0.80	0.65	0.77	0.72	0.71	0.62
Pakistan	0.89	48	0.60	0.70	0.70	0.52	0.69	0.73	0.61	0.79	0.76	0.73	0.75
Philippines	0.91	52	0.53	0.60	0.56	0.67	0.75	0.87	0.81	0.83	0.79	0.78	0.66
Poland	0.88	46	0.60	0.67	0.65	0.67	0.68	0.72	0.63	0.76	0.78	0.68	0.61
Portugal	0.88	47	0.58	0.67	0.72	0.61	0.72	0.78	0.67	0.78	0.72	0.67	0.60
Qatar	0.91	53	0.64	0.67	0.74	0.74	0.69	0.79	0.63	0.76	0.80	0.81	0.69
Russian Federation	0.87	43	0.62	0.72	0.63	0.61	0.73	0.69	0.51	0.71	0.70	0.72	0.58
Saudi Arabia	0.90	51	0.68	0.71	0.66	0.63	0.71	0.78	0.69	0.80	0.79	0.77	0.62
Serbia	0.86	42	0.66	0.75	0.66	0.67	0.60	0.74	0.47	0.64	0.65	0.62	0.60
Singapore	0.91	53	0.55	0.71	0.77	0.58	0.72	0.80	0.81	0.79	0.81	0.80	0.62
Slovak Republic	0.87	44	0.59	0.69	0.59	0.55	0.68	0.74	0.54	0.79	0.81	0.71	0.53
South Africa (5)	0.90	49	0.58	0.66	0.65	0.67	0.77	0.78	0.72	0.79	0.67	0.67	0.72
Spain	0.89	47	0.47	0.65	0.74	0.63	0.78	0.82	0.75	0.81	0.63	0.69	0.52
Sweden	0.91	54	0.67	0.76	0.76	0.65	0.76	0.81	0.76	0.81	0.72	0.80	0.52
Turkey (5)	0.89	49	0.62	0.69	0.72	0.60	0.78	0.81	0.63	0.78	0.70	0.74	0.57
United Arab Emirates	0.94	63	0.81	0.82	0.84	0.81	0.75	0.83	0.67	0.81	0.83	0.83	0.73
United States	0.92	55	0.61	0.69	0.66	0.68	0.80	0.86	0.76	0.85	0.79	0.74	0.70
Benchmarking Participants													
Ontario, Canada	0.91	53	0.58	0.67	0.69	0.69	0.80	0.81	0.70	0.84	0.74	0.78	0.69
Quebec, Canada	0.89	47	0.71	0.70	0.69	0.64	0.69	0.71	0.75	0.75	0.71	0.64	0.58
Moscow City, Russian Fed.	0.87	43	0.63	0.66	0.62	0.76	0.69	0.74	0.39	0.67	0.70	0.63	0.64
Madrid, Spain	0.89	48	0.43	0.68	0.79	0.50	0.78	0.84	0.75	0.81	0.70	0.67	0.57
Abu Dhabi, UAE	0.94	64	0.83	0.80	0.85	0.79	0.74	0.82	0.65	0.85	0.84	0.85	0.77
Dubai, UAE	0.94	62	0.82	0.80	0.84	0.80	0.78	0.84	0.72	0.82	0.78	0.80	0.67

Relationship Between the TIMSS 2019 *School Emphasis on Academic Success—Principals' Reports* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Albania	0.21	0.04	0.03	0.22	0.05	0.03
Armenia	0.06	0.00	0.00	0.04	0.00	0.00
Australia	0.22	0.05	0.04	0.21	0.04	0.04
Austria	0.21	0.04	0.03	0.26	0.07	0.06
Azerbaijan	-0.01	0.00	0.00	0.01	0.00	0.00
Bahrain	0.05	0.00	0.00	0.13	0.02	0.03
Belgium (Flemish)	0.16	0.03	0.01	0.16	0.02	0.01
Bosnia and Herzegovina	0.02	0.00	0.01	0.03	0.00	0.01
Bulgaria	0.39	0.15	0.12	0.42	0.18	0.15
Canada	0.14	0.02	0.02	0.13	0.02	0.01
Chile	0.24	0.06	0.05	0.22	0.05	0.04
Chinese Taipei	0.10	0.01	0.01	0.09	0.01	0.01
Croatia	0.06	0.00	0.00	0.07	0.00	0.01
Cyprus	0.16	0.03	0.02	0.14	0.02	0.02
Czech Republic	0.21	0.05	0.04	0.19	0.04	0.03
Denmark	0.10	0.01	0.01	0.09	0.01	0.01
England	0.21	0.05	0.05	0.25	0.06	0.07
Finland	0.11	0.01	0.01	0.13	0.02	0.01
France	0.22	0.05	0.03	0.23	0.05	0.04
Georgia	0.07	0.01	0.01	0.06	0.00	0.01
Germany	0.28	0.08	0.05	0.27	0.07	0.05
Hong Kong SAR	0.17	0.03	0.02	0.21	0.05	0.03
Hungary	0.29	0.08	0.04	0.26	0.07	0.03
Iran, Islamic Rep. of	0.20	0.04	0.02	0.22	0.05	0.03
Ireland	0.17	0.03	0.03	0.15	0.02	0.03
Italy	0.06	0.00	0.00	0.08	0.01	0.00
Japan	0.13	0.02	0.01	0.11	0.01	0.01
Kazakhstan	0.03	0.00	0.01	0.02	0.00	0.01
Korea, Rep. of	0.18	0.03	0.02	0.16	0.03	0.02
Kosovo	0.13	0.02	0.01	0.15	0.02	0.02
Kuwait	0.19	0.04	0.04	0.20	0.04	0.04
Latvia	0.11	0.01	0.01	0.12	0.01	0.01
Lithuania	0.19	0.04	0.01	0.19	0.04	0.01
Malta	0.23	0.05	0.04	0.26	0.07	0.05
Montenegro	0.02	0.00	0.00	0.03	0.00	0.00
Morocco	0.21	0.04	0.04	0.20	0.04	0.03
Netherlands	0.10	0.01	0.01	0.14	0.02	0.02

Relationship Between the TIMSS 2019 *School Emphasis on Academic Success—Principals' Reports* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
New Zealand	0.27	0.07	0.07	0.27	0.07	0.07
North Macedonia	0.21	0.04	0.02	0.23	0.05	0.03
Northern Ireland	0.18	0.03	0.03	0.14	0.02	0.02
Norway (5)	0.15	0.02	0.01	0.17	0.03	0.01
Oman	0.17	0.03	0.02	0.20	0.04	0.02
Pakistan	0.23	0.05	0.03	0.24	0.06	0.05
Philippines	0.09	0.01	0.01	0.10	0.01	0.01
Poland	0.14	0.02	0.01	0.14	0.02	0.01
Portugal	0.18	0.03	0.02	0.16	0.02	0.02
Qatar	0.08	0.01	0.00	0.09	0.01	0.01
Russian Federation	0.14	0.02	0.03	0.13	0.02	0.03
Saudi Arabia	0.14	0.02	0.02	0.18	0.03	0.03
Serbia	0.14	0.02	0.02	0.14	0.02	0.02
Singapore	0.15	0.02	0.02	0.16	0.03	0.02
Slovak Republic	0.31	0.10	0.04	0.35	0.12	0.04
South Africa (5)	0.26	0.07	0.06	0.26	0.07	0.06
Spain	0.21	0.04	0.03	0.19	0.04	0.03
Sweden	0.23	0.05	0.05	0.25	0.06	0.06
Turkey (5)	0.30	0.09	0.07	0.28	0.08	0.06
United Arab Emirates	0.37	0.14	0.13	0.40	0.16	0.16
United States	0.26	0.07	0.06	0.28	0.08	0.06
International Median	**0.17**	**0.03**	**0.02**	**0.16**	**0.03**	**0.03**
Benchmarking Participants						
Ontario, Canada	0.14	0.02	0.02	0.11	0.01	0.01
Quebec, Canada	0.14	0.02	0.02	0.13	0.02	0.02
Moscow City, Russian Fed.	0.04	0.00	0.00	0.05	0.00	0.00
Madrid, Spain	0.22	0.05	0.04	0.20	0.04	0.03
Abu Dhabi, UAE	0.39	0.15	0.17	0.43	0.18	0.20
Dubai, UAE	0.20	0.04	0.04	0.20	0.04	0.04

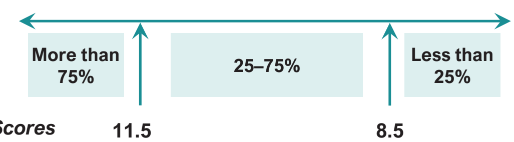
Schools Where Students Enter with Literacy and Numeracy Skills – Grade 4

About the Scale

The *Schools Where Students Enter with Literacy and Numeracy Skills* scale was created based on principals' responses to twelve items listed below.

About how many of the students in your school can do the following when they begin the first grade of primary/elementary school?

		More than 75%	51–75%	25–50%	Less than 25%
ACBG17A T	1) Recognize most of the letters of the alphabet	○	○	○	○
ACBG17B T	2) Read some words	○	○	○	○
ACBG17C T	3) Read sentences	○	○	○	○
ACBG17D T	4) Write letters of the alphabet	○	○	○	○
ACBG17E	5) Write their names	○	○	○	○
ACBG17F	6) Write words other than their names	○	○	○	○
ACBG17G T	7) Count up to 100 or higher	○	○	○	○
ACBG17H T	8) Recognize written numbers from 1-10	○	○	○	○
ACBG17I T	9) Recognize written numbers higher than 10	○	○	○	○
ACBG17J T	10) Write numbers from 1-10	○	○	○	○
ACBG17K T	11) Do simple addition	○	○	○	○
ACBG17L T	12) Do simple subtraction	○	○	○	○



T Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Schools Where Students Enter with Literacy and Numeracy Skills* Scale – Grade 4

Item	delta	tau_1	tau_2	tau_3	Infit
ACBG17A	-0.72531	-1.60124	0.01778	1.58346	1.28
ACBG17B	0.28400	-1.70182	-0.01923	1.72105	1.14
ACBG17C	1.50852	-1.39085	-0.20204	1.59289	1.12
ACBG17D	-0.58439	-1.95844	0.18365	1.77479	1.16
ACBG17E	-1.66731	-1.47535	0.11321	1.36214	1.61
ACBG17F	0.54604	-1.72628	-0.06684	1.79312	0.97
ACBG17G	1.07922	-1.89888	-0.16212	2.06100	1.46
ACBG17H	-1.89417	-2.02701	0.38834	1.63867	1.15
ACBG17I	0.07560	-1.89253	-0.09470	1.98723	1.08
ACBG17J	-0.88472	-1.53692	0.16155	1.37537	1.08
ACBG17K	0.86027	-1.92773	-0.05223	1.97996	1.03
ACBG17L	1.40225	-1.74952	-0.13408	1.88360	1.16

Scale Transformation Constants for the TIMSS 2019 *Schools Where Students Enter with Literacy and Numeracy Skills* Scale – Grade 4

Scale Transformation Constants	
A = 10.004848	Transformed Scale Score = 10.004848 + 0.6920950 • Logit Scale Score
B = 0.6920950	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Schools Where Students Enter with Literacy and Numeracy Skills* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	6.11923	
1	6.99979	
2	7.45729	
3	7.78264	
4	8.04034	
5	8.25669	
6	8.44586	8.5
7	8.61529	
8	8.77012	
9	8.91382	
10	9.04893	
11	9.17718	
12	9.30061	
13	9.42002	
14	9.53638	
15	9.65056	
16	9.76331	
17	9.87533	
18	9.98694	
19	10.09967	
20	10.21320	
21	10.32848	
22	10.44619	
23	10.56707	
24	10.69194	
25	10.82142	
26	10.95725	
27	11.10057	
28	11.25305	
29	11.41684	
30	11.59476	11.5
31	11.79087	
32	12.01080	
33	12.26650	
34	12.58081	
35	13.01404	
36	13.85372	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Schools Where Students Enter with Literacy and Numeracy Skills* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item											
			ACBG17A	ACBG17B	ACBG17C	ACBG17D	ACBG17E	ACBG17F	ACBG17G	ACBG17H	ACBG17I	ACBG17J	ACBG17K	ACBG17L
Albania	0.94	59	0.73	0.76	0.78	0.80	0.72	0.82	0.68	0.73	0.80	0.74	0.84	0.81
Armenia	0.95	65	0.72	0.85	0.83	0.85	0.83	0.88	0.76	0.70	0.77	0.79	0.84	0.84
Australia	0.97	76	0.86	0.91	0.90	0.87	0.73	0.91	0.85	0.85	0.91	0.91	0.85	0.86
Austria	0.91	52	0.82	0.83	0.75	0.66	0.41	0.77	0.67	0.59	0.75	0.75	0.80	0.76
Azerbaijan	0.97	72	0.76	0.88	0.82	0.88	0.88	0.88	0.75	0.82	0.86	0.85	0.90	0.91
Bahrain	0.97	73	0.82	0.90	0.87	0.88	0.88	0.90	0.83	0.79	0.90	0.82	0.84	0.82
Belgium (Flemish)	0.86	42	0.63	0.68	0.46	0.74	0.56	0.66	0.49	0.62	0.69	0.66	0.73	0.73
Bosnia and Herzegovina	0.94	61	0.78	0.83	0.77	0.81	0.73	0.84	0.74	0.71	0.82	0.80	0.74	0.75
Bulgaria	0.97	76	0.85	0.86	0.81	0.89	0.89	0.89	0.86	0.84	0.90	0.88	0.89	0.89
Canada	0.95	64	0.79	0.85	0.77	0.83	0.77	0.86	0.75	0.80	0.82	0.83	0.78	0.78
Chile	0.96	70	0.82	0.86	0.83	0.86	0.81	0.88	0.80	0.79	0.89	0.82	0.83	0.82
Chinese Taipei	0.96	71	0.88	0.85	0.81	0.88	0.82	0.85	0.82	0.82	0.85	0.80	0.86	0.87
Croatia	0.94	60	0.80	0.84	0.75	0.81	0.71	0.79	0.65	0.74	0.80	0.82	0.79	0.76
Cyprus	0.93	59	0.77	0.84	0.80	0.76	0.62	0.81	0.79	0.70	0.79	0.71	0.83	0.78
Czech Republic	0.88	45	0.73	0.74	0.58	0.68	0.58	0.67	0.55	0.72	0.69	0.76	0.67	0.62
Denmark	0.92	55	0.80	0.82	0.68	0.64	0.55	0.70	0.77	0.76	0.81	0.80	0.79	0.75
England	0.96	72	0.89	0.88	0.82	0.83	0.86	0.86	0.76	0.85	0.87	0.89	0.83	0.80
Finland	0.91	50	0.68	0.69	0.67	0.66	0.55	0.73	0.63	0.72	0.75	0.78	0.80	0.79
France	0.91	51	0.70	0.74	0.67	0.75	0.53	0.81	0.73	0.65	0.80	0.71	0.76	0.72
Georgia	0.97	74	0.78	0.83	0.90	0.91	0.86	0.94	0.84	0.73	0.91	0.83	0.86	0.87
Germany	0.87	43	0.68	0.71	0.62	0.67	0.56	0.61	0.52	0.69	0.70	0.73	0.68	0.69
Hong Kong SAR	0.95	64	0.77	0.87	0.82	0.83	0.81	0.83	0.78	0.73	0.78	0.74	0.83	0.80
Hungary	0.90	51	0.76	0.72	0.55	0.75	0.80	0.72	0.60	0.79	0.75	0.81	0.66	0.57
Iran, Islamic Rep. of	0.96	69	0.83	0.85	0.85	0.86	0.89	0.71	0.77	0.81	0.85	0.80	0.89	0.86
Ireland	0.88	54	0.84	0.83	0.84	0.82	0.63	0.82	0.41	0.76	0.62	0.76	0.82	0.54
Italy	0.93	57	0.72	0.78	0.79	0.71	0.52	0.82	0.72	0.80	0.83	0.79	0.79	0.75
Japan	0.94	62	0.70	0.77	0.85	0.81	0.68	0.84	0.74	0.80	0.84	0.84	0.80	0.78
Kazakhstan	0.95	63	0.63	0.77	0.75	0.76	0.86	0.87	0.78	0.77	0.82	0.78	0.88	0.85
Korea, Rep. of	0.96	73	0.84	0.85	0.88	0.90	0.72	0.89	0.86	0.80	0.88	0.86	0.88	0.87
Kosovo	0.95	64	0.77	0.84	0.79	0.84	0.76	0.85	0.74	0.71	0.81	0.81	0.84	0.83
Kuwait	0.97	77	0.86	0.90	0.87	0.90	0.86	0.92	0.85	0.82	0.92	0.87	0.90	0.85
Latvia	0.89	46	0.48	0.63	0.74	0.66	0.67	0.69	0.60	0.74	0.74	0.75	0.72	0.68
Lithuania	0.94	62	0.73	0.79	0.84	0.79	0.62	0.81	0.73	0.75	0.86	0.79	0.86	0.86
Malta	0.96	68	0.79	0.88	0.84	0.89	0.87	0.87	0.64	0.76	0.85	0.82	0.87	0.81
Montenegro	0.92	54	0.76	0.75	0.71	0.78	0.76	0.80	0.57	0.72	0.77	0.75	0.73	0.67
Morocco	0.97	77	0.91	0.93	0.89	0.92	0.91	0.90	0.84	0.89	0.90	0.87	0.86	0.68
Netherlands	0.93	57	0.79	0.73	0.79	0.70	0.57	0.79	0.72	0.76	0.77	0.79	0.83	0.79
New Zealand	0.95	66	0.85	0.87	0.81	0.85	0.74	0.88	0.79	0.85	0.84	0.85	0.75	0.69
North Macedonia	0.93	56	0.76	0.78	0.67	0.82	0.68	0.74	0.78	0.59	0.78	0.71	0.82	0.78
Northern Ireland	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Schools Where Students Enter with Literacy and Numeracy Skills* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item											
			ACBG17A	ACBG17B	ACBG17C	ACBG17D	ACBG17E	ACBG17F	ACBG17G	ACBG17H	ACBG17I	ACBG17J	ACBG17K	ACBG17L
Norway (5)	0.92	53	0.78	0.77	0.62	0.75	0.67	0.75	0.63	0.80	0.78	0.78	0.73	0.67
Oman	0.95	66	0.75	0.85	0.81	0.83	0.80	0.86	0.81	0.75	0.84	0.76	0.86	0.81
Pakistan	0.96	70	0.78	0.79	0.76	0.91	0.90	0.87	0.85	0.81	0.85	0.80	0.88	0.86
Philippines	0.96	70	0.72	0.88	0.83	0.84	0.77	0.82	0.88	0.81	0.87	0.85	0.90	0.90
Poland	0.93	58	0.74	0.74	0.72	0.82	0.77	0.82	0.63	0.69	0.78	0.85	0.80	0.79
Portugal	0.94	59	0.76	0.82	0.72	0.75	0.66	0.85	0.73	0.72	0.84	0.78	0.79	0.79
Qatar	0.98	79	0.89	0.91	0.89	0.91	0.91	0.93	0.88	0.85	0.89	0.87	0.89	0.87
Russian Federation	0.96	68	0.79	0.86	0.84	0.83	0.81	0.84	0.71	0.82	0.87	0.81	0.84	0.84
Saudi Arabia	0.96	72	0.82	0.89	0.81	0.89	0.87	0.90	0.83	0.78	0.86	0.84	0.86	0.86
Serbia	0.93	58	0.77	0.79	0.68	0.77	0.81	0.79	0.63	0.80	0.78	0.81	0.80	0.72
Singapore	0.95	71	0.85	0.88	0.80	0.93	0.90	0.87	0.59	0.87	0.86	0.90	0.83	0.79
Slovak Republic	0.91	52	0.71	0.80	0.63	0.74	0.69	0.79	0.64	0.72	0.67	0.75	0.76	0.70
South Africa (5)	0.94	60	0.75	0.80	0.69	0.79	0.76	0.83	0.70	0.79	0.81	0.79	0.78	0.80
Spain	0.90	51	0.69	0.85	0.82	0.73	0.60	0.87	0.62	0.64	0.63	0.67	0.73	0.64
Sweden	0.95	66	0.74	0.82	0.81	0.81	0.79	0.83	0.79	0.80	0.79	0.84	0.86	0.82
Turkey (5)	0.98	79	0.88	0.92	0.92	0.92	0.90	0.93	0.84	0.79	0.90	0.84	0.92	0.91
United Arab Emirates	0.98	80	0.89	0.93	0.88	0.92	0.90	0.92	0.89	0.85	0.91	0.89	0.89	0.88
United States	0.98	83	0.91	0.93	0.87	0.93	0.91	0.93	0.91	0.92	0.94	0.92	0.90	0.85
Benchmarking Participants														
Ontario, Canada	0.96	69	0.82	0.88	0.82	0.87	0.82	0.90	0.78	0.82	0.83	0.84	0.80	0.80
Quebec, Canada	0.93	58	0.78	0.81	0.71	0.79	0.70	0.78	0.70	0.75	0.80	0.80	0.74	0.73
Moscow City, Russian Fed	0.94	60	0.69	0.78	0.81	0.79	0.77	0.83	0.65	0.68	0.72	0.85	0.83	0.84
Madrid, Spain	0.92	56	0.61	0.79	0.82	0.73	0.59	0.86	0.59	0.72	0.82	0.79	0.82	0.78
Abu Dhabi, UAE	0.98	82	0.91	0.92	0.90	0.92	0.90	0.90	0.89	0.88	0.93	0.89	0.91	0.90
Dubai, UAE	0.98	84	0.95	0.95	0.89	0.94	0.89	0.94	0.90	0.88	0.92	0.92	0.90	0.88

A dash (–) indicates comparable data not available.

Relationship Between the TIMSS 2019 *Schools Where Students Enter with Literacy and Numeracy Skills* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Albania	0.06	0.00	0.01	0.07	0.00	0.01
Armenia	-0.01	0.00	0.01	-0.01	0.00	0.01
Australia	0.20	0.04	0.03	0.21	0.05	0.03
Austria	0.11	0.01	0.01	0.16	0.02	0.02
Azerbaijan	0.02	0.00	0.00	0.03	0.00	0.00
Bahrain	0.04	0.00	0.00	0.11	0.01	0.01
Belgium (Flemish)	0.14	0.02	0.01	0.16	0.03	0.01
Bosnia and Herzegovina	0.05	0.00	0.00	0.03	0.00	0.00
Bulgaria	0.39	0.15	0.12	0.44	0.19	0.15
Canada	0.10	0.01	0.01	0.10	0.01	0.01
Chile	0.22	0.05	0.05	0.19	0.04	0.04
Chinese Taipei	0.03	0.00	0.00	0.04	0.00	0.00
Croatia	0.07	0.01	0.00	0.10	0.01	0.00
Cyprus	0.10	0.01	0.01	0.11	0.01	0.01
Czech Republic	0.09	0.01	0.01	0.08	0.01	0.01
Denmark	0.06	0.00	0.00	0.05	0.00	0.00
England	0.21	0.04	0.01	0.24	0.06	0.02
Finland	0.08	0.01	0.01	0.08	0.01	0.01
France	0.17	0.03	0.01	0.16	0.02	0.01
Georgia	0.07	0.00	0.01	0.09	0.01	0.01
Germany	0.18	0.03	0.02	0.17	0.03	0.02
Hong Kong SAR	0.13	0.02	0.01	0.13	0.02	0.01
Hungary	0.15	0.02	0.01	0.12	0.01	0.01
Iran, Islamic Rep. of	0.03	0.00	0.00	0.05	0.00	0.00
Ireland	0.14	0.02	0.02	0.14	0.02	0.01
Italy	-0.03	0.00	0.00	-0.03	0.00	0.00
Japan	0.06	0.00	0.00	0.06	0.00	0.00
Kazakhstan	0.11	0.01	0.01	0.08	0.01	0.00
Korea, Rep. of	0.08	0.01	0.01	0.08	0.01	0.01
Kosovo	0.04	0.00	0.00	0.06	0.00	0.00
Kuwait	0.14	0.02	0.03	0.09	0.01	0.01
Latvia	0.02	0.00	0.00	0.02	0.00	0.00
Lithuania	0.12	0.01	0.01	0.11	0.01	0.01
Malta	0.14	0.02	0.01	0.14	0.02	0.02
Montenegro	0.01	0.00	0.00	0.02	0.00	0.00
Morocco	0.20	0.04	0.02	0.19	0.04	0.02
Netherlands	0.05	0.00	0.00	0.05	0.00	0.00

Relationship Between the TIMSS 2019 *Schools Where Students Enter with Literacy and Numeracy Skills* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
New Zealand	0.24	0.06	0.05	0.25	0.06	0.05
North Macedonia	0.11	0.01	0.02	0.07	0.01	0.02
Northern Ireland	-	-	-	-	-	-
Norway (5)	0.10	0.01	0.02	0.13	0.02	0.02
Oman	0.10	0.01	0.01	0.12	0.01	0.01
Pakistan	0.09	0.01	0.05	0.14	0.02	0.04
Philippines	0.13	0.02	0.01	0.16	0.02	0.01
Poland	0.03	0.00	0.00	0.05	0.00	0.00
Portugal	0.04	0.00	0.00	0.05	0.00	0.00
Qatar	0.18	0.03	0.03	0.18	0.03	0.03
Russian Federation	0.15	0.02	0.02	0.18	0.03	0.02
Saudi Arabia	0.10	0.01	0.01	0.11	0.01	0.02
Serbia	0.19	0.04	0.02	0.20	0.04	0.02
Singapore	0.10	0.01	0.00	0.11	0.01	0.00
Slovak Republic	0.31	0.10	0.07	0.35	0.12	0.08
South Africa (5)	0.01	0.00	0.04	0.00	0.00	0.05
Spain	0.20	0.04	0.04	0.20	0.04	0.03
Sweden	0.13	0.02	0.01	0.17	0.03	0.02
Turkey (5)	0.14	0.02	0.02	0.15	0.02	0.02
United Arab Emirates	0.30	0.09	0.09	0.30	0.09	0.09
United States	0.15	0.02	0.02	0.15	0.02	0.02
International Median	**0.10**	**0.01**	**0.01**	**0.11**	**0.01**	**0.01**
Benchmarking Participants						
Ontario, Canada	0.16	0.03	0.02	0.13	0.02	0.02
Quebec, Canada	0.03	0.00	0.00	0.05	0.00	0.00
Moscow City, Russian Fed.	0.02	0.00	0.00	0.02	0.00	0.00
Madrid, Spain	0.08	0.01	0.01	0.08	0.01	0.01
Abu Dhabi, UAE	0.32	0.10	0.10	0.35	0.12	0.13
Dubai, UAE	0.28	0.08	0.05	0.26	0.07	0.05

A dash (–) indicates comparable data not available.

Scales Based on Teachers' Reports

Classroom Teaching Limited by Students Not Ready for Instruction – Grade 4

About the Scale

The *Classroom Teaching Limited by Students Not Ready for Instruction* scale was created based on teachers' responses to eight items listed below.

In your view, to what extent do the following limit how you teach this class?

		Not at all	Some	A lot
ATBG13A T	1) Students lacking prerequisite knowledge or skills	○	○	○
ATBG13B T	2) Students suffering from lack of basic nutrition	○	○	○
ATBG13C T	3) Students suffering from not enough sleep	○	○	○
ATBG13D	4) Students absent from class	○	○	○
ATBG13E T	5) Disruptive students	○	○	○
ATBG13F T	6) Uninterested students	○	○	○
ATBG13G T	7) Students with mental, emotional, or psychological impairment	○	○	○
ATBG13H	8) Students with difficulties understanding the language of instruction	○	○	○

Very Little	Some	A Lot

Scale Cut Scores **10.8** **6.8**

[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Classroom Teaching Limited by Students Not Ready for Instruction* Scale – Grade 4

Item	delta	tau_1	tau_2	Infit
ATBG13A	0.91125	-2.07515	2.07515	1.04
ATBG13B	-1.09075	-1.07877	1.07877	1.08
ATBG13C	-0.28607	-1.55578	1.55578	0.96
ATBG13D	-0.02520	-1.59451	1.59451	1.00
ATBG13E	0.52548	-1.39807	1.39807	0.97
ATBG13F	0.51002	-1.82889	1.82889	0.94
ATBG13G	-0.04183	-1.36843	1.36843	1.02
ATBG13H	-0.50290	-1.21050	1.21050	1.05

Scale Transformation Constants for the TIMSS 2019 *Classroom Teaching Limited by Students Not Ready for Instruction* Scale – Grade 4

Scale Transformation Constants	
A = 8.854136	Transformed Scale Score = 8.854136 + 1.257770 • Logit Scale Score
B = 1.257770	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Classroom Teaching Limited by Students Not Ready for Instruction* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	3.29994	
1	4.83892	
2	5.65218	
3	6.26163	
4	6.78649	6.8
5	7.27293	
6	7.74710	
7	8.22648	
8	8.72059	
9	9.23472	
10	9.76704	
11	10.31572	
12	10.88870	10.8
13	11.50809	
14	12.22222	
15	13.14707	
16	14.80460	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Classroom Teaching Limited by Students Not Ready for Instruction* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			ATBG13A	ATBG13B	ATBG13C	ATBG13D	ATBG13E	ATBG13F	ATBG13G	ATBG13H
Albania	0.81	43	0.46	0.69	0.77	0.61	0.61	0.75	0.73	0.59
Armenia	0.80	42	0.39	0.61	0.71	0.70	0.67	0.72	0.66	0.65
Australia	0.81	43	0.66	0.71	0.72	0.69	0.69	0.62	0.61	0.51
Austria	0.75	37	0.74	0.18	0.71	0.61	0.51	0.60	0.64	0.66
Azerbaijan	0.71	34	0.52	0.41	0.64	0.65	0.58	0.58	0.63	0.63
Bahrain	0.76	38	0.55	0.50	0.64	0.55	0.62	0.74	0.66	0.62
Belgium (Flemish)	0.77	38	0.67	0.52	0.57	0.55	0.70	0.67	0.67	0.59
Bosnia and Herzegovina	0.82	45	0.47	0.66	0.71	0.68	0.69	0.71	0.68	0.72
Bulgaria	0.83	46	0.69	0.66	0.63	0.70	0.63	0.74	0.65	0.70
Canada	0.78	40	0.62	0.71	0.70	0.61	0.61	0.58	0.69	0.53
Chile	0.81	43	0.58	0.72	0.76	0.69	0.56	0.74	0.68	0.46
Chinese Taipei	0.82	45	0.53	0.60	0.65	0.68	0.70	0.72	0.71	0.74
Croatia	0.80	41	0.53	0.61	0.70	0.63	0.72	0.70	0.59	0.64
Cyprus	0.76	38	0.47	0.54	0.64	0.64	0.62	0.69	0.67	0.64
Czech Republic	0.68	32	0.61	0.44	0.64	0.49	0.52	0.70	0.59	0.48
Denmark	0.76	38	0.60	0.64	0.63	0.56	0.68	0.62	0.71	0.42
England	0.76	37	0.62	0.52	0.74	0.53	0.55	0.58	0.70	0.57
Finland	0.74	36	0.59	0.53	0.64	0.49	0.67	0.60	0.69	0.57
France	0.78	39	0.48	0.62	0.66	0.65	0.71	0.66	0.51	0.68
Georgia	0.71	34	0.57	0.49	0.58	0.59	0.66	0.60	0.54	0.59
Germany	0.78	41	0.72	0.62	0.67	0.57	0.55	0.67	0.67	0.63
Hong Kong SAR	0.76	38	0.66	0.54	0.61	0.64	0.57	0.64	0.66	0.62
Hungary	0.78	40	0.65	0.59	0.67	0.58	0.56	0.70	0.71	0.62
Iran, Islamic Rep. of	0.84	48	0.56	0.74	0.75	0.67	0.69	0.70	0.77	0.64
Ireland	0.79	40	0.61	0.61	0.64	0.72	0.67	0.66	0.54	0.60
Italy	0.82	44	0.70	0.58	0.64	0.74	0.69	0.65	0.64	0.67
Japan	0.76	38	0.67	0.29	0.50	0.47	0.69	0.77	0.75	0.66
Kazakhstan	0.87	53	0.53	0.56	0.75	0.71	0.80	0.78	0.81	0.81
Korea, Rep. of	0.87	53	0.50	0.70	0.81	0.80	0.66	0.73	0.80	0.76
Kosovo	0.65	30	0.62	0.61	0.69	0.46	0.32	0.61	0.54	0.43
Kuwait	0.75	37	0.55	0.57	0.67	0.46	0.70	0.68	0.58	0.62
Latvia	0.75	37	0.57	0.38	0.60	0.68	0.69	0.68	0.71	0.46
Lithuania	0.83	46	0.49	0.69	0.74	0.72	0.77	0.70	0.66	0.64
Malta	0.85	50	0.49	0.74	0.76	0.74	0.68	0.74	0.68	0.76
Montenegro	0.78	40	0.40	0.59	0.65	0.72	0.62	0.69	0.68	0.69
Morocco	0.73	35	0.57	0.61	0.56	0.45	0.65	0.66	0.53	0.64
Netherlands	0.79	41	0.49	0.61	0.75	0.67	0.76	0.63	0.52	0.61
New Zealand	0.76	37	0.46	0.65	0.68	0.61	0.70	0.66	0.62	0.46
North Macedonia	0.65	29	0.50	0.33	0.47	0.59	0.69	0.68	0.61	0.32
Northern Ireland	0.71	35	0.45	0.56	0.74	0.60	0.65	0.73	0.56	0.27

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Classroom Teaching Limited by Students Not Ready for Instruction* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			ATBG13A	ATBG13B	ATBG13C	ATBG13D	ATBG13E	ATBG13F	ATBG13G	ATBG13H
Norway (5)	0.81	43	0.62	0.65	0.74	0.67	0.72	0.76	0.62	0.42
Oman	0.84	46	0.53	0.65	0.67	0.72	0.70	0.75	0.69	0.71
Pakistan	0.66	29	0.57	0.70	0.52	0.34	0.48	0.54	0.63	0.45
Philippines	0.81	43	0.63	0.65	0.66	0.66	0.76	0.75	0.58	0.50
Poland	0.74	36	0.66	0.39	0.58	0.60	0.60	0.75	0.68	0.47
Portugal	0.83	47	0.47	0.78	0.78	0.76	0.74	0.74	0.54	0.59
Qatar	0.78	40	0.59	0.59	0.72	0.59	0.66	0.66	0.59	0.62
Russian Federation	0.86	52	0.48	0.64	0.73	0.81	0.76	0.79	0.78	0.70
Saudi Arabia	0.71	33	0.48	0.47	0.59	0.59	0.66	0.75	0.50	0.51
Serbia	0.81	44	0.40	0.62	0.69	0.71	0.72	0.65	0.75	0.71
Singapore	0.83	47	0.65	0.59	0.69	0.74	0.63	0.66	0.71	0.77
Slovak Republic	0.87	53	0.70	0.65	0.75	0.65	0.80	0.76	0.74	0.75
South Africa (5)	0.78	40	0.60	0.61	0.69	0.59	0.68	0.75	0.63	0.45
Spain	0.77	38	0.62	0.57	0.68	0.62	0.73	0.64	0.58	0.50
Sweden	0.68	32	0.53	0.08	0.51	0.55	0.72	0.64	0.67	0.58
Turkey (5)	0.82	44	0.55	0.62	0.69	0.62	0.74	0.72	0.69	0.67
United Arab Emirates	0.83	46	0.65	0.60	0.72	0.62	0.73	0.75	0.67	0.68
United States	0.79	42	0.64	0.68	0.73	0.65	0.67	0.67	0.66	0.41
Benchmarking Participants										
Ontario, Canada	0.82	45	0.62	0.78	0.73	0.63	0.66	0.59	0.71	0.63
Quebec, Canada	0.69	32	0.64	0.59	0.58	0.54	0.52	0.59	0.62	0.40
Moscow City, Russian Fed.	0.84	49	0.40	0.69	0.76	0.77	0.68	0.71	0.73	0.76
Madrid, Spain	0.77	39	0.68	0.57	0.69	0.57	0.63	0.67	0.59	0.56
Abu Dhabi, UAE	0.83	46	0.67	0.54	0.68	0.64	0.71	0.73	0.69	0.72
Dubai, UAE	0.82	45	0.68	0.66	0.71	0.60	0.71	0.74	0.63	0.61

Relationship Between the TIMSS 2019 *Classroom Teaching Limited by Students Not Ready for Instruction* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Albania	0.09	0.01	0.02	0.07	0.01	0.02
Armenia	0.05	0.00	0.01	0.04	0.00	0.00
Australia	0.28	0.08	0.06	0.26	0.07	0.06
Austria	0.22	0.05	0.03	0.25	0.06	0.04
Azerbaijan	0.01	0.00	0.00	-0.01	0.00	0.00
Bahrain	0.05	0.00	0.00	0.18	0.03	0.02
Belgium (Flemish)	0.16	0.03	0.02	0.17	0.03	0.03
Bosnia and Herzegovina	0.06	0.00	0.00	0.05	0.00	0.00
Bulgaria	0.23	0.05	0.07	0.27	0.07	0.08
Canada	0.22	0.05	0.04	0.19	0.04	0.03
Chile	0.18	0.03	0.04	0.20	0.04	0.04
Chinese Taipei	0.04	0.00	0.00	0.04	0.00	0.00
Croatia	-0.01	0.00	0.00	-0.01	0.00	0.00
Cyprus	0.14	0.02	0.02	0.12	0.02	0.01
Czech Republic	0.22	0.05	0.03	0.19	0.04	0.03
Denmark	0.09	0.01	0.01	0.07	0.01	0.00
England	0.09	0.01	0.01	0.10	0.01	0.02
Finland	0.14	0.02	0.02	0.15	0.02	0.03
France	0.21	0.04	0.04	0.20	0.04	0.04
Georgia	0.01	0.00	0.00	0.03	0.00	0.00
Germany	0.26	0.07	0.07	0.22	0.05	0.05
Hong Kong SAR	0.26	0.07	0.03	0.26	0.07	0.05
Hungary	0.21	0.05	0.03	0.20	0.04	0.02
Iran, Islamic Rep. of	0.16	0.03	0.02	0.14	0.02	0.01
Ireland	0.15	0.02	0.01	0.14	0.02	0.01
Italy	0.06	0.00	0.00	0.07	0.00	0.00
Japan	0.06	0.00	0.00	0.03	0.00	0.00
Kazakhstan	0.07	0.01	0.00	0.06	0.00	0.00
Korea, Rep. of	0.06	0.00	0.01	0.08	0.01	0.01
Kosovo	0.05	0.00	0.00	0.06	0.00	0.00
Kuwait	0.16	0.02	0.03	0.15	0.02	0.02
Latvia	0.14	0.02	0.01	0.05	0.00	0.01
Lithuania	0.13	0.02	0.02	0.14	0.02	0.02
Malta	0.18	0.03	0.04	0.18	0.03	0.04
Montenegro	0.10	0.01	0.01	0.10	0.01	0.01
Morocco	0.09	0.01	0.02	0.03	0.00	0.01
Netherlands	0.15	0.02	0.03	0.19	0.04	0.04

Relationship Between the TIMSS 2019 *Classroom Teaching Limited by Students Not Ready for Instruction* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
New Zealand	0.23	0.05	0.04	0.23	0.05	0.04
North Macedonia	0.14	0.02	0.02	0.13	0.02	0.02
Northern Ireland	0.17	0.03	0.02	0.15	0.02	0.02
Norway (5)	0.13	0.02	0.02	0.13	0.02	0.02
Oman	0.08	0.01	0.02	0.08	0.01	0.01
Pakistan	-0.01	0.00	0.01	-0.04	0.00	0.01
Philippines	0.15	0.02	0.02	0.19	0.04	0.04
Poland	0.16	0.03	0.02	0.08	0.01	0.01
Portugal	0.14	0.02	0.02	0.12	0.01	0.01
Qatar	0.05	0.00	0.01	0.10	0.01	0.01
Russian Federation	0.02	0.00	0.00	0.03	0.00	0.00
Saudi Arabia	0.12	0.02	0.01	0.07	0.00	0.00
Serbia	0.01	0.00	0.00	0.02	0.00	0.00
Singapore	0.40	0.16	0.16	0.40	0.16	0.12
Slovak Republic	0.21	0.04	0.03	0.28	0.08	0.07
South Africa (5)	0.12	0.01	0.02	0.12	0.01	0.01
Spain	0.18	0.03	0.02	0.17	0.03	0.02
Sweden	0.16	0.03	0.01	0.16	0.03	0.01
Turkey (5)	0.27	0.07	0.07	0.19	0.04	0.04
United Arab Emirates	0.27	0.07	0.08	0.29	0.08	0.07
United States	0.26	0.07	0.04	0.27	0.07	0.05
International Median	**0.14**	**0.02**	**0.02**	**0.13**	**0.02**	**0.02**
Benchmarking Participants						
Ontario, Canada	0.26	0.07	0.05	0.21	0.04	0.03
Quebec, Canada	0.12	0.02	0.01	0.13	0.02	0.01
Moscow City, Russian Fed.	0.09	0.01	0.01	0.09	0.01	0.01
Madrid, Spain	0.15	0.02	0.01	0.21	0.04	0.03
Abu Dhabi, UAE	0.28	0.08	0.08	0.29	0.09	0.06
Dubai, UAE	0.24	0.06	0.06	0.25	0.06	0.05

Safe and Orderly School – Grade 4

About the Scale

The *Safe and Orderly School* scale was created based on teachers' responses to eight items listed below.[1]

Thinking about your current school, indicate the extent to which you agree or disagree with each of the following statements.

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
ATBG07A [T]	1) This school is located in a safe neighborhood	○	○	○	○
ATBG07B [T]	2) I feel safe at this school	○	○	○	○
ATBG07C [T]	3) This school's security policies and practices are sufficient	○	○	○	○
ATBG07D [T]	4) The students behave in an orderly manner	○	○	○	○
ATBG07E [T]	5) The students are respectful of the teachers	○	○	○	○
ATBG07F [T]	6) The students respect school property	○	○	○	○
ATBG07G [T]	7) This school has clear rules about student conduct	○	○	○	○
ATBG07H [T]	8) This school's rules are enforced in a fair and consistent manner	○	○	○	○

Very Safe and Orderly | Somewhat Safe and Orderly | Less than Safe and Orderly

Scale Cut Scores 9.9 6.8

[1] For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Disagree a little" and "Disagree a lot" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Safe and Orderly School* Scale – Grade 4

Item	delta	tau_1	tau_2	Infit
ATBG07A	-0.57100	-1.06224	1.06224	1.32
ATBG07B	-1.45786	-1.30539	1.30539	0.99
ATBG07C	-0.57584	-1.41575	1.41575	1.12
ATBG07D	1.15251	-2.01386	2.01386	0.87
ATBG07E	0.69114	-1.95117	1.95117	0.87
ATBG07F	1.29829	-1.92798	1.92798	0.90
ATBG07G	-0.54040	-1.37123	1.37123	1.03
ATBG07H	0.00316	-1.51773	1.51773	1.00

Scale Transformation Constants for the TIMSS 2019 *Safe and Orderly School* Scale – Grade 4

Scale Transformation Constants	
A = 8.379152	Transformed Scale Score = 8.379152 + 0.972455 • Logit Scale Score
B = 0.972455	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Safe and Orderly School* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	3.88793	
1	5.11715	
2	5.77693	
3	6.27669	
4	6.70234	6.8
5	7.09225	
6	7.46639	
7	7.83877	
8	8.21647	
9	8.61043	
10	9.02958	
11	9.48469	
12	9.99265	9.9
13	10.56968	
14	11.22966	
15	12.02318	
16	13.35176	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Safe and Orderly School* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			ATBG07A	ATBG07B	ATBG07C	ATBG07D	ATBG07E	ATBG07F	ATBG07G	ATBG07H
Albania	0.76	41	0.59	0.66	0.43	0.73	0.74	0.71	0.68	0.50
Armenia	0.75	38	0.59	0.54	0.58	0.62	0.64	0.60	0.70	0.62
Australia	0.91	61	0.75	0.68	0.72	0.87	0.88	0.85	0.70	0.78
Austria	0.86	50	0.68	0.68	0.70	0.82	0.70	0.81	0.52	0.71
Azerbaijan	0.81	43	0.37	0.60	0.69	0.70	0.74	0.60	0.70	0.75
Bahrain	0.89	57	0.51	0.66	0.74	0.80	0.80	0.81	0.80	0.84
Belgium (Flemish)	0.83	47	0.52	0.66	0.68	0.73	0.69	0.75	0.69	0.75
Bosnia and Herzegovina	0.85	50	0.55	0.65	0.71	0.73	0.76	0.79	0.68	0.78
Bulgaria	0.85	50	0.67	0.71	0.75	0.73	0.69	0.75	0.69	0.67
Canada	0.89	57	0.59	0.69	0.68	0.84	0.84	0.82	0.75	0.78
Chile	0.88	54	0.70	0.69	0.67	0.82	0.75	0.79	0.76	0.72
Chinese Taipei	0.89	56	0.67	0.74	0.75	0.79	0.76	0.74	0.77	0.78
Croatia	0.89	56	0.67	0.76	0.75	0.79	0.82	0.79	0.62	0.77
Cyprus	0.85	49	0.51	0.66	0.72	0.76	0.78	0.75	0.66	0.72
Czech Republic	0.81	43	0.53	0.63	0.57	0.75	0.76	0.72	0.61	0.67
Denmark	0.88	55	0.59	0.69	0.67	0.80	0.80	0.80	0.74	0.80
England	0.84	48	0.49	0.50	0.55	0.75	0.78	0.75	0.82	0.80
Finland	0.84	49	0.54	0.66	0.54	0.83	0.81	0.72	0.67	0.78
France	0.85	49	0.72	0.69	0.66	0.81	0.77	0.75	0.56	0.58
Georgia	0.81	45	0.47	0.68	0.69	0.73	0.73	0.70	0.72	0.63
Germany	0.86	50	0.68	0.71	0.64	0.80	0.78	0.76	0.62	0.66
Hong Kong SAR	0.86	51	0.55	0.61	0.65	0.71	0.77	0.78	0.82	0.78
Hungary	0.86	51	0.57	0.65	0.74	0.78	0.72	0.77	0.70	0.74
Iran, Islamic Rep. of	0.86	50	0.66	0.67	0.67	0.75	0.68	0.76	0.78	0.67
Ireland	0.87	54	0.51	0.58	0.60	0.84	0.88	0.81	0.74	0.81
Italy	0.81	44	0.55	0.68	0.65	0.68	0.81	0.75	0.49	0.64
Japan	0.79	40	0.65	0.63	0.62	0.74	0.60	0.69	0.50	0.63
Kazakhstan	0.87	53	0.48	0.68	0.77	0.77	0.78	0.75	0.75	0.80
Korea, Rep. of	0.90	59	0.68	0.77	0.81	0.78	0.79	0.74	0.79	0.77
Kosovo	0.74	38	0.53	0.68	0.47	0.80	0.64	0.67	0.45	0.60
Kuwait	0.81	45	0.36	0.67	0.65	0.74	0.74	0.72	0.70	0.69
Latvia	0.79	41	0.37	0.60	0.62	0.67	0.73	0.74	0.65	0.69
Lithuania	0.83	46	0.49	0.59	0.70	0.78	0.78	0.76	0.56	0.72
Malta	0.85	50	0.36	0.68	0.66	0.80	0.81	0.79	0.72	0.74
Montenegro	0.86	51	0.63	0.68	0.72	0.78	0.79	0.74	0.65	0.72
Morocco	0.87	53	0.62	0.72	0.74	0.79	0.69	0.82	0.71	0.71
Netherlands	0.87	53	0.67	0.73	0.76	0.80	0.77	0.73	0.68	0.67
New Zealand	0.87	54	0.50	0.60	0.70	0.81	0.84	0.80	0.76	0.79
North Macedonia	0.82	46	0.60	0.65	0.60	0.74	0.74	0.68	0.65	0.77
Northern Ireland	0.88	55	0.63	0.52	0.50	0.83	0.85	0.85	0.81	0.81

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Safe and Orderly School* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			ATBG07A	ATBG07B	ATBG07C	ATBG07D	ATBG07E	ATBG07F	ATBG07G	ATBG07H
Norway (5)	0.86	52	0.57	0.71	0.63	0.80	0.81	0.82	0.66	0.70
Oman	0.81	45	0.42	0.57	0.67	0.76	0.74	0.73	0.68	0.71
Pakistan	0.84	48	0.69	0.66	0.63	0.75	0.69	0.63	0.71	0.74
Philippines	0.86	50	0.56	0.63	0.73	0.75	0.80	0.79	0.69	0.68
Poland	0.85	49	0.52	0.70	0.73	0.77	0.76	0.74	0.70	0.69
Portugal	0.82	44	0.67	0.68	0.66	0.77	0.76	0.74	0.50	0.48
Qatar	0.88	55	0.53	0.59	0.63	0.83	0.83	0.81	0.80	0.85
Russian Federation	0.83	47	0.58	0.69	0.60	0.70	0.77	0.77	0.54	0.78
Saudi Arabia	0.85	50	0.51	0.58	0.71	0.76	0.71	0.77	0.79	0.79
Serbia	0.86	51	0.58	0.69	0.75	0.75	0.72	0.69	0.69	0.81
Singapore	0.88	55	0.57	0.62	0.68	0.80	0.82	0.80	0.77	0.80
Slovak Republic	0.83	46	0.43	0.61	0.63	0.73	0.74	0.76	0.73	0.73
South Africa (5)	0.87	53	0.69	0.73	0.71	0.81	0.76	0.77	0.61	0.73
Spain	0.83	47	0.59	0.62	0.54	0.82	0.80	0.75	0.63	0.67
Sweden	0.84	48	0.63	0.71	0.74	0.80	0.80	0.68	0.56	0.62
Turkey (5)	0.89	58	0.73	0.76	0.77	0.78	0.79	0.79	0.72	0.73
United Arab Emirates	0.89	58	0.42	0.55	0.69	0.86	0.87	0.87	0.83	0.85
United States	0.89	58	0.64	0.68	0.67	0.85	0.85	0.85	0.73	0.78
Benchmarking Participants										
Ontario, Canada	0.91	62	0.61	0.75	0.74	0.87	0.86	0.85	0.78	0.81
Quebec, Canada	0.82	45	0.60	0.67	0.63	0.73	0.73	0.63	0.69	0.69
Moscow City, Russian Fed.	0.88	55	0.47	0.72	0.71	0.83	0.81	0.84	0.70	0.80
Madrid, Spain	0.86	51	0.58	0.64	0.67	0.82	0.83	0.75	0.71	0.71
Abu Dhabi, UAE	0.90	59	0.41	0.54	0.67	0.87	0.88	0.88	0.85	0.87
Dubai, UAE	0.88	55	0.40	0.52	0.71	0.82	0.84	0.85	0.83	0.84

Relationship Between the TIMSS 2019 *Safe and Orderly School* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		*r*	r^2	
Albania	0.01	0.00	0.00	-0.01	0.00	0.00
Armenia	0.09	0.01	0.01	0.04	0.00	0.00
Australia	0.18	0.03	0.03	0.20	0.04	0.04
Austria	0.18	0.03	0.03	0.22	0.05	0.05
Azerbaijan	-0.04	0.00	0.00	-0.04	0.00	0.00
Bahrain	0.03	0.00	0.00	0.21	0.04	0.02
Belgium (Flemish)	0.13	0.02	0.02	0.14	0.02	0.02
Bosnia and Herzegovina	0.04	0.00	0.00	0.06	0.00	0.00
Bulgaria	0.13	0.02	0.01	0.19	0.04	0.03
Canada	0.07	0.00	0.01	0.15	0.02	0.02
Chile	0.18	0.03	0.02	0.20	0.04	0.02
Chinese Taipei	0.01	0.00	0.00	0.04	0.00	0.00
Croatia	-0.01	0.00	0.00	0.01	0.00	0.00
Cyprus	0.02	0.00	0.00	0.05	0.00	0.00
Czech Republic	0.14	0.02	0.01	0.16	0.02	0.01
Denmark	0.11	0.01	0.01	0.10	0.01	0.01
England	0.12	0.01	0.01	0.13	0.02	0.01
Finland	0.12	0.01	0.01	0.10	0.01	0.01
France	0.19	0.03	0.02	0.18	0.03	0.02
Georgia	0.02	0.00	0.01	0.01	0.00	0.00
Germany	0.23	0.05	0.04	0.19	0.04	0.03
Hong Kong SAR	0.18	0.03	0.04	0.12	0.01	0.00
Hungary	0.17	0.03	0.03	0.10	0.01	0.02
Iran, Islamic Rep. of	0.09	0.01	0.00	0.09	0.01	0.00
Ireland	0.13	0.02	0.02	0.15	0.02	0.02
Italy	0.09	0.01	0.00	0.10	0.01	0.01
Japan	0.06	0.00	0.00	0.05	0.00	0.00
Kazakhstan	0.04	0.00	0.00	0.07	0.01	0.01
Korea, Rep. of	0.12	0.01	0.01	0.10	0.01	0.01
Kosovo	-0.02	0.00	0.00	-0.01	0.00	0.00
Kuwait	0.16	0.03	0.01	0.18	0.03	0.02
Latvia	0.08	0.01	0.02	0.05	0.00	0.00
Lithuania	0.09	0.01	0.00	0.12	0.01	0.01
Malta	0.13	0.02	0.01	0.12	0.01	0.01
Montenegro	0.00	0.00	0.00	-0.01	0.00	0.00
Morocco	0.26	0.07	0.04	0.23	0.05	0.03
Netherlands	0.12	0.01	0.01	0.17	0.03	0.02

Relationship Between the TIMSS 2019 *Safe and Orderly School* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
New Zealand	0.16	0.03	0.02	0.17	0.03	0.02
North Macedonia	0.03	0.00	0.01	0.02	0.00	0.01
Northern Ireland	0.13	0.02	0.02	0.12	0.01	0.02
Norway (5)	0.07	0.01	0.00	0.04	0.00	0.00
Oman	0.13	0.02	0.01	0.09	0.01	0.00
Pakistan	0.20	0.04	0.06	0.15	0.02	0.04
Philippines	0.15	0.02	0.02	0.14	0.02	0.00
Poland	0.02	0.00	0.00	-0.02	0.00	0.00
Portugal	0.16	0.03	0.03	0.15	0.02	0.02
Qatar	0.02	0.00	0.00	0.05	0.00	0.00
Russian Federation	0.03	0.00	0.00	-0.01	0.00	0.00
Saudi Arabia	0.11	0.01	0.02	0.14	0.02	0.03
Serbia	-0.08	0.01	0.01	-0.06	0.00	0.00
Singapore	0.06	0.00	0.00	0.03	0.00	0.00
Slovak Republic	0.09	0.01	0.01	0.11	0.01	0.01
South Africa (5)	0.12	0.01	0.01	0.08	0.01	0.01
Spain	0.18	0.03	0.04	0.17	0.03	0.03
Sweden	0.22	0.05	0.03	0.20	0.04	0.02
Turkey (5)	0.23	0.05	0.04	0.13	0.02	0.03
United Arab Emirates	0.29	0.08	0.07	0.34	0.11	0.09
United States	0.23	0.05	0.05	0.24	0.06	0.05
International Median	**0.12**	**0.01**	**0.01**	**0.12**	**0.01**	**0.01**
Benchmarking Participants						
Ontario, Canada	0.18	0.03	0.02	0.19	0.03	0.03
Quebec, Canada	0.04	0.00	0.00	0.07	0.00	0.00
Moscow City, Russian Fed.	0.04	0.00	0.00	0.02	0.00	0.00
Madrid, Spain	0.17	0.03	0.04	0.20	0.04	0.04
Abu Dhabi, UAE	0.37	0.14	0.12	0.37	0.14	0.11
Dubai, UAE	0.15	0.02	0.01	0.14	0.02	0.02

Teachers' Emphasis on Science Investigation – Grade 4

About the Scale

The *Teachers' Emphasis on Science Investigation* scale was created based on teachers' responses to eight items listed below.[1]

In teaching science to the students in this class, how often do you ask them to do the following?

		Every or almost every lesson	About half the lessons	Some lessons	Never
ATBS02B T	1) Observe natural phenomena such as the weather or a plant growing and describe what they see	○	○	○	○
ATBS02C T	2) Watch me demonstrate an experiment or investigation	○	○	○	○
ATBS02D T	3) Design or plan experiments or investigations	○	○	○	○
ATBS02E T	4) Conduct experiments or investigations	○	○	○	○
ATBS02F T	5) Present data from experiments or investigations	○	○	○	○
ATBS02G T	6) Interpret data from experiments or investigations	○	○	○	○
ATBS02H T	7) Use evidence from experiments or investigations to support conclusions	○	○	○	○
ATBS02K T	8) Do field work outside the class	○	○	○	○

About Half the Lessons or More | Less than Half the Lessons

Scale Cut Scores 11.3

[1] For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Some lessons" and "Never" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

T Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Teachers' Emphasis on Science Investigation* Scale – Grade 4

Item	delta	tau_1	tau_2	Infit
ATBS02B	-0.99302	-0.87415	0.87415	1.47
ATBS02C	-0.36179	-0.42328	0.42328	1.36
ATBS02D	0.14006	-0.74966	0.74966	0.83
ATBS02E	-0.03386	-0.74195	0.74195	0.82
ATBS02F	0.14169	-0.64943	0.64943	0.73
ATBS02G	-0.00064	-0.58904	0.58904	0.73
ATBS02H	-0.33626	-0.65045	0.65045	0.95
ATBS02K	1.44382	-0.54317	0.54317	1.53

Scale Transformation Constants for the TIMSS 2019 *Teachers' Emphasis on Science Investigation* Scale – Grade 4

Scale Transformation Constants	
A = 11.370579	Transformed Scale Score = 11.370579 + 1.2943720 • Logit Scale Score
B = 1.2943720	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Teachers' Emphasis on Science Investigation* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	6.59220	
1	8.14968	
2	8.93739	
3	9.49744	
4	9.94650	
5	10.33915	
6	10.69474	
7	11.03003	
8	11.35666	11.3
9	11.68739	
10	12.02985	
11	12.39638	
12	12.79933	
13	13.26245	
14	13.83022	
15	14.61161	
16	16.13766	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Teachers' Emphasis on Science Investigation* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			ATBS02B	ATBS02C	ATBS02D	ATBS02E	ATBS02F	ATBS02G	ATBS02H	ATBS02K
Albania	0.84	49	0.64	0.42	0.77	0.72	0.80	0.84	0.80	0.49
Armenia	0.86	55	0.56	0.68	0.84	0.79	0.87	0.85	0.72	0.55
Australia	0.88	56	0.63	0.60	0.82	0.87	0.88	0.88	0.72	0.50
Austria	0.86	55	0.54	0.73	0.81	0.88	0.87	0.86	0.66	0.46
Azerbaijan	0.86	51	0.61	0.66	0.77	0.76	0.83	0.82	0.72	0.48
Bahrain	0.85	50	0.54	0.47	0.71	0.85	0.85	0.84	0.80	0.41
Belgium (Flemish)	0.88	59	0.50	0.42	0.83	0.89	0.90	0.91	0.86	0.66
Bosnia and Herzegovina	0.90	63	0.54	0.71	0.89	0.89	0.89	0.87	0.82	0.66
Bulgaria	0.89	61	0.59	0.69	0.87	0.92	0.90	0.86	0.68	0.67
Canada	0.85	49	0.65	0.59	0.72	0.80	0.78	0.81	0.74	0.48
Chile	0.87	54	0.58	0.69	0.84	0.85	0.83	0.84	0.59	0.59
Chinese Taipei	0.88	56	0.68	0.60	0.74	0.80	0.86	0.87	0.86	0.49
Croatia	0.92	66	0.71	0.71	0.86	0.93	0.93	0.91	0.70	0.70
Cyprus	0.82	48	0.42	0.31	0.82	0.78	0.92	0.89	0.80	0.20
Czech Republic	0.80	49	0.55	0.65	0.84	0.83	0.82	0.78	0.57	0.46
Denmark	0.85	50	0.54	0.51	0.83	0.79	0.84	0.80	0.72	0.50
England	0.76	39	0.27	0.46	0.68	0.71	0.77	0.82	0.75	0.29
Finland	0.83	53	0.47	0.60	0.78	0.81	0.87	0.87	0.74	0.56
France	0.88	56	0.52	0.30	0.81	0.83	0.92	0.91	0.89	0.58
Georgia	0.87	56	0.54	0.60	0.86	0.91	0.88	0.89	0.72	0.42
Germany	0.84	49	0.62	0.54	0.67	0.85	0.80	0.83	0.76	0.43
Hong Kong SAR	0.91	62	0.67	0.70	0.79	0.86	0.86	0.87	0.88	0.63
Hungary	0.83	52	0.53	0.61	0.79	0.82	0.84	0.76	0.68	0.66
Iran, Islamic Rep. of	0.87	54	0.61	0.45	0.80	0.82	0.83	0.84	0.81	0.61
Ireland	0.83	47	0.44	0.45	0.65	0.82	0.85	0.83	0.83	0.43
Italy	0.90	60	0.64	0.67	0.90	0.89	0.87	0.87	0.78	0.50
Japan	0.83	47	0.65	0.46	0.74	0.81	0.77	0.80	0.71	0.45
Kazakhstan	0.90	58	0.62	0.71	0.82	0.77	0.85	0.83	0.72	0.73
Korea, Rep. of	0.81	47	0.45	0.44	0.75	0.84	0.85	0.86	0.70	0.35
Kosovo	0.85	51	0.42	0.58	0.77	0.79	0.86	0.84	0.76	0.58
Kuwait	0.88	57	0.40	0.63	0.85	0.88	0.88	0.88	0.81	0.51
Latvia	0.87	55	0.61	0.53	0.79	0.86	0.89	0.82	0.80	0.49
Lithuania	0.90	63	0.50	0.74	0.88	0.90	0.92	0.87	0.75	0.71
Malta	0.90	60	0.57	0.69	0.84	0.89	0.91	0.90	0.81	0.46
Montenegro	0.91	65	0.56	0.70	0.86	0.90	0.91	0.89	0.85	0.72
Morocco	0.85	49	0.52	0.51	0.78	0.81	0.88	0.81	0.74	0.37
Netherlands	0.88	57	0.62	0.48	0.91	0.92	0.85	0.86	0.81	0.34
New Zealand	0.86	52	0.64	0.51	0.75	0.82	0.85	0.84	0.76	0.51
North Macedonia	0.89	58	0.57	0.56	0.80	0.86	0.89	0.88	0.84	0.59
Northern Ireland	0.89	60	0.44	0.62	0.86	0.90	0.90	0.87	0.88	0.55

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Teachers' Emphasis on Science Investigation* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			ATBS02B	ATBS02C	ATBS02D	ATBS02E	ATBS02F	ATBS02G	ATBS02H	ATBS02K
Norway (5)	0.83	49	0.46	0.52	0.75	0.80	0.85	0.83	0.76	0.49
Oman	0.81	45	0.40	0.52	0.64	0.82	0.84	0.82	0.70	0.44
Pakistan	0.86	54	0.54	0.69	0.81	0.81	0.86	0.87	0.81	0.25
Philippines	0.89	58	0.55	0.62	0.80	0.86	0.87	0.87	0.86	0.55
Poland	0.83	51	0.54	0.63	0.67	0.74	0.85	0.81	0.82	0.56
Portugal	0.91	64	0.66	0.74	0.90	0.91	0.93	0.91	0.70	0.57
Qatar	0.88	55	0.62	0.54	0.79	0.85	0.88	0.87	0.77	0.47
Russian Federation	0.92	66	0.65	0.74	0.86	0.93	0.90	0.93	0.78	0.68
Saudi Arabia	0.86	51	0.52	0.53	0.74	0.77	0.85	0.89	0.83	0.44
Serbia	0.92	65	0.66	0.78	0.88	0.90	0.89	0.89	0.75	0.64
Singapore	0.81	45	0.54	0.41	0.65	0.76	0.84	0.82	0.71	0.48
Slovak Republic	0.89	60	0.59	0.57	0.88	0.90	0.91	0.90	0.78	0.53
South Africa (5)	0.91	63	0.64	0.75	0.82	0.86	0.87	0.90	0.83	0.63
Spain	0.89	59	0.57	0.66	0.81	0.87	0.85	0.87	0.82	0.66
Sweden	0.88	57	0.57	0.67	0.88	0.82	0.87	0.85	0.76	0.55
Turkey (5)	0.90	60	0.53	0.42	0.85	0.91	0.88	0.90	0.87	0.70
United Arab Emirates	0.91	61	0.60	0.66	0.84	0.89	0.91	0.90	0.79	0.58
United States	0.90	60	0.69	0.64	0.84	0.85	0.89	0.88	0.79	0.53
Benchmarking Participants										
Ontario, Canada	0.87	54	0.70	0.64	0.78	0.81	0.82	0.83	0.72	0.53
Quebec, Canada	0.81	44	0.59	0.61	0.62	0.73	0.75	0.76	0.74	0.43
Moscow City, Russian Fed.	0.89	61	0.55	0.70	0.90	0.91	0.91	0.91	0.67	0.57
Madrid, Spain	0.86	53	0.44	0.65	0.84	0.84	0.86	0.83	0.79	0.41
Abu Dhabi, UAE	0.91	63	0.67	0.74	0.85	0.89	0.89	0.88	0.79	0.59
Dubai, UAE	0.91	62	0.68	0.62	0.81	0.88	0.92	0.91	0.80	0.61

Relationship Between the TIMSS 2019 *Teachers' Emphasis on Science Investigation* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Albania	0.02	0.00	0.00
Armenia	0.06	0.00	0.00
Australia	-0.01	0.00	0.00
Austria	0.02	0.00	0.00
Azerbaijan	0.02	0.00	0.00
Bahrain	0.05	0.00	0.00
Belgium (Flemish)	0.02	0.00	0.00
Bosnia and Herzegovina	-0.02	0.00	0.00
Bulgaria	0.03	0.00	0.01
Canada	-0.02	0.00	0.00
Chile	0.06	0.00	0.00
Chinese Taipei	-0.02	0.00	0.00
Croatia	-0.01	0.00	0.00
Cyprus	-0.01	0.00	0.00
Czech Republic	-0.02	0.00	0.00
Denmark	0.01	0.00	0.00
England	-0.02	0.00	0.00
Finland	-0.01	0.00	0.00
France	0.02	0.00	0.00
Georgia	-0.04	0.00	0.00
Germany	-0.04	0.00	0.00
Hong Kong SAR	0.16	0.03	0.00
Hungary	-0.05	0.00	0.00
Iran, Islamic Rep. of	0.06	0.00	0.01
Ireland	-0.03	0.00	0.00
Italy	0.01	0.00	0.00
Japan	-0.02	0.00	0.00
Kazakhstan	-0.04	0.00	0.00
Korea, Rep. of	0.07	0.00	0.00
Kosovo	0.03	0.00	0.00
Kuwait	0.01	0.00	0.00
Latvia	0.01	0.00	0.00
Lithuania	-0.01	0.00	0.00
Malta	0.00	0.00	0.00
Montenegro	-0.01	0.00	0.00
Morocco	0.04	0.00	0.00
Netherlands	-0.01	0.00	0.00
New Zealand	-0.01	0.00	0.00

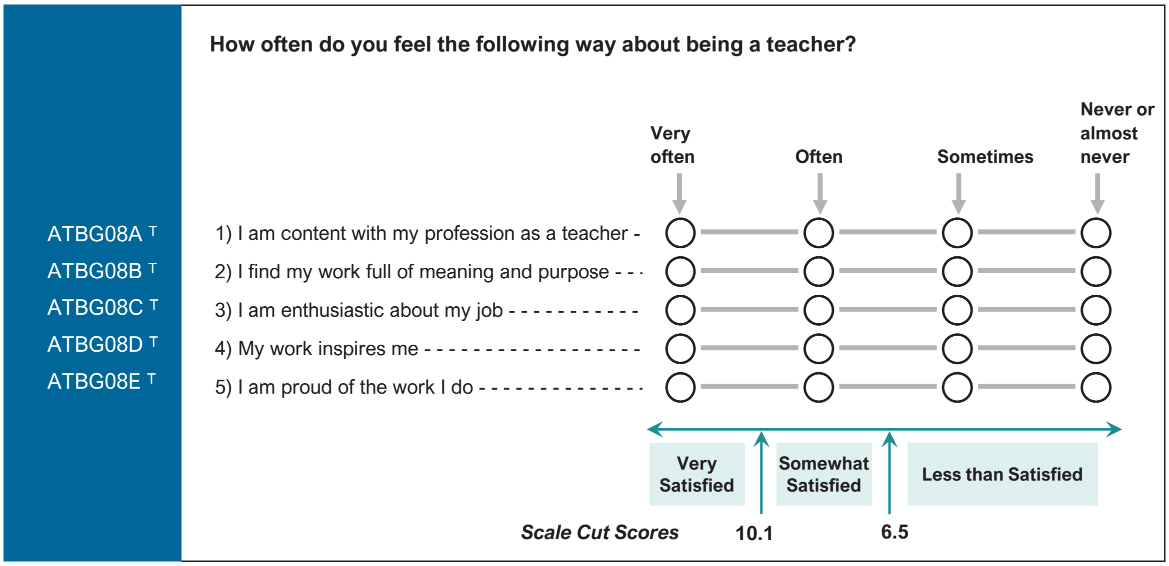
Relationship Between the TIMSS 2019 *Teachers' Emphasis on Science Investigation* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
North Macedonia	0.04	0.00	0.00
Northern Ireland	0.05	0.00	0.00
Norway (5)	0.03	0.00	0.00
Oman	-0.01	0.00	0.01
Pakistan	0.20	0.04	0.01
Philippines	-0.03	0.00	0.00
Poland	0.02	0.00	0.00
Portugal	-0.05	0.00	0.00
Qatar	-0.14	0.02	0.01
Russian Federation	0.01	0.00	0.00
Saudi Arabia	0.15	0.02	0.01
Serbia	0.05	0.00	0.00
Singapore	0.07	0.00	0.00
Slovak Republic	0.07	0.00	0.01
South Africa (5)	-0.08	0.01	0.00
Spain	-0.05	0.00	0.00
Sweden	0.03	0.00	0.00
Turkey (5)	-0.02	0.00	0.00
United Arab Emirates	0.14	0.02	0.02
United States	0.08	0.01	0.00
International Median	**0.01**	**0.00**	**0.00**
Benchmarking Participants			
Ontario, Canada	-0.03	0.00	0.00
Quebec, Canada	-0.03	0.00	0.00
Moscow City, Russian Fed.	0.03	0.00	0.00
Madrid, Spain	0.04	0.00	0.00
Abu Dhabi, UAE	0.10	0.01	0.01
Dubai, UAE	0.07	0.00	0.01

Teachers’ Job Satisfaction – Grade 4

About the Scale

The *Teachers' Job Satisfaction* scale was created based on students’ responses to five items listed below.[1]



[1] For the purpose of scaling, categories in which there were very few respondents were combined. The categories “Sometimes” and “Never or almost never” were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Teachers' Job Satisfaction* Scale – Grade 4

Item	delta	tau_1	tau_2	Infit
ATBG08A	0.26211	-2.33491	2.33491	1.09
ATBG08B	-0.67089	-2.27457	2.27457	1.10
ATBG08C	0.10245	-2.28986	2.28986	0.94
ATBG08D	0.53886	-2.12788	2.12788	0.91
ATBG08E	-0.23253	-1.97222	1.97222	1.10

Scale Transformation Constants for the TIMSS 2019 *Teachers' Job Satisfaction* Scale – Grade 4

Scale Transformation Constants	
A = 8.286639	Transformed Scale Score = 8.286639 + 0.732139 • Logit Scale Score
B = 0.732139	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Teachers' Job Satisfaction* Scale – Grade 4

Raw Score	Transformed Scale Score	Cutpoint
0	4.84518	
1	5.82502	
2	6.41362	6.5
3	6.92137	
4	7.47949	
5	8.28271	
6	9.09041	
7	9.65527	
8	10.16724	10.1
9	10.75160	
10	11.72115	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Teachers' Job Satisfaction* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item				
			ATBG08A	ATBG08B	ATBG08C	ATBG08D	ATBG08E
Albania	0.74	49	0.65	0.81	0.68	0.76	0.59
Armenia	0.77	53	0.69	0.66	0.73	0.78	0.77
Australia	0.92	76	0.87	0.88	0.90	0.91	0.79
Austria	0.87	66	0.79	0.77	0.87	0.85	0.78
Azerbaijan	0.79	56	0.77	0.72	0.81	0.71	0.74
Bahrain	0.86	65	0.83	0.72	0.87	0.84	0.78
Belgium (Flemish)	0.89	70	0.82	0.77	0.88	0.86	0.83
Bosnia and Herzegovina	0.86	65	0.77	0.78	0.86	0.83	0.78
Bulgaria	0.89	69	0.81	0.75	0.86	0.89	0.84
Canada	0.92	76	0.87	0.85	0.91	0.89	0.82
Chile	0.90	71	0.86	0.81	0.85	0.90	0.80
Chinese Taipei	0.93	78	0.88	0.90	0.91	0.85	0.88
Croatia	0.88	67	0.84	0.77	0.89	0.85	0.75
Cyprus	0.91	75	0.84	0.89	0.91	0.90	0.78
Czech Republic	0.91	73	0.84	0.88	0.88	0.89	0.78
Denmark	0.89	70	0.81	0.79	0.87	0.87	0.83
England	0.82	59	0.87	0.83	0.78	0.85	0.43
Finland	0.89	69	0.83	0.76	0.89	0.87	0.79
France	0.87	67	0.77	0.74	0.88	0.89	0.79
Georgia	0.77	53	0.70	0.57	0.77	0.81	0.75
Germany	0.86	65	0.77	0.72	0.85	0.83	0.85
Hong Kong SAR	0.93	79	0.87	0.90	0.91	0.89	0.88
Hungary	0.90	72	0.83	0.81	0.87	0.89	0.84
Iran, Islamic Rep. of	0.81	60	0.77	0.79	0.78	0.86	0.65
Ireland	0.92	75	0.84	0.89	0.87	0.89	0.83
Italy	0.85	64	0.76	0.74	0.84	0.85	0.80
Japan	0.90	72	0.84	0.86	0.84	0.87	0.81
Kazakhstan	0.86	64	0.72	0.80	0.82	0.83	0.81
Korea, Rep. of	0.93	79	0.87	0.92	0.87	0.91	0.88
Kosovo	0.58	38	0.48	0.68	0.70	0.62	0.59
Kuwait	0.85	63	0.83	0.77	0.87	0.80	0.69
Latvia	0.85	62	0.82	0.71	0.74	0.84	0.83
Lithuania	0.90	72	0.85	0.70	0.90	0.90	0.88
Malta	0.94	80	0.90	0.90	0.91	0.92	0.85
Montenegro	0.84	62	0.81	0.73	0.83	0.81	0.73
Morocco	0.90	71	0.86	0.77	0.89	0.87	0.81
Netherlands	0.87	67	0.77	0.71	0.88	0.84	0.86
New Zealand	0.91	74	0.83	0.85	0.89	0.91	0.82
North Macedonia	0.86	65	0.72	0.82	0.78	0.83	0.87
Northern Ireland	0.93	80	0.87	0.90	0.93	0.93	0.82

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Teachers' Job Satisfaction* Scale – Grade 4

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item: ATBG08A	ATBG08B	ATBG08C	ATBG08D	ATBG08E
Norway (5)	0.89	70	0.82	0.84	0.85	0.88	0.80
Oman	0.82	59	0.80	0.66	0.82	0.81	0.76
Pakistan	0.78	53	0.70	0.70	0.79	0.68	0.77
Philippines	0.91	73	0.82	0.89	0.85	0.87	0.83
Poland	0.91	74	0.86	0.84	0.86	0.87	0.85
Portugal	0.87	66	0.82	0.77	0.88	0.87	0.71
Qatar	0.93	78	0.90	0.86	0.88	0.90	0.88
Russian Federation	0.89	70	0.78	0.83	0.88	0.86	0.85
Saudi Arabia	0.82	58	0.77	0.68	0.82	0.82	0.71
Serbia	0.86	65	0.79	0.76	0.82	0.84	0.83
Singapore	0.95	83	0.90	0.91	0.93	0.93	0.90
Slovak Republic	0.91	73	0.79	0.85	0.87	0.88	0.87
South Africa (5)	0.90	70	0.85	0.84	0.86	0.85	0.79
Spain	0.86	64	0.78	0.78	0.80	0.81	0.84
Sweden	0.89	71	0.79	0.78	0.89	0.91	0.84
Turkey (5)	0.93	77	0.84	0.83	0.94	0.92	0.87
United Arab Emirates	0.91	74	0.83	0.87	0.89	0.89	0.82
United States	0.93	79	0.83	0.92	0.92	0.93	0.84
Benchmarking Participants							
Ontario, Canada	0.92	77	0.87	0.90	0.91	0.91	0.79
Quebec, Canada	0.91	75	0.87	0.80	0.90	0.89	0.84
Moscow City, Russian Fed.	0.90	72	0.80	0.82	0.86	0.91	0.83
Madrid, Spain	0.88	68	0.83	0.83	0.84	0.85	0.78
Abu Dhabi, UAE	0.92	77	0.85	0.89	0.90	0.91	0.83
Dubai, UAE	0.92	75	0.84	0.86	0.90	0.91	0.82

Relationship Between the TIMSS 2019 *Teachers' Job Satisfaction* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		*r*	r^2	
Albania	0.03	0.00	0.00	0.01	0.00	0.00
Armenia	0.06	0.00	0.00	0.07	0.01	0.00
Australia	0.00	0.00	0.00	-0.03	0.00	0.00
Austria	0.05	0.00	0.00	0.06	0.00	0.00
Azerbaijan	0.05	0.00	0.00	0.04	0.00	0.00
Bahrain	0.03	0.00	0.00	0.07	0.01	0.00
Belgium (Flemish)	0.00	0.00	0.00	0.00	0.00	0.00
Bosnia and Herzegovina	0.03	0.00	0.00	0.05	0.00	0.00
Bulgaria	0.15	0.02	0.02	0.18	0.03	0.03
Canada	-0.04	0.00	0.00	0.04	0.00	0.00
Chile	0.03	0.00	0.00	-0.04	0.00	0.00
Chinese Taipei	-0.01	0.00	0.00	-0.02	0.00	0.00
Croatia	0.01	0.00	0.00	0.03	0.00	0.00
Cyprus	0.03	0.00	0.00	0.07	0.01	0.00
Czech Republic	0.05	0.00	0.00	0.04	0.00	0.00
Denmark	0.01	0.00	0.00	0.03	0.00	0.00
England	0.06	0.00	0.01	0.08	0.01	0.01
Finland	0.02	0.00	0.00	0.02	0.00	0.00
France	0.02	0.00	0.00	0.04	0.00	0.00
Georgia	-0.07	0.01	0.00	-0.03	0.00	0.00
Germany	-0.03	0.00	0.00	-0.01	0.00	0.00
Hong Kong SAR	0.18	0.03	0.02	0.10	0.01	0.01
Hungary	0.05	0.00	0.00	0.01	0.00	0.00
Iran, Islamic Rep. of	0.05	0.00	0.00	0.04	0.00	0.00
Ireland	0.04	0.00	0.00	0.03	0.00	0.01
Italy	0.01	0.00	0.00	-0.01	0.00	0.00
Japan	-0.03	0.00	0.00	-0.03	0.00	0.00
Kazakhstan	0.02	0.00	0.00	0.01	0.00	0.00
Korea, Rep. of	0.04	0.00	0.00	0.03	0.00	0.00
Kosovo	0.04	0.00	0.00	0.04	0.00	0.00
Kuwait	0.04	0.00	0.00	0.06	0.00	0.01
Latvia	-0.02	0.00	0.00	-0.01	0.00	0.00
Lithuania	0.10	0.01	0.01	0.11	0.01	0.01
Malta	0.07	0.00	0.01	0.08	0.01	0.01
Montenegro	-0.03	0.00	0.00	-0.05	0.00	0.00
Morocco	0.20	0.04	0.03	0.19	0.03	0.04
Netherlands	-0.01	0.00	0.00	0.02	0.00	0.00

Relationship Between the TIMSS 2019 *Teachers' Job Satisfaction* Scale and TIMSS 2019 Achievement – Grade 4

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
New Zealand	0.09	0.01	0.01	0.08	0.01	0.01
North Macedonia	-0.01	0.00	0.00	-0.03	0.00	0.00
Northern Ireland	0.04	0.00	0.00	0.03	0.00	0.00
Norway (5)	0.03	0.00	0.00	0.04	0.00	0.00
Oman	0.06	0.00	0.01	0.02	0.00	0.01
Pakistan	0.15	0.02	0.00	0.12	0.01	0.01
Philippines	-0.04	0.00	0.00	-0.04	0.00	0.00
Poland	0.01	0.00	0.00	0.00	0.00	0.00
Portugal	0.05	0.00	0.00	0.06	0.00	0.00
Qatar	-0.02	0.00	0.00	-0.01	0.00	0.00
Russian Federation	-0.01	0.00	0.00	-0.02	0.00	0.00
Saudi Arabia	0.12	0.01	0.02	0.11	0.01	0.01
Serbia	-0.04	0.00	0.00	-0.04	0.00	0.00
Singapore	0.01	0.00	0.01	-0.04	0.00	0.00
Slovak Republic	0.10	0.01	0.01	0.01	0.00	0.01
South Africa (5)	-0.07	0.00	0.01	-0.04	0.00	0.00
Spain	0.03	0.00	0.00	0.05	0.00	0.00
Sweden	0.04	0.00	0.00	0.01	0.00	0.00
Turkey (5)	0.07	0.01	0.01	0.06	0.00	0.01
United Arab Emirates	0.16	0.03	0.03	0.16	0.03	0.03
United States	0.03	0.00	0.00	0.02	0.00	0.00
International Median	**0.03**	**0.00**	**0.00**	**0.03**	**0.00**	**0.00**
Benchmarking Participants						
Ontario, Canada	0.01	0.00	0.00	0.03	0.00	0.00
Quebec, Canada	-0.02	0.00	0.00	0.02	0.00	0.00
Moscow City, Russian Fed.	0.05	0.00	0.01	0.04	0.00	0.01
Madrid, Spain	0.08	0.01	0.01	0.15	0.02	0.03
Abu Dhabi, UAE	0.22	0.05	0.05	0.19	0.04	0.04
Dubai, UAE	0.06	0.00	0.00	0.07	0.00	0.01

Appendix 16B: TIMSS 2019 Context Questionnaire Scales—Grade 8

Scales Based on Students' Reports

Disorderly Behavior During Mathematics Lessons – Grade 8

About the Scale

The *Disorderly Behavior During Mathematics Lessons* scale was created based on students' responses to six items listed below.

How often do these things happen in your mathematics lessons?

		Never	Some lessons	About half the lessons	Every or almost every lesson
BSBM18A	1) Students don't listen to what the teacher says	○	○	○	○
BSBM18B	2) There is disruptive noise	○	○	○	○
BSBM18C	3) It is too disorderly for students to work well	○	○	○	○
BSBM18D	4) My teacher has to wait a long time for students to quiet down	○	○	○	○
BSBM18E	5) Students interrupt the teacher	○	○	○	○
BSBM18F	6) My teacher has to keep telling us to follow the classroom rules	○	○	○	○

	Few or No Lessons		Some Lessons		Most Lessons
Scale Cut Scores		11.6		7.9	

Item Parameters for the TIMSS 2019 *Disorderly Behavior During Mathematics Lessons* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBM18A	0.37476	-1.32069	-0.62479	1.94548	1.04
BSBM18B	0.31283	-1.02077	-0.64913	1.66990	0.89
BSBM18C	-0.46280	-0.91951	-0.27631	1.19582	0.95
BSBM18D	-0.13515	-0.67454	-0.57593	1.25047	0.88
BSBM18E	-0.23701	-0.67678	-0.61471	1.29149	0.90
BSBM18F	0.14737	-0.45870	-0.63749	1.09619	1.28

Scale Transformation Constants for the TIMSS 2019 *Disorderly Behavior During Mathematics Lessons* Scale – Grade 8

Scale Transformation Constants	
A = 9.663799	Transformed Scale Score = 9.663799 + 1.213350 • Logit Scale Score
B = 1.213350	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Disorderly Behavior During Mathematics Lessons* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	5.56633	
1	6.82233	
2	7.41866	
3	7.82995	7.9
4	8.15891	
5	8.44435	
6	8.70559	
7	8.95449	
8	9.19981	
9	9.45061	
10	9.71101	
11	9.99445	
12	10.31153	
13	10.67499	
14	11.10344	
15	11.61733	11.6
16	12.25252	
17	13.10820	
18	14.67370	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Disorderly Behavior During Mathematics Lessons* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item					
			BSBM18A	BSBM18B	BSBM18C	BSBM18D	BSBM18E	BSBM18F
Australia	0.92	71	0.83	0.84	0.82	0.87	0.87	0.82
Bahrain	0.84	57	0.76	0.81	0.76	0.79	0.80	0.57
Chile	0.90	67	0.76	0.82	0.81	0.85	0.87	0.78
Chinese Taipei	0.90	67	0.77	0.83	0.88	0.85	0.85	0.73
Cyprus	0.87	62	0.78	0.84	0.73	0.84	0.83	0.68
Egypt	0.77	48	0.73	0.78	0.77	0.72	0.76	0.26
England	0.92	73	0.83	0.86	0.85	0.87	0.87	0.84
Finland	0.91	70	0.78	0.84	0.87	0.88	0.86	0.79
France	0.89	65	0.76	0.82	0.84	0.83	0.79	0.80
Georgia	0.86	59	0.70	0.80	0.65	0.83	0.82	0.77
Hong Kong SAR	0.93	74	0.82	0.85	0.90	0.90	0.85	0.83
Hungary	0.90	67	0.78	0.84	0.76	0.87	0.81	0.85
Iran, Islamic Rep. of	0.86	59	0.71	0.79	0.83	0.82	0.78	0.68
Ireland	0.91	69	0.79	0.86	0.83	0.86	0.85	0.79
Israel	0.90	66	0.79	0.85	0.85	0.86	0.82	0.70
Italy	0.90	67	0.74	0.87	0.87	0.85	0.82	0.76
Japan	0.88	63	0.75	0.79	0.82	0.77	0.80	0.82
Jordan	0.83	55	0.77	0.79	0.76	0.78	0.78	0.54
Kazakhstan	0.81	54	0.72	0.82	0.77	0.76	0.79	0.50
Korea, Rep. of	0.90	67	0.83	0.86	0.88	0.83	0.87	0.60
Kuwait	0.81	53	0.75	0.82	0.77	0.77	0.78	0.40
Lebanon	0.83	54	0.70	0.78	0.73	0.78	0.78	0.62
Lithuania	0.90	67	0.81	0.86	0.83	0.86	0.80	0.75
Malaysia	0.73	44	0.68	0.75	0.75	0.71	0.69	0.28
Morocco	0.78	49	0.68	0.76	0.73	0.75	0.73	0.50
New Zealand	0.91	69	0.82	0.83	0.83	0.86	0.87	0.79
Norway (9)	0.91	69	0.77	0.85	0.83	0.87	0.85	0.82
Oman	0.82	53	0.69	0.77	0.76	0.77	0.75	0.61
Portugal	0.92	72	0.75	0.87	0.88	0.88	0.85	0.87
Qatar	0.88	62	0.77	0.82	0.77	0.82	0.81	0.71
Romania	0.87	61	0.68	0.84	0.79	0.83	0.82	0.71
Russian Federation	0.91	70	0.79	0.87	0.88	0.87	0.82	0.80
Saudi Arabia	0.82	54	0.76	0.80	0.72	0.76	0.79	0.52
Singapore	-	-	-	-	-	-	-	-
South Africa (9)	0.80	51	0.69	0.75	0.69	0.76	0.79	0.58
Sweden	0.90	66	0.77	0.84	0.83	0.86	0.83	0.75
Turkey	0.87	61	0.77	0.83	0.74	0.79	0.77	0.77
United Arab Emirates	0.89	65	0.80	0.84	0.81	0.83	0.83	0.72
United States	0.92	70	0.81	0.84	0.82	0.87	0.86	0.83

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Disorderly Behavior During Mathematics Lessons* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item					
			BSBM18A	BSBM18B	BSBM18C	BSBM18D	BSBM18E	BSBM18F
Benchmarking Participants								
Ontario, Canada	0.91	70	0.82	0.84	0.80	0.87	0.86	0.82
Quebec, Canada	0.90	66	0.78	0.78	0.84	0.85	0.82	0.80
Moscow City, Russian Fed.	0.91	70	0.81	0.87	0.88	0.86	0.80	0.79
Gauteng, RSA (9)	0.85	57	0.73	0.80	0.73	0.79	0.81	0.66
Western Cape, RSA (9)	0.88	62	0.72	0.81	0.80	0.83	0.82	0.74
Abu Dhabi, UAE	0.89	63	0.80	0.83	0.79	0.82	0.83	0.70
Dubai, UAE	0.89	65	0.79	0.82	0.80	0.83	0.83	0.77

A dash (–) indicates comparable data not available.

Relationship Between the TIMSS 2019 *Disorderly Behavior During Mathematics Lessons* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Australia	0.23	0.05	0.05
Bahrain	0.07	0.01	0.00
Chile	0.07	0.00	0.00
Chinese Taipei	0.02	0.00	0.00
Cyprus	0.13	0.02	0.01
Egypt	0.13	0.02	0.01
England	0.28	0.08	0.06
Finland	0.04	0.00	0.00
France	0.06	0.00	0.00
Georgia	0.10	0.01	0.01
Hong Kong SAR	0.04	0.00	0.01
Hungary	0.17	0.03	0.02
Iran, Islamic Rep. of	0.12	0.01	0.01
Ireland	0.29	0.08	0.07
Israel	0.23	0.05	0.04
Italy	0.12	0.01	0.01
Japan	0.08	0.01	0.01
Jordan	0.16	0.02	0.02
Kazakhstan	0.10	0.01	0.01
Korea, Rep. of	-0.07	0.01	0.00
Kuwait	0.07	0.01	0.01
Lebanon	0.07	0.00	0.00
Lithuania	0.08	0.01	0.01
Malaysia	0.27	0.07	0.03
Morocco	0.10	0.01	0.01
New Zealand	0.17	0.03	0.02
Norway (9)	0.05	0.00	0.00
Oman	0.08	0.01	0.00
Portugal	0.04	0.00	0.00
Qatar	0.16	0.03	0.02
Romania	0.21	0.04	0.04
Russian Federation	0.12	0.01	0.01
Saudi Arabia	0.08	0.01	0.01
Singapore	-	-	-
South Africa (9)	0.14	0.02	0.02
Sweden	0.04	0.00	0.00
Turkey	0.10	0.01	0.01
United Arab Emirates	0.19	0.04	0.03
United States	0.25	0.06	0.05
International Median	**0.10**	**0.01**	**0.01**

Relationship Between the TIMSS 2019 *Disorderly Behavior During Mathematics Lessons* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Benchmarking Participants			
Ontario, Canada	0.18	0.03	0.03
Quebec, Canada	0.12	0.02	0.02
Moscow City, Russian Fed.	0.13	0.02	0.01
Gauteng, RSA (9)	0.25	0.06	0.05
Western Cape, RSA (9)	0.28	0.08	0.07
Abu Dhabi, UAE	0.19	0.04	0.03
Dubai, UAE	0.19	0.04	0.03

A dash (–) indicates comparable data not available.

Home Educational Resources – Grade 8

About the Scale

The *Home Educational Resources* scale was created based on students' reports regarding the availability of three resources listed below.

BSBG04 [T]

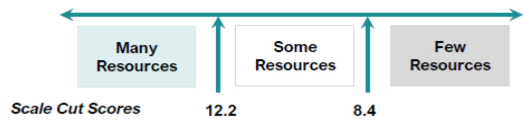
Number of books in the home:
1) 0-10
2) 11-25
3) 26-100
4) 101-200
5) More than 200

BSDG05S [T 1]

Number of home study supports:
1) None
2) Internet connection or own room
3) Both internet connection and own room

BSDGEDUP [T 1]

Highest level of education of either parent:
1) Finished some primary or lower secondary or did not go to school
2) Finished lower secondary
3) Finished upper secondary
4) Finished post-secondary education
5) Finished university or higher



[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.
[1] Derived variable. For details, see Supplement 3 of the *TIMSS 2019 User Guide for the International Database* .

Item Parameters for the TIMSS 2019 *Home Educational Resources* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	tau_4	Infit
BSBG04	1.03579	-0.89014	-0.25434	0.71392	0.43056	1.00
BSDG05S	-0.94808	-0.62039	0.62039			0.95
BSDGEDUP	-0.08771	-0.72810	-0.53330	0.92403	0.33737	0.99

Scale Transformation Constants for the TIMSS 2019 *Home Educational Resources* Scale – Grade 8

Scale Transformation Constants	
A = 9.168781	Transformed Scale Score = 9.168781 + 1.639257 • Logit Scale Score
B = 1.639257	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Home Educational Resources* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	4.55208	
1	6.52207	
2	7.56439	
3	8.35810	8.4
4	9.03613	
5	9.64761	
6	10.23835	
7	10.84499	
8	11.48712	
9	12.25655	12.2
10	13.51543	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Home Educational Resources* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item		
			BSBG04	BSDG05S	BSDGEDUP
Australia	0.40	46	0.77	0.44	0.78
Bahrain	0.40	45	0.72	0.50	0.77
Chile	0.41	46	0.71	0.55	0.76
Chinese Taipei	0.48	49	0.81	0.43	0.80
Cyprus	0.47	49	0.80	0.40	0.81
Egypt	0.35	44	0.57	0.68	0.73
England	0.44	48	0.77	0.47	0.79
Finland	0.36	44	0.76	0.35	0.79
France	0.46	49	0.79	0.44	0.80
Georgia	0.48	50	0.82	0.36	0.83
Hong Kong SAR	0.49	50	0.76	0.53	0.79
Hungary	0.51	51	0.86	0.24	0.86
Iran, Islamic Rep. of	0.55	53	0.73	0.65	0.80
Ireland	0.42	47	0.79	0.34	0.81
Israel	0.43	64	0.80	-	0.80
Italy	0.44	47	0.81	0.29	0.83
Japan	0.25	41	0.76	0.22	0.77
Jordan	0.40	46	0.63	0.64	0.76
Kazakhstan	0.40	46	0.71	0.60	0.72
Korea, Rep. of	0.41	47	0.75	0.50	0.78
Kuwait	0.30	41	0.73	0.41	0.73
Lebanon	0.39	45	0.73	0.44	0.79
Lithuania	0.44	49	0.79	0.45	0.80
Malaysia	0.48	49	0.73	0.57	0.78
Morocco	0.56	54	0.75	0.66	0.78
New Zealand	0.44	48	0.79	0.46	0.78
Norway (9)	0.37	46	0.76	0.46	0.78
Oman	0.41	46	0.70	0.51	0.79
Portugal	0.48	49	0.84	0.23	0.85
Qatar	0.38	45	0.72	0.47	0.78
Romania	0.54	53	0.81	0.47	0.84
Russian Federation	0.41	46	0.77	0.40	0.79
Saudi Arabia	0.37	44	0.68	0.51	0.77
Singapore	0.43	47	0.73	0.51	0.79
South Africa (9)	0.42	47	0.65	0.69	0.71
Sweden	0.41	47	0.80	0.42	0.77
Turkey	0.64	59	0.78	0.72	0.81
United Arab Emirates	0.35	44	0.72	0.54	0.72
United States	0.45	49	0.74	0.54	0.79

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Home Educational Resources* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item		
			BSBG04	BSDG05S	BSDGEDUP
Benchmarking Participants					
Ontario, Canada	0.36	45	0.74	0.44	0.78
Quebec, Canada	0.35	44	0.79	0.27	0.80
Moscow City, Russian Fed.	0.29	42	0.71	0.44	0.76
Gauteng, RSA (9)	0.44	48	0.62	0.72	0.73
Western Cape, RSA (9)	0.59	56	0.75	0.75	0.75
Abu Dhabi, UAE	0.35	44	0.68	0.54	0.74
Dubai, UAE	0.37	45	0.74	0.58	0.68

A dash (–) indicates comparable data not available.

Relationship Between the TIMSS 2019 *Home Educational Resources* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Australia	0.41	0.17	0.12	0.46	0.22	0.14
Bahrain	0.15	0.02	0.01	0.28	0.08	0.05
Chile	0.40	0.16	0.11	0.38	0.15	0.10
Chinese Taipei	0.41	0.16	0.13	0.40	0.16	0.13
Cyprus	0.43	0.19	0.12	0.42	0.18	0.11
Egypt	0.18	0.03	0.03	0.18	0.03	0.03
England	0.41	0.17	0.10	0.44	0.20	0.12
Finland	0.37	0.14	0.09	0.39	0.15	0.10
France	0.46	0.22	0.13	0.49	0.24	0.14
Georgia	0.26	0.07	0.04	0.26	0.07	0.04
Hong Kong SAR	0.31	0.09	0.06	0.28	0.08	0.05
Hungary	0.56	0.32	0.24	0.54	0.30	0.22
Iran, Islamic Rep. of	0.44	0.19	0.14	0.44	0.19	0.15
Ireland	0.47	0.22	0.15	0.48	0.23	0.16
Israel	0.43	0.18	0.08	0.41	0.17	0.07
Italy	0.36	0.13	0.10	0.40	0.16	0.12
Japan	0.35	0.12	0.07	0.35	0.12	0.07
Jordan	0.22	0.05	0.03	0.24	0.06	0.04
Kazakhstan	0.25	0.06	0.03	0.25	0.06	0.03
Korea, Rep. of	0.36	0.13	0.11	0.31	0.09	0.08
Kuwait	0.17	0.03	0.02	0.17	0.03	0.02
Lebanon	0.34	0.11	0.07	0.38	0.15	0.10
Lithuania	0.46	0.21	0.12	0.44	0.19	0.12
Malaysia	0.41	0.16	0.10	0.38	0.15	0.09
Morocco	0.25	0.06	0.07	0.15	0.02	0.03
New Zealand	0.44	0.19	0.11	0.50	0.25	0.16
Norway (9)	0.35	0.12	0.09	0.39	0.15	0.10
Oman	0.32	0.10	0.06	0.30	0.09	0.05
Portugal	0.42	0.18	0.12	0.39	0.15	0.11
Qatar	0.37	0.14	0.08	0.34	0.11	0.06
Romania	0.47	0.22	0.16	0.42	0.18	0.13
Russian Federation	0.30	0.09	0.05	0.31	0.09	0.04
Saudi Arabia	0.25	0.06	0.05	0.23	0.05	0.04
Singapore	0.40	0.16	0.10	0.44	0.19	0.12
South Africa (9)	0.26	0.07	0.06	0.28	0.08	0.06
Sweden	0.46	0.22	0.14	0.50	0.25	0.16
Turkey	0.48	0.23	0.18	0.47	0.22	0.17
United Arab Emirates	0.30	0.09	0.05	0.31	0.10	0.06
United States	0.46	0.21	0.13	0.45	0.20	0.13
International Median	**0.37**	**0.14**	**0.10**	**0.39**	**0.15**	**0.10**

Relationship Between the TIMSS 2019 *Home Educational Resources* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Benchmarking Participants						
Ontario, Canada	0.37	0.14	0.08	0.41	0.17	0.11
Quebec, Canada	0.35	0.12	0.07	0.40	0.16	0.10
Moscow City, Russian Fed.	0.29	0.08	0.05	0.29	0.08	0.05
Gauteng, RSA (9)	0.30	0.09	0.06	0.32	0.10	0.06
Western Cape, RSA (9)	0.47	0.22	0.18	0.47	0.22	0.19
Abu Dhabi, UAE	0.30	0.09	0.06	0.32	0.10	0.06
Dubai, UAE	0.31	0.10	0.06	0.32	0.10	0.06

Instructional Clarity in Biology Lessons – Grade 8

About the Scale

The *Instructional Clarity in Biology Lessons* scale was created based on students' responses to seven items listed below.

How much do you agree with these statements about your biology lessons?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BSBB24A	1) I know what my teacher expects me to do	○	○	○	○
BSBB24B	2) My teacher is easy to understand	○	○	○	○
BSBB24C	3) My teacher has clear answers to my questions	○	○	○	○
BSBB24D	4) My teacher is good at explaining biology	○	○	○	○
BSBB24E	5) My teacher does a variety of things to help us learn	○	○	○	○
BSBB24F	6) My teacher links new lessons to what I already know	○	○	○	○
BSBB24G	7) My teacher explains a topic again when we don't understand	○	○	○	○

	High Clarity		Moderate Clarity		Low Clarity
Scale Cut Scores		10.3		7.7	

Item Parameters for the TIMSS 2019 *Instructional Clarity in Biology Lessons* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBB24A	0.39375	-1.92236	-0.50414	2.42650	1.46
BSBB24B	0.12682	-1.69599	-0.52211	2.21810	0.90
BSBB24C	-0.12835	-1.55327	-0.39952	1.95279	0.82
BSBB24D	-0.42182	-1.28981	-0.57979	1.86960	0.74
BSBB24E	-0.02631	-1.65061	-0.36700	2.01761	0.92
BSBB24F	0.26751	-1.83617	-0.37156	2.20773	1.10
BSBB24G	-0.21160	-1.31757	-0.47165	1.78922	1.00

Scale Transformation Constants for the TIMSS 2019 *Instructional Clarity in Biology Lessons* Scale – Grade 8

Scale Transformation Constants	
A = 7.994887	Transformed Scale Score = 7.994887 + 0.979530 • Logit Scale Score
B = 0.979530	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Instructional Clarity in Biology Lessons* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	3.80573	
1	4.92740	
2	5.47995	
3	5.86758	
4	6.18006	
5	6.45197	
6	6.70084	
7	6.93739	
8	7.16933	
9	7.40311	
10	7.64551	7.7
11	7.90030	
12	8.17575	
13	8.47625	
14	8.80537	
15	9.16053	
16	9.53839	
17	9.93996	
18	10.37818	10.3
19	10.88385	
20	11.54787	
21	12.77607	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instructional Clarity in Biology Lessons Scale – Grade 8*

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item						
			BSBB24A	BSBB24B	BSBB24C	BSBB24D	BSBB24E	BSBB24F	BSBB24G
Cyprus	0.92	67	0.67	0.81	0.88	0.89	0.85	0.75	0.84
Finland	0.93	71	0.69	0.88	0.89	0.90	0.87	0.81	0.86
France	0.91	66	0.66	0.87	0.88	0.89	0.82	0.73	0.82
Georgia	0.89	62	0.58	0.84	0.85	0.86	0.81	0.75	0.78
Hungary	0.92	67	0.69	0.85	0.88	0.88	0.81	0.80	0.81
Kazakhstan	0.89	62	0.62	0.80	0.86	0.85	0.82	0.76	0.78
Lebanon	0.90	62	0.65	0.83	0.86	0.85	0.83	0.74	0.74
Lithuania	0.89	62	0.69	0.83	0.84	0.84	0.82	0.73	0.75
Morocco	0.87	57	0.53	0.70	0.81	0.82	0.83	0.78	0.79
Portugal	0.90	62	0.63	0.82	0.85	0.87	0.75	0.77	0.80
Romania	0.91	67	0.66	0.85	0.86	0.86	0.84	0.78	0.86
Russian Federation	0.91	66	0.67	0.83	0.87	0.86	0.86	0.77	0.80
Sweden	0.92	68	0.64	0.86	0.82	0.88	0.88	0.84	0.80
Benchmarking Participants									
Moscow City, Russian Fed.	0.91	65	0.68	0.83	0.85	0.86	0.85	0.78	0.77

Relationship Between the TIMSS 2019 *Instructional Clarity in Biology Lessons Scale and TIMSS 2019 Achievement – Grade 8*

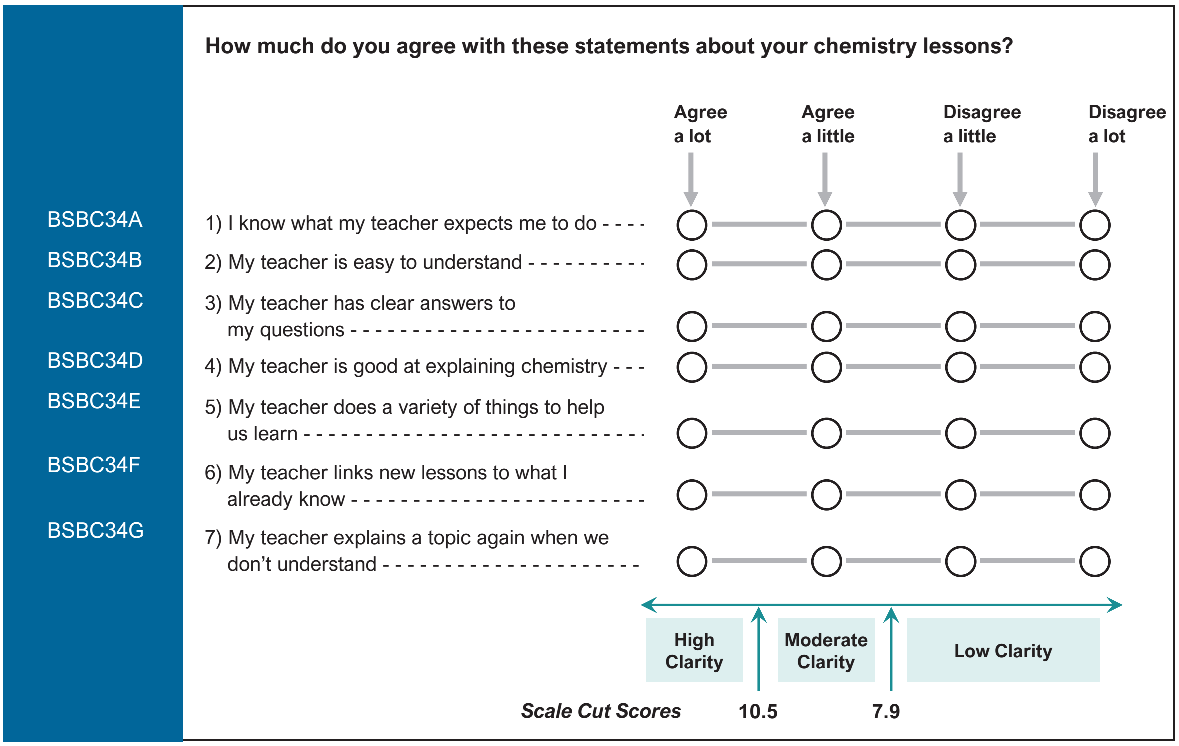
Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Cyprus	0.16	0.03	0.02
Finland	0.19	0.03	0.03
France	0.04	0.00	0.00
Georgia	0.12	0.01	0.01
Hungary	0.04	0.00	0.00
Kazakhstan	0.15	0.02	0.02
Lebanon	0.18	0.03	0.03
Lithuania	0.04	0.00	0.00
Morocco	0.11	0.01	0.01
Portugal	0.06	0.00	0.01
Romania	0.01	0.00	0.00
Russian Federation	0.05	0.00	0.00
Sweden	0.09	0.01	0.01
International Median	**0.09**	**0.01**	**0.01**
Benchmarking Participants			
Moscow City, Russian Fed.	0.05	0.00	0.00

Scale was created for TIMSS 2019 countries where science is taught as separate subjects.

Instructional Clarity in Chemistry Lessons – Grade 8

About the Scale

The *Instructional Clarity in Chemistry Lessons* scale was created based on students' responses to seven items listed below.



Item Parameters for the TIMSS 2019 *Instructional Clarity in Chemistry Lessons* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBC34A	0.30367	-2.02142	-0.58108	2.60250	1.57
BSBC34B	0.29616	-1.88144	-0.52571	2.40715	0.89
BSBC34C	-0.06984	-1.72577	-0.53226	2.25803	0.77
BSBC34D	-0.24544	-1.52782	-0.58802	2.11584	0.76
BSBC34E	-0.10563	-1.79645	-0.47708	2.27353	0.89
BSBC34F	0.16892	-1.95437	-0.46326	2.41763	1.17
BSBC34G	-0.34784	-1.40043	-0.66098	2.06141	1.02

Scale Transformation Constants for the TIMSS 2019 *Instructional Clarity in Chemistry Lessons* Scale – Grade 8

Scale Transformation Constants	
A = 8.258522	Transformed Scale Score = 8.258522 + 0.859737 • Logit Scale Score
B = 0.859737	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Instructional Clarity in Chemistry Lessons* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	4.44982	
1	5.44094	
2	5.93238	
3	6.27893	
4	6.55867	
5	6.80233	
6	7.02593	
7	7.23849	
8	7.44984	
9	7.66458	
10	7.88907	7.9
11	8.13037	
12	8.39611	
13	8.69402	
14	9.02624	
15	9.38399	
16	9.75318	
17	10.12995	
18	10.52773	10.5
19	10.97816	
20	11.56142	
21	12.63681	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instructional Clarity in Chemistry Lessons* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item						
			BSBC34A	BSBC34B	BSBC34C	BSBC34D	BSBC34E	BSBC34F	BSBC34G
Cyprus	0.93	71	0.73	0.85	0.90	0.90	0.88	0.76	0.86
Finland	0.94	75	0.71	0.90	0.91	0.92	0.89	0.85	0.86
France	0.93	70	0.69	0.89	0.90	0.90	0.87	0.77	0.82
Georgia	0.93	70	0.64	0.88	0.89	0.90	0.86	0.82	0.85
Hungary	0.93	71	0.72	0.86	0.89	0.90	0.86	0.81	0.84
Kazakhstan	0.92	67	0.67	0.83	0.88	0.88	0.86	0.79	0.82
Lebanon	0.92	67	0.70	0.86	0.88	0.88	0.85	0.78	0.77
Lithuania	0.92	68	0.71	0.86	0.88	0.88	0.86	0.75	0.80
Morocco	0.90	63	0.58	0.75	0.85	0.84	0.85	0.81	0.81
Portugal	0.92	68	0.67	0.86	0.88	0.89	0.83	0.82	0.81
Romania	0.93	71	0.68	0.87	0.89	0.88	0.87	0.81	0.85
Russian Federation	0.92	70	0.72	0.84	0.89	0.88	0.88	0.80	0.83
Sweden	0.94	73	0.68	0.89	0.91	0.91	0.90	0.85	0.83
Benchmarking Participants									
Moscow City, Russian Fed.	0.93	71	0.73	0.85	0.89	0.89	0.88	0.81	0.83

Relationship Between the TIMSS 2019 *Instructional Clarity in Chemistry Lessons* Scale and TIMSS 2019 Achievement – Grade 8

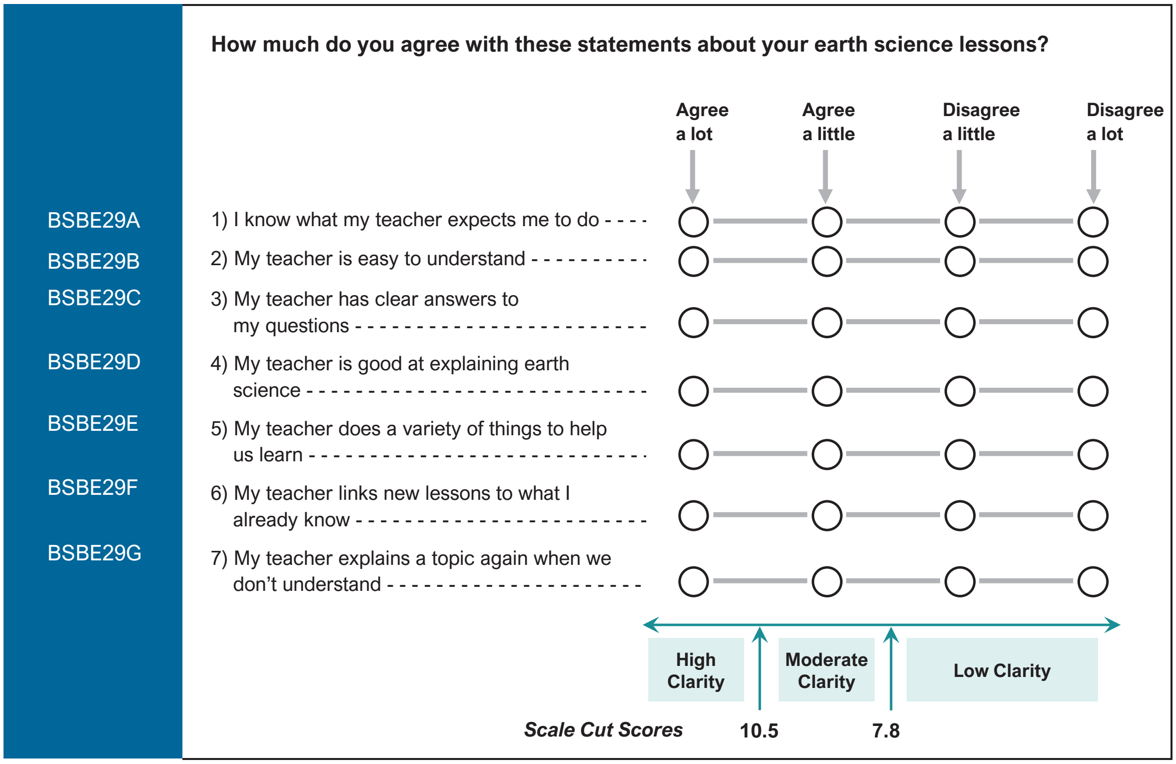
Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Cyprus	0.18	0.03	0.03
Finland	0.27	0.07	0.06
France	0.05	0.00	0.00
Georgia	0.18	0.03	0.02
Hungary	0.04	0.00	0.00
Kazakhstan	0.18	0.03	0.03
Lebanon	0.19	0.03	0.03
Lithuania	0.12	0.02	0.01
Morocco	0.15	0.02	0.03
Portugal	0.12	0.01	0.01
Romania	0.10	0.01	0.01
Russian Federation	0.14	0.02	0.02
Sweden	0.12	0.01	0.01
International Median	**0.14**	**0.02**	**0.02**
Benchmarking Participants			
Moscow City, Russian Fed.	0.11	0.01	0.01

Scale was created for TIMSS 2019 countries where science is taught as separate subjects.

Instructional Clarity in Earth Science Lessons – Grade 8

About the Scale

The *Instructional Clarity in Earth Science Lessons* scale was created based on students' responses to nine items listed below.



Item Parameters for the TIMSS 2019 *Instructional Clarity in Earth Science Lessons* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBE29A	0.37167	-1.95461	-0.56539	2.52000	1.52
BSBE29B	0.08286	-1.72660	-0.62489	2.35149	0.91
BSBE29C	-0.22308	-1.57867	-0.58651	2.16518	0.77
BSBE29D	-0.40135	-1.38541	-0.67655	2.06196	0.76
BSBE29E	0.04460	-1.71685	-0.45018	2.16703	0.95
BSBE29F	0.29232	-1.91852	-0.43425	2.35277	1.13
BSBE29G	-0.16702	-1.38739	-0.59333	1.98072	0.97

Scale Transformation Constants for the TIMSS 2019 *Instructional Clarity in Earth Science Lessons* Scale – Grade 8

Scale Transformation Constants	
A = 8.161485	Transformed Scale Score = 8.161485 + 0.914376 • Logit Scale Score
B = 0.914376	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Instructional Clarity in Earth Science Lessons* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	4.19800	
1	5.23994	
2	5.75344	
3	6.11353	
4	6.40348	
5	6.65566	
6	6.88683	
7	7.10638	
8	7.32462	
9	7.54622	
10	7.77771	7.8
11	8.02627	
12	8.29947	
13	8.60494	
14	8.94504	
15	9.31254	
16	9.69531	
17	10.09005	
18	10.51012	10.5
19	10.98774	
20	11.60849	
21	12.75360	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instructional Clarity in Earth Science Lessons Scale – Grade 8*

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item						
			BSBE29A	BSBE29B	BSBE29C	BSBE29D	BSBE29E	BSBE29F	BSBE29G
Cyprus	0.94	73	0.74	0.86	0.91	0.91	0.87	0.80	0.87
Finland	0.94	74	0.71	0.89	0.91	0.91	0.88	0.82	0.87
France	0.91	66	0.66	0.87	0.88	0.89	0.82	0.73	0.82
Georgia	0.91	66	0.61	0.84	0.87	0.87	0.83	0.80	0.85
Hungary	0.93	69	0.71	0.87	0.89	0.89	0.83	0.80	0.83
Kazakhstan	0.91	65	0.66	0.81	0.88	0.86	0.85	0.76	0.82
Lebanon	-	-	-	-	-	-	-	-	-
Lithuania	0.92	67	0.69	0.86	0.87	0.87	0.85	0.77	0.80
Morocco	0.89	60	0.55	0.73	0.84	0.84	0.83	0.80	0.80
Portugal	0.90	62	0.63	0.82	0.85	0.87	0.75	0.77	0.80
Romania	0.90	64	0.63	0.86	0.86	0.85	0.80	0.78	0.82
Russian Federation	0.92	68	0.69	0.84	0.88	0.87	0.87	0.77	0.82
Sweden	-	-	-	-	-	-	-	-	-
Benchmarking Participants									
Moscow City, Russian Fed.	0.93	70	0.73	0.85	0.88	0.89	0.88	0.80	0.81

Relationship Between the TIMSS 2019 *Instructional Clarity in Earth Science Lessons Scale and TIMSS 2019 Achievement – Grade 8*

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Cyprus	0.14	0.02	0.02
Finland	0.18	0.03	0.03
France	0.04	0.00	0.00
Georgia	0.10	0.01	0.01
Hungary	0.03	0.00	0.01
Kazakhstan	0.13	0.02	0.01
Lebanon	-	-	-
Lithuania	0.05	0.00	0.00
Morocco	0.12	0.02	0.02
Portugal	0.07	0.00	0.01
Romania	0.09	0.01	0.01
Russian Federation	0.02	0.00	0.00
Sweden	-	-	-
International Median	**0.09**	**0.01**	**0.01**
Benchmarking Participants			
Moscow City, Russian Fed.	0.05	0.00	0.00

Scale was created for TIMSS 2019 countries where science is taught as separate subjects.

A dash (–) indicates comparable data not available.

Instructional Clarity in Mathematics Lessons – Grade 8

About the Scale

The *Instructional Clarity in Mathematics Lessons* scale was created based on students' responses to seven items listed below.

How much do you agree with these statements about your mathematics lessons?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BSBM17A	1) I know what my teacher expects me to do	○	○	○	○
BSBM17B	2) My teacher is easy to understand	○	○	○	○
BSBM17C	3) My teacher has clear answers to my questions	○	○	○	○
BSBM17D	4) My teacher is good at explaining mathematics	○	○	○	○
BSBM17E	5) My teacher does a variety of things to help us learn	○	○	○	○
BSBM17F	6) My teacher links new lessons to what I already know	○	○	○	○
BSBM17G	7) My teacher explains a topic again when we don't understand	○	○	○	○

High Clarity | Moderate Clarity | Low Clarity

Scale Cut Scores 10.3 7.8

Item Parameters for the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBM17A	0.05459	-1.49723	-0.59435	2.09158	1.43
BSBM17B	0.32080	-1.45404	-0.41713	1.87117	0.92
BSBM17C	0.00334	-1.36560	-0.33731	1.70291	0.81
BSBM17D	-0.11300	-1.01045	-0.39445	1.40490	0.77
BSBM17E	-0.03564	-1.22811	-0.35992	1.58803	0.93
BSBM17F	0.11754	-1.42524	-0.44389	1.86913	1.04
BSBM17G	-0.34763	-0.88348	-0.46134	1.34482	0.99

Scale Transformation Constants for the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* Scale – Grade 8

Scale Transformation Constants	
A = 8.053331	Transformed Scale Score = 8.053331 + 1.109981 • Logit Scale Score
B = 1.109981	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	3.69567	
1	4.93373	
2	5.53063	
3	5.94336	
4	6.27085	
5	6.55198	
6	6.80608	
7	7.04464	
8	7.27558	
9	7.50516	
10	7.73983	7.8
11	7.98224	
12	8.24029	
13	8.51896	
14	8.82274	
15	9.15605	
16	9.52219	
17	9.92675	
18	10.38390	10.3
19	10.92487	
20	11.65115	
21	13.01885	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item						
			BSBM17A	BSBM17B	BSBM17C	BSBM17D	BSBM17E	BSBM17F	BSBM17G
Australia	0.92	67	0.69	0.87	0.87	0.88	0.82	0.80	0.79
Bahrain	0.88	59	0.57	0.79	0.84	0.84	0.81	0.77	0.74
Chile	0.90	62	0.61	0.81	0.86	0.86	0.80	0.77	0.77
Chinese Taipei	0.91	65	0.73	0.80	0.81	0.81	0.83	0.85	0.82
Cyprus	0.90	62	0.63	0.78	0.85	0.88	0.83	0.72	0.80
Egypt	0.81	48	0.43	0.62	0.78	0.77	0.79	0.70	0.66
England	0.91	66	0.68	0.85	0.88	0.88	0.82	0.77	0.78
Finland	0.92	67	0.60	0.87	0.87	0.90	0.85	0.79	0.82
France	0.88	58	0.55	0.83	0.85	0.87	0.76	0.66	0.77
Georgia	0.87	58	0.52	0.80	0.83	0.85	0.80	0.71	0.77
Hong Kong SAR	0.94	73	0.75	0.86	0.89	0.88	0.86	0.88	0.83
Hungary	0.91	66	0.66	0.84	0.87	0.89	0.82	0.77	0.81
Iran, Islamic Rep. of	0.88	57	0.56	0.76	0.82	0.83	0.77	0.78	0.75
Ireland	0.91	64	0.63	0.87	0.86	0.89	0.80	0.76	0.77
Israel	0.89	60	0.61	0.83	0.85	0.86	0.76	0.72	0.74
Italy	0.86	54	0.46	0.73	0.84	0.84	0.77	0.73	0.72
Japan	0.90	64	0.55	0.85	0.84	0.88	0.85	0.77	0.80
Jordan	0.87	57	0.54	0.71	0.83	0.83	0.81	0.75	0.75
Kazakhstan	0.86	55	0.53	0.73	0.80	0.83	0.83	0.70	0.76
Korea, Rep. of	0.91	65	0.54	0.80	0.85	0.89	0.85	0.84	0.83
Kuwait	0.87	56	0.58	0.77	0.82	0.83	0.76	0.74	0.72
Lebanon	0.86	55	0.59	0.80	0.82	0.81	0.79	0.65	0.70
Lithuania	0.89	60	0.62	0.83	0.83	0.83	0.80	0.71	0.76
Malaysia	0.86	54	0.56	0.76	0.79	0.81	0.77	0.71	0.73
Morocco	0.88	59	0.52	0.70	0.83	0.85	0.85	0.78	0.79
New Zealand	0.92	66	0.67	0.86	0.87	0.87	0.83	0.77	0.80
Norway (9)	0.90	63	0.58	0.85	0.84	0.87	0.80	0.80	0.79
Oman	0.84	51	0.46	0.71	0.80	0.80	0.74	0.71	0.72
Portugal	0.90	64	0.63	0.83	0.85	0.88	0.79	0.78	0.81
Qatar	0.90	63	0.64	0.83	0.86	0.85	0.82	0.77	0.76
Romania	0.90	62	0.60	0.82	0.84	0.84	0.81	0.77	0.81
Russian Federation	0.87	58	0.59	0.75	0.82	0.83	0.82	0.73	0.75
Saudi Arabia	0.86	56	0.50	0.76	0.82	0.83	0.80	0.73	0.72
Singapore	0.90	62	0.67	0.84	0.85	0.86	0.76	0.76	0.75
South Africa (9)	0.83	49	0.54	0.76	0.76	0.80	0.68	0.62	0.71
Sweden	0.92	66	0.60	0.86	0.88	0.89	0.87	0.81	0.77
Turkey	0.85	54	0.57	0.78	0.78	0.82	0.72	0.73	0.70
United Arab Emirates	0.90	62	0.64	0.83	0.84	0.85	0.81	0.77	0.77
United States	0.91	66	0.67	0.86	0.86	0.88	0.82	0.77	0.80

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item						
			BSBM17A	BSBM17B	BSBM17C	BSBM17D	BSBM17E	BSBM17F	BSBM17G
Benchmarking Participants									
Ontario, Canada	0.90	63	0.64	0.85	0.84	0.85	0.80	0.77	0.76
Quebec, Canada	0.90	63	0.62	0.86	0.85	0.87	0.80	0.77	0.75
Moscow City, Russian Fed.	0.88	59	0.58	0.79	0.83	0.85	0.84	0.72	0.74
Gauteng, RSA (9)	0.86	54	0.53	0.81	0.81	0.83	0.74	0.65	0.74
Western Cape, RSA (9)	0.87	55	0.56	0.80	0.80	0.82	0.75	0.70	0.72
Abu Dhabi, UAE	0.90	62	0.64	0.83	0.83	0.84	0.81	0.76	0.76
Dubai, UAE	0.89	61	0.64	0.83	0.84	0.85	0.79	0.76	0.74

Relationship Between the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Australia	0.23	0.05	0.05
Bahrain	0.15	0.02	0.02
Chile	0.12	0.01	0.01
Chinese Taipei	0.27	0.07	0.06
Cyprus	0.23	0.05	0.04
Egypt	0.13	0.02	0.02
England	0.10	0.01	0.01
Finland	0.23	0.05	0.05
France	0.11	0.01	0.02
Georgia	0.14	0.02	0.02
Hong Kong SAR	0.17	0.03	0.03
Hungary	0.21	0.05	0.04
Iran, Islamic Rep. of	0.15	0.02	0.02
Ireland	0.04	0.00	0.00
Israel	0.09	0.01	0.01
Italy	0.09	0.01	0.01
Japan	0.19	0.04	0.03
Jordan	0.16	0.03	0.02
Kazakhstan	0.17	0.03	0.02
Korea, Rep. of	0.31	0.10	0.07
Kuwait	0.12	0.01	0.01
Lebanon	0.20	0.04	0.04
Lithuania	0.18	0.03	0.03
Malaysia	0.14	0.02	0.01
Morocco	0.11	0.01	0.01
New Zealand	0.12	0.02	0.02
Norway (9)	0.22	0.05	0.05
Oman	0.22	0.05	0.05
Portugal	0.12	0.01	0.01
Qatar	0.15	0.02	0.02
Romania	0.16	0.03	0.02
Russian Federation	0.15	0.02	0.02
Saudi Arabia	0.16	0.03	0.02
Singapore	0.18	0.03	0.03
South Africa (9)	0.07	0.00	0.00
Sweden	0.11	0.01	0.01
Turkey	0.23	0.05	0.04
United Arab Emirates	0.23	0.05	0.05
United States	0.15	0.02	0.02
International Median	**0.15**	**0.02**	**0.02**

Relationship Between the TIMSS 2019 *Instructional Clarity in Mathematics Lessons* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Benchmarking Participants			
Ontario, Canada	0.18	0.03	0.03
Quebec, Canada	0.13	0.02	0.02
Moscow City, Russian Fed.	0.17	0.03	0.02
Gauteng, RSA (9)	0.05	0.00	0.00
Western Cape, RSA (9)	-0.01	0.00	0.00
Abu Dhabi, UAE	0.25	0.06	0.06
Dubai, UAE	0.20	0.04	0.03

Instructional Clarity in Physics Lessons – Grade 8

About the Scale

The *Instructional Clarity in Physics Lessons* scale was created based on students' responses to seven items listed below.

How much do you agree with these statements about your physics lessons?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BSBP39A	1) I know what my teacher expects me to do	○	○	○	○
BSBP39B	2) My teacher is easy to understand	○	○	○	○
BSBP39C	3) My teacher has clear answers to my questions	○	○	○	○
BSBP39D	4) My teacher is good at explaining physics	○	○	○	○
BSBP39E	5) My teacher does a variety of things to help us learn	○	○	○	○
BSBP39F	6) My teacher links new lessons to what I already know	○	○	○	○
BSBP39G	7) My teacher explains a topic again when we don't understand	○	○	○	○

High Clarity	Moderate Clarity	Low Clarity

Scale Cut Scores **10.5** **8.0**

Item Parameters for the TIMSS 2019 *Instructional Clarity in Physics Lessons* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBP39A	0.34432	-2.05050	-0.59666	2.64716	1.64
BSBP39B	0.27344	-1.97617	-0.55340	2.52957	0.91
BSBP39C	-0.08583	-1.79388	-0.54214	2.33602	0.76
BSBP39D	-0.22852	-1.61545	-0.60200	2.21745	0.78
BSBP39E	-0.13662	-1.86085	-0.51526	2.37611	0.89
BSBP39F	0.15357	-2.00034	-0.49510	2.49544	1.13
BSBP39G	-0.32036	-1.51653	-0.62527	2.14180	1.03

Scale Transformation Constants for the TIMSS 2019 *Instructional Clarity in Physics Lessons* Scale – Grade 8

Scale Transformation Constants	
A = 8.328119	Transformed Scale Score = 8.328119 + 0.832906 • Logit Scale Score
B = 0.832906	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Instructional Clarity in Physics Lessons* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	4.57593	
1	5.54107	
2	6.02148	
3	6.36159	
4	6.63685	
5	6.87705	
6	7.09773	
7	7.30766	
8	7.51670	
9	7.72934	
10	7.95209	8.0
11	8.19235	
12	8.45841	
13	8.75915	
14	9.09682	
15	9.45971	
16	9.82967	
17	10.20206	
18	10.59164	10.5
19	11.03337	
20	11.59772	
21	12.64074	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instructional Clarity in Physics Lessons* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item: BSBP39A	BSBP39B	BSBP39C	BSBP39D	BSBP39E	BSBP39F	BSBP39G
Cyprus	0.94	73	0.74	0.85	0.91	0.91	0.90	0.80	0.87
Finland	0.95	77	0.72	0.91	0.92	0.93	0.91	0.86	0.88
France	0.93	70	0.69	0.89	0.90	0.90	0.87	0.77	0.82
Georgia	0.92	69	0.62	0.86	0.89	0.88	0.86	0.84	0.84
Hungary	0.94	72	0.72	0.86	0.90	0.90	0.86	0.83	0.85
Kazakhstan	0.92	68	0.67	0.83	0.89	0.89	0.86	0.79	0.83
Lebanon	0.92	67	0.68	0.85	0.88	0.88	0.84	0.79	0.78
Lithuania	0.93	69	0.70	0.87	0.90	0.89	0.87	0.77	0.82
Morocco	0.90	63	0.58	0.76	0.85	0.85	0.86	0.82	0.81
Portugal	0.92	68	0.67	0.86	0.88	0.89	0.83	0.82	0.81
Romania	0.93	71	0.66	0.87	0.90	0.88	0.87	0.84	0.86
Russian Federation	0.93	71	0.72	0.86	0.89	0.89	0.88	0.80	0.84
Sweden	0.94	74	0.71	0.89	0.91	0.90	0.90	0.85	0.83
Benchmarking Participants									
Moscow City, Russian Fed.	0.94	72	0.75	0.86	0.90	0.89	0.89	0.82	0.84

Relationship Between the TIMSS 2019 *Instructional Clarity in Physics Lessons* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Cyprus	0.20	0.04	0.04
Finland	0.25	0.06	0.05
France	0.05	0.00	0.00
Georgia	0.14	0.02	0.01
Hungary	0.12	0.02	0.01
Kazakhstan	0.18	0.03	0.02
Lebanon	0.17	0.03	0.03
Lithuania	0.05	0.00	0.00
Morocco	0.15	0.02	0.03
Portugal	0.12	0.01	0.01
Romania	0.09	0.01	0.01
Russian Federation	0.07	0.01	0.01
Sweden	0.13	0.02	0.02
International Median	**0.13**	**0.02**	**0.01**
Benchmarking Participants			
Moscow City, Russian Fed.	0.12	0.01	0.01

Scale was created for TIMSS 2019 countries where science is taught as separate subjects.

Instructional Clarity in Science Lessons – Grade 8

About the Scale

The *Instructional Clarity in Science Lessons* scale was created based on students' responses to seven items listed below.

How much do you agree with these statements about your science lessons?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BSBS23A	1) I know what my teacher expects me to do	○	○	○	○
BSBS23B	2) My teacher is easy to understand	○	○	○	○
BSBS23C	3) My teacher has clear answers to my questions	○	○	○	○
BSBS23D	4) My teacher is good at explaining science	○	○	○	○
BSBS23E	5) My teacher does a variety of things to help us learn	○	○	○	○
BSBS23F	6) My teacher links new lessons to what I already know	○	○	○	○
BSBS23G	7) My teacher explains a topic again when we don't understand	○	○	○	○

High Clarity	Moderate Clarity	Low Clarity

Scale Cut Scores **10.3** **7.8**

Item Parameters for the TIMSS 2019 *Instructional Clarity in Science Lessons* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBS23A	0.15452	-1.66449	-0.56400	2.22849	1.45
BSBS23B	0.19752	-1.56622	-0.45528	2.02150	0.89
BSBS23C	-0.05676	-1.54981	-0.33550	1.88531	0.80
BSBS23D	-0.18594	-1.24450	-0.44753	1.69203	0.78
BSBS23E	-0.08849	-1.49870	-0.38792	1.88662	0.90
BSBS23F	0.18342	-1.66061	-0.41828	2.07889	1.02
BSBS23G	-0.20427	-1.21307	-0.48440	1.69747	1.00

Scale Transformation Constants for the TIMSS 2019 *Instructional Clarity in Science Lessons* Scale – Grade 8

Scale Transformation Constants	
A = 8.069663	Transformed Scale Score = 8.069663 + 0.974947 • Logit Scale Score
B = 0.974947	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Instructional Clarity in Science Lessons* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	4.02273	
1	5.13310	
2	5.67786	
3	6.05827	
4	6.36240	
5	6.62529	
6	6.86434	
7	7.09012	
8	7.31015	
9	7.53057	
10	7.75780	7.8
11	7.99517	
12	8.25053	
13	8.52820	
14	8.83177	
15	9.16063	
16	9.51303	
17	9.89078	
18	10.30639	10.3
19	10.78969	
20	11.43106	
21	12.63386	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instructional Clarity in Science Lessons* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item						
			BSBS23A	BSBS23B	BSBS23C	BSBS23D	BSBS23E	BSBS23F	BSBS23G
Australia	0.94	74	0.76	0.89	0.90	0.90	0.87	0.84	0.84
Bahrain	0.91	65	0.66	0.83	0.86	0.86	0.83	0.79	0.78
Chile	0.92	67	0.69	0.82	0.85	0.88	0.87	0.80	0.81
Chinese Taipei	0.92	69	0.75	0.83	0.83	0.86	0.85	0.85	0.83
Cyprus	-	-	-	-	-	-	-	-	-
Egypt	0.83	51	0.44	0.65	0.80	0.79	0.79	0.75	0.70
England	0.93	71	0.75	0.88	0.89	0.88	0.85	0.82	0.82
Finland	-	-	-	-	-	-	-	-	-
France	-	-	-	-	-	-	-	-	-
Georgia	-	-	-	-	-	-	-	-	-
Hong Kong SAR	0.95	77	0.81	0.90	0.91	0.90	0.89	0.87	0.86
Hungary	-	-	-	-	-	-	-	-	-
Iran, Islamic Rep. of	0.89	61	0.59	0.79	0.83	0.85	0.81	0.77	0.78
Ireland	0.93	71	0.71	0.89	0.89	0.90	0.85	0.82	0.81
Israel	0.93	70	0.74	0.88	0.89	0.89	0.83	0.80	0.81
Italy	0.88	58	0.49	0.78	0.86	0.83	0.79	0.76	0.76
Japan	0.91	67	0.58	0.86	0.86	0.89	0.87	0.80	0.83
Jordan	0.89	60	0.56	0.74	0.85	0.84	0.84	0.79	0.80
Kazakhstan	-	-	-	-	-	-	-	-	-
Korea, Rep. of	0.92	70	0.65	0.83	0.87	0.88	0.86	0.87	0.85
Kuwait	0.89	60	0.61	0.80	0.85	0.84	0.79	0.76	0.77
Lebanon	-	-	-	-	-	-	-	-	-
Lithuania	-	-	-	-	-	-	-	-	-
Malaysia	0.88	59	0.63	0.79	0.82	0.83	0.80	0.74	0.75
Morocco	-	-	-	-	-	-	-	-	-
New Zealand	0.93	71	0.73	0.88	0.89	0.89	0.87	0.81	0.83
Norway (9)	0.92	69	0.65	0.87	0.87	0.89	0.85	0.83	0.82
Oman	0.86	56	0.48	0.76	0.82	0.82	0.79	0.74	0.76
Portugal	-	-	-	-	-	-	-	-	-
Qatar	0.93	69	0.71	0.86	0.88	0.88	0.86	0.82	0.80
Romania	-	-	-	-	-	-	-	-	-
Russian Federation	-	-	-	-	-	-	-	-	-
Saudi Arabia	0.89	60	0.61	0.78	0.84	0.83	0.82	0.77	0.76
Singapore	0.91	65	0.71	0.86	0.86	0.86	0.79	0.77	0.78
South Africa (9)	0.86	54	0.60	0.79	0.78	0.80	0.72	0.68	0.72
Sweden	-	-	-	-	-	-	-	-	-
Turkey	0.90	63	0.65	0.84	0.85	0.86	0.80	0.78	0.77
United Arab Emirates	0.91	66	0.67	0.84	0.86	0.86	0.84	0.80	0.79
United States	0.93	71	0.73	0.87	0.89	0.90	0.86	0.81	0.83

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instructional Clarity in Science Lessons* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item: BSBS23A	BSBS23B	BSBS23C	BSBS23D	BSBS23E	BSBS23F	BSBS23G
Benchmarking Participants									
Ontario, Canada	0.93	69	0.69	0.88	0.88	0.88	0.85	0.82	0.81
Quebec, Canada	0.92	69	0.65	0.87	0.89	0.88	0.85	0.81	0.82
Moscow City, Russian Fed.	-	-	-	-	-	-	-	-	-
Gauteng, RSA (9)	0.88	58	0.61	0.82	0.83	0.84	0.78	0.71	0.74
Western Cape, RSA (9)	0.89	60	0.63	0.82	0.83	0.83	0.79	0.75	0.76
Abu Dhabi, UAE	0.92	66	0.69	0.85	0.86	0.86	0.84	0.80	0.79
Dubai, UAE	0.91	65	0.68	0.85	0.86	0.86	0.82	0.79	0.79

A dash (–) indicates comparable data not available.

Relationship Between the TIMSS 2019 *Instructional Clarity in Science Lessons* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Australia	0.19	0.04	0.03
Bahrain	0.20	0.04	0.04
Chile	0.06	0.00	0.00
Chinese Taipei	0.23	0.05	0.05
Cyprus	-	-	-
Egypt	0.24	0.06	0.06
England	0.17	0.03	0.03
Finland	-	-	-
France	-	-	-
Georgia	-	-	-
Hong Kong SAR	0.11	0.01	0.01
Hungary	-	-	-
Iran, Islamic Rep. of	0.14	0.02	0.02
Ireland	0.12	0.01	0.01
Israel	0.07	0.01	0.00
Italy	0.06	0.00	0.00
Japan	0.21	0.05	0.03
Jordan	0.20	0.04	0.04
Kazakhstan	-	-	-
Korea, Rep. of	0.31	0.09	0.09
Kuwait	0.16	0.02	0.02
Lebanon	-	-	-
Lithuania	-	-	-
Malaysia	0.22	0.05	0.04
Morocco	-	-	-
New Zealand	0.13	0.02	0.02
Norway (9)	0.15	0.02	0.03
Oman	0.24	0.06	0.06
Portugal	-	-	-
Qatar	0.14	0.02	0.02
Romania	-	-	-
Russian Federation	-	-	-
Saudi Arabia	0.17	0.03	0.03
Singapore	0.16	0.02	0.02
South Africa (9)	0.07	0.01	0.01
Sweden	-	-	-
Turkey	0.14	0.02	0.01
United Arab Emirates	0.23	0.05	0.05
United States	0.07	0.00	0.00
International Median	**0.16**	**0.02**	**0.02**

Relationship Between the TIMSS 2019 *Instructional Clarity in Science Lessons* Scale and TIMSS 2019 Achievement – Grade 8

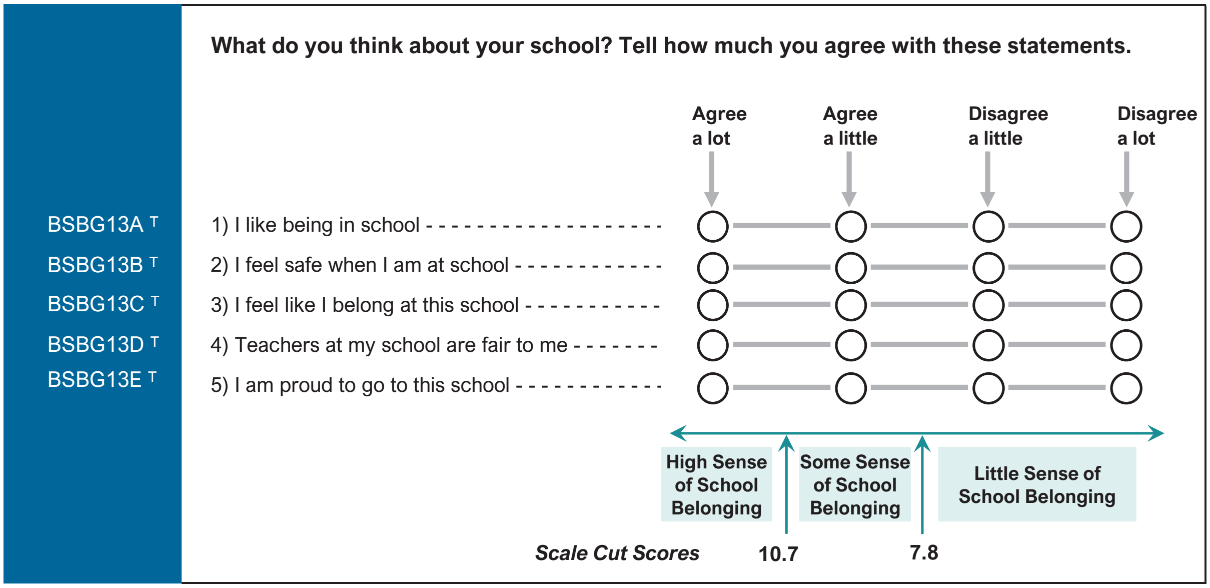
Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Benchmarking Participants			
Ontario, Canada	0.09	0.01	0.01
Quebec, Canada	0.14	0.02	0.02
Moscow City, Russian Fed.	-	-	-
Gauteng, RSA (9)	0.03	0.00	0.00
Western Cape, RSA (9)	-0.02	0.00	0.00
Abu Dhabi, UAE	0.29	0.09	0.08
Dubai, UAE	0.13	0.02	0.01

A dash (–) indicates comparable data not available.

Sense of School Belonging – Grade 8

About the Scale

The *Sense of School Belonging* scale was created based on students' responses to five items listed below.



[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Sense of School Belonging* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBG13A	0.32714	-1.10477	-0.73290	1.83767	1.01
BSBG13B	-0.22817	-1.02570	-0.55000	1.57570	0.98
BSBG13C	-0.06550	-0.90206	-0.56127	1.46333	0.95
BSBG13D	-0.07279	-1.03563	-0.53608	1.57171	1.13
BSBG13E	0.03932	-0.82245	-0.53816	1.36061	0.93

Scale Transformation Constants for the TIMSS 2019 *Sense of School Belonging* Scale – Grade 8

Scale Transformation Constants	
A = 8.156530	Transformed Scale Score = 8.156530 + 1.281319 • Logit Scale Score
B = 1.281319	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Sense of School Belonging* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	3.92440	
1	5.25880	
2	5.90263	
3	6.36071	
4	6.73999	
5	7.08155	
6	7.40784	
7	7.73555	7.8
8	8.08091	
9	8.46096	
10	8.89725	
11	9.40801	
12	10.00496	
13	10.70956	10.7
14	11.62252	
15	13.27199	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Sense of School Belonging* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item				
			BSBG13A	BSBG13B	BSBG13C	BSBG13D	BSBG13E
Australia	0.85	62	0.79	0.77	0.82	0.70	0.84
Bahrain	0.80	56	0.74	0.74	0.82	0.61	0.81
Chile	0.81	58	0.73	0.79	0.81	0.66	0.80
Chinese Taipei	0.81	58	0.73	0.79	0.80	0.68	0.79
Cyprus	0.80	55	0.74	0.77	0.74	0.66	0.80
Egypt	0.78	53	0.71	0.75	0.78	0.63	0.77
England	0.82	59	0.76	0.76	0.79	0.70	0.82
Finland	0.82	58	0.79	0.76	0.80	0.70	0.77
France	0.72	47	0.73	0.68	0.59	0.63	0.77
Georgia	0.76	51	0.68	0.67	0.76	0.65	0.79
Hong Kong SAR	0.86	65	0.81	0.84	0.87	0.67	0.81
Hungary	0.82	59	0.72	0.76	0.81	0.68	0.84
Iran, Islamic Rep. of	0.71	48	0.73	0.76	0.77	0.67	0.50
Ireland	0.81	58	0.74	0.77	0.78	0.67	0.82
Israel	0.82	58	0.77	0.74	0.80	0.69	0.82
Italy	0.77	53	0.71	0.73	0.73	0.65	0.80
Japan	0.85	63	0.80	0.82	0.83	0.68	0.83
Jordan	0.80	56	0.74	0.74	0.83	0.61	0.80
Kazakhstan	0.81	57	0.75	0.69	0.78	0.72	0.81
Korea, Rep. of	0.84	61	0.79	0.81	0.80	0.68	0.81
Kuwait	0.80	56	0.76	0.73	0.80	0.64	0.79
Lebanon	0.76	51	0.71	0.73	0.77	0.59	0.76
Lithuania	0.79	54	0.75	0.73	0.73	0.66	0.79
Malaysia	0.72	47	0.72	0.67	0.64	0.67	0.74
Morocco	0.77	53	0.73	0.72	0.77	0.60	0.79
New Zealand	0.81	57	0.76	0.74	0.81	0.66	0.79
Norway (9)	0.83	60	0.76	0.77	0.83	0.68	0.82
Oman	0.76	51	0.69	0.72	0.76	0.63	0.75
Portugal	0.78	53	0.76	0.73	0.73	0.63	0.79
Qatar	0.84	60	0.77	0.78	0.82	0.70	0.82
Romania	0.80	56	0.73	0.73	0.75	0.69	0.82
Russian Federation	0.80	55	0.74	0.75	0.75	0.67	0.81
Saudi Arabia	0.78	53	0.72	0.69	0.81	0.59	0.79
Singapore	0.83	59	0.76	0.76	0.84	0.66	0.81
South Africa (9)	0.70	46	0.68	0.69	0.77	0.48	0.75
Sweden	0.80	55	0.71	0.70	0.83	0.64	0.81
Turkey	0.76	51	0.70	0.75	0.77	0.58	0.75
United Arab Emirates	0.83	59	0.76	0.77	0.81	0.68	0.81
United States	0.83	59	0.74	0.75	0.81	0.68	0.84

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Sense of School Belonging* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item				
			BSBG13A	BSBG13B	BSBG13C	BSBG13D	BSBG13E
Benchmarking Participants							
Ontario, Canada	0.82	58	0.75	0.73	0.79	0.69	0.83
Quebec, Canada	0.81	56	0.76	0.75	0.78	0.66	0.79
Moscow City, Russian Fed.	0.79	54	0.77	0.74	0.78	0.60	0.79
Gauteng, RSA (9)	0.73	49	0.67	0.69	0.80	0.53	0.78
Western Cape, RSA (9)	0.73	49	0.68	0.68	0.80	0.50	0.79
Abu Dhabi, UAE	0.81	56	0.74	0.76	0.80	0.64	0.80
Dubai, UAE	0.81	57	0.74	0.76	0.79	0.67	0.80

Relationship Between the TIMSS 2019 *Sense of School Belonging* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Australia	0.30	0.09	0.08	0.29	0.09	0.08
Bahrain	0.10	0.01	0.01	0.16	0.02	0.02
Chile	0.07	0.00	0.01	0.07	0.01	0.01
Chinese Taipei	0.08	0.01	0.01	0.06	0.00	0.01
Cyprus	0.18	0.03	0.03	0.18	0.03	0.03
Egypt	0.01	0.00	0.00	0.05	0.00	0.00
England	0.29	0.08	0.08	0.32	0.10	0.10
Finland	0.21	0.04	0.04	0.21	0.04	0.04
France	0.12	0.01	0.02	0.13	0.02	0.02
Georgia	0.01	0.00	0.00	0.01	0.00	0.00
Hong Kong SAR	0.16	0.02	0.03	0.14	0.02	0.03
Hungary	0.18	0.03	0.03	0.15	0.02	0.02
Iran, Islamic Rep. of	0.05	0.00	0.01	0.04	0.00	0.00
Ireland	0.27	0.07	0.07	0.25	0.06	0.06
Israel	0.02	0.00	0.01	0.04	0.00	0.00
Italy	0.07	0.00	0.01	0.04	0.00	0.00
Japan	0.10	0.01	0.01	0.06	0.00	0.01
Jordan	0.04	0.00	0.00	0.06	0.00	0.00
Kazakhstan	0.03	0.00	0.00	0.01	0.00	0.00
Korea, Rep. of	0.14	0.02	0.02	0.13	0.02	0.02
Kuwait	0.13	0.02	0.02	0.15	0.02	0.03
Lebanon	0.13	0.02	0.02	0.14	0.02	0.02
Lithuania	0.07	0.00	0.01	0.07	0.00	0.01
Malaysia	0.01	0.00	0.00	0.06	0.00	0.00
Morocco	-0.04	0.00	0.00	0.02	0.00	0.00
New Zealand	0.23	0.06	0.05	0.21	0.04	0.04
Norway (9)	0.16	0.03	0.03	0.15	0.02	0.03
Oman	0.11	0.01	0.01	0.13	0.02	0.02
Portugal	0.10	0.01	0.01	0.10	0.01	0.01
Qatar	0.15	0.02	0.03	0.15	0.02	0.03
Romania	0.05	0.00	0.00	0.04	0.00	0.00
Russian Federation	0.07	0.00	0.00	0.02	0.00	0.00
Saudi Arabia	0.04	0.00	0.00	0.03	0.00	0.00
Singapore	0.20	0.04	0.04	0.18	0.03	0.03
South Africa (9)	-0.01	0.00	0.00	-0.04	0.00	0.00
Sweden	0.17	0.03	0.03	0.16	0.03	0.02
Turkey	0.02	0.00	0.00	0.01	0.00	0.00
United Arab Emirates	0.18	0.03	0.03	0.19	0.04	0.04
United States	0.26	0.07	0.06	0.22	0.05	0.05
International Median	**0.10**	**0.01**	**0.01**	**0.13**	**0.02**	**0.02**

Relationship Between the TIMSS 2019 *Sense of School Belonging* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Benchmarking Participants						
Ontario, Canada	0.18	0.03	0.03	0.18	0.03	0.03
Quebec, Canada	0.24	0.06	0.06	0.22	0.05	0.04
Moscow City, Russian Fed.	0.09	0.01	0.01	0.07	0.01	0.00
Gauteng, RSA (9)	-0.04	0.00	0.00	-0.06	0.00	0.00
Western Cape, RSA (9)	-0.01	0.00	0.00	-0.05	0.00	0.00
Abu Dhabi, UAE	0.19	0.04	0.03	0.19	0.03	0.03
Dubai, UAE	0.17	0.03	0.03	0.17	0.03	0.03

Student Bullying – Grade 8

About the Scale

The *Student Bullying* scale was created based on students' responses to fourteen items listed below.

During this school year, how often have other students from your school done any of the following things to you, including through texting or the Internet?

		Never	A few times a year	Once or twice a month	At least once a week
BSBG14A	1) Said mean things about my physical appearance (e.g., my hair, my size)	○	○	○	○
BSBG14B	2) Spread lies about me	○	○	○	○
BSBG14C	3) Shared my secrets with others	○	○	○	○
BSBG14D	4) Refused to talk to me	○	○	○	○
BSBG14E	5) Insulted a member of my family	○	○	○	○
BSBG14F	6) Stole something from me	○	○	○	○
BSBG14G	7) Made me do things I didn't want to do	○	○	○	○
BSBG14H	8) Sent me nasty or hurtful messages online	○	○	○	○
BSBG14I	9) Shared nasty or hurtful things about me online	○	○	○	○
BSBG14J	10) Shared embarrassing photos of me online	○	○	○	○
BSBG14K	11) Threatened me	○	○	○	○
BSBG14L	12) Physically hurt me	○	○	○	○
BSBG14M	13) Excluded me from their group (e.g., parties, messaging)	○	○	○	○
BSBG14N	14) Damaged something of mine on purpose	○	○	○	○

Never or Almost Never	About Monthly	About Weekly

Scale Cut Scores 8.8 7.2

Item Parameters for the TIMSS 2019 *Student Bullying* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBG14A	0.71305	0.04645	-0.26859	0.22214	1.10
BSBG14B	0.52658	-0.41912	-0.19205	0.61117	0.98
BSBG14C	0.39392	-0.22178	-0.15770	0.37948	1.10
BSBG14D	0.23331	-0.01799	-0.12110	0.13909	1.09
BSBG14E	0.07453	0.24301	0.02265	-0.26566	1.07
BSBG14F	0.08557	0.07582	-0.11162	0.03580	1.10
BSBG14G	-0.19467	0.10093	0.07979	-0.18072	0.99
BSBG14H	-0.11689	0.33415	0.03433	-0.36848	1.05
BSBG14I	-0.43644	0.39686	0.16165	-0.55851	0.81
BSBG14J	-0.43928	0.18892	0.08301	-0.27193	1.00
BSBG14K	-0.36957	0.31464	0.10156	-0.41620	0.82
BSBG14L	-0.27648	0.24465	-0.00493	-0.23972	0.91
BSBG14M	-0.02422	0.13198	-0.17425	0.04227	1.02
BSBG14N	-0.16941	0.20537	-0.10652	-0.09885	0.88

Scale Transformation Constants for the TIMSS 2019 *Student Bullying* Scale – Grade 8

Scale Transformation Constants	
A = 7.220866	Transformed Scale Score = 7.220866 + 1.652921 • Logit Scale Score
B = 1.652921	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Student Bullying* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	1.95344	
1	3.57732	
2	4.25794	
3	4.68054	
4	4.98853	
5	5.23124	
6	5.43399	
7	5.60919	
8	5.76449	
9	5.90507	
10	6.03321	
11	6.15477	
12	6.26941	
13	6.37845	
14	6.48302	
15	6.58405	
16	6.68234	
17	6.77861	
18	6.87349	
19	6.96754	
20	7.06124	
21	7.15513	7.2
22	7.25088	
23	7.34528	
24	7.44254	
25	7.54198	
26	7.64418	
27	7.74978	
28	7.85954	
29	7.97425	
30	8.09367	
31	8.22128	
32	8.35784	
33	8.50552	
34	8.66719	
35	8.84686	8.8
36	9.05015	
37	9.28593	
38	9.56975	
39	9.92809	
40	10.41803	
41	11.18784	
42	12.93736	

TIMSS 2019

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Student Bullying* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item													
			BSBG14A	BSBG14B	BSBG14C	BSBG14D	BSBG14E	BSBG14F	BSBG14G	BSBG14H	BSBG14I	BSBG14J	BSBG14K	BSBG14L	BSBG14M	BSBG14N
Australia	0.91	47	0.67	0.77	0.68	0.69	0.67	0.61	0.68	0.74	0.74	0.56	0.74	0.64	0.67	0.72
Bahrain	0.90	44	0.61	0.67	0.62	0.62	0.68	0.58	0.67	0.72	0.69	0.66	0.73	0.70	0.66	0.70
Chile	0.87	40	0.64	0.65	0.63	0.62	0.63	0.51	0.64	0.73	0.62	0.54	0.68	0.68	0.58	0.68
Chinese Taipei	0.83	34	0.59	0.66	0.60	0.59	0.54	0.53	0.63	0.61	0.60	0.51	0.63	0.58	0.53	0.57
Cyprus	0.87	39	0.60	0.64	0.54	0.61	0.64	0.54	0.64	0.67	0.69	0.58	0.70	0.64	0.61	0.67
Egypt	0.92	50	0.64	0.63	0.58	0.62	0.72	0.66	0.73	0.74	0.78	0.78	0.76	0.78	0.72	0.74
England	0.90	44	0.67	0.73	0.68	0.66	0.66	0.57	0.63	0.75	0.73	0.54	0.73	0.64	0.64	0.66
Finland	0.88	43	0.67	0.70	0.63	0.62	0.61	0.61	0.61	0.72	0.72	0.58	0.70	0.64	0.62	0.68
France	0.83	33	0.63	0.68	0.43	0.59	0.62	0.49	0.53	0.62	0.56	0.39	0.65	0.63	0.56	0.59
Georgia	0.84	38	0.56	0.62	0.58	0.59	0.63	0.50	0.63	0.65	0.70	0.67	0.68	0.59	0.64	0.61
Hong Kong SAR	0.92	50	0.64	0.70	0.60	0.66	0.69	0.66	0.76	0.79	0.79	0.68	0.76	0.73	0.67	0.76
Hungary	0.84	35	0.64	0.67	0.57	0.53	0.63	0.50	0.59	0.69	0.59	0.45	0.63	0.60	0.54	0.64
Iran, Islamic Rep. of	0.85	38	0.58	0.64	0.59	0.38	0.67	0.55	0.67	0.59	0.66	0.68	0.72	0.66	0.49	0.66
Ireland	0.88	40	0.66	0.73	0.63	0.62	0.64	0.53	0.60	0.69	0.69	0.52	0.68	0.60	0.60	0.66
Israel	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Italy	0.83	33	0.61	0.67	0.58	0.59	0.57	0.43	0.56	0.63	0.57	0.42	0.59	0.57	0.58	0.62
Japan	0.80	31	0.58	0.66	0.63	0.62	0.53	0.49	0.64	0.51	0.47	0.42	0.52	0.50	0.57	0.56
Jordan	0.92	51	0.65	0.63	0.60	0.67	0.74	0.64	0.75	0.70	0.79	0.76	0.78	0.78	0.75	0.75
Kazakhstan	0.78	34	0.56	0.61	0.55	0.59	0.59	0.47	0.55	0.28	0.66	0.55	0.61	0.63	0.66	0.68
Korea, Rep. of	0.82	35	0.53	0.67	0.61	0.48	0.55	0.50	-	0.69	0.66	0.57	0.63	0.60	0.57	0.62
Kuwait	0.89	44	0.62	0.61	0.63	0.62	0.67	0.56	0.65	0.67	0.72	0.65	0.73	0.72	0.69	0.69
Lebanon	0.88	41	0.62	0.59	0.43	0.58	0.64	0.58	0.64	0.69	0.72	0.67	0.72	0.71	0.63	0.65
Lithuania	0.85	35	0.66	0.67	0.60	0.60	0.63	0.45	0.52	0.65	0.57	0.48	0.63	0.64	0.56	0.58
Malaysia	0.82	30	0.46	0.56	0.46	0.48	0.53	0.51	0.58	0.63	0.64	0.50	0.58	0.54	0.56	0.58
Morocco	0.85	36	0.52	0.58	0.55	0.59	0.64	0.47	0.63	0.58	0.67	0.62	0.65	0.66	0.61	0.62
New Zealand	0.90	44	0.65	0.73	0.67	0.65	0.63	0.60	0.65	0.74	0.75	0.54	0.72	0.61	0.66	0.69
Norway (9)	0.89	42	0.68	0.73	0.67	0.63	0.59	0.55	0.62	0.72	0.72	0.58	0.68	0.59	0.61	0.64
Oman	0.88	41	0.62	0.60	0.61	0.55	0.63	0.53	0.64	0.69	0.71	0.67	0.70	0.67	0.62	0.68
Portugal	0.86	39	0.59	0.66	0.61	0.60	0.62	0.56	0.61	0.66	0.62	0.56	0.68	0.64	0.60	0.67
Qatar	0.92	52	0.66	0.70	0.67	0.68	0.72	0.65	0.72	0.74	0.78	0.73	0.78	0.75	0.70	0.75
Romania	0.90	44	0.66	0.68	0.63	0.63	0.67	0.58	0.62	0.76	0.71	0.60	0.72	0.68	0.65	0.69
Russian Federation	0.86	38	0.63	0.65	0.60	0.52	0.64	0.51	0.59	0.68	0.69	0.55	0.66	0.63	0.63	0.60
Saudi Arabia	0.90	46	0.64	0.65	0.63	0.66	0.73	0.58	0.70	0.62	0.69	0.69	0.75	0.73	0.70	0.71
Singapore	0.89	41	0.64	0.72	0.60	0.61	0.65	0.59	0.62	0.74	0.69	0.52	0.69	0.63	0.61	0.68
South Africa (9)	0.84	33	0.52	0.54	0.51	0.56	0.59	0.32	0.59	0.67	0.69	0.58	0.62	0.62	0.58	0.59
Sweden	0.89	44	0.67	0.70	0.64	0.65	0.62	0.56	0.64	0.74	0.76	0.57	0.69	0.64	0.65	0.71
Turkey	0.84	36	0.52	0.60	0.58	0.53	0.59	0.51	0.51	0.64	0.67	0.59	0.68	0.67	0.61	0.68
United Arab Emirates	0.92	51	0.67	0.70	0.64	0.66	0.72	0.66	0.72	0.76	0.76	0.70	0.77	0.74	0.70	0.75
United States	0.90	46	0.68	0.72	0.67	0.68	0.68	0.60	0.64	0.73	0.73	0.57	0.71	0.62	0.68	0.70

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Student Bullying* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item													
			BSBG14A	BSBG14B	BSBG14C	BSBG14D	BSBG14E	BSBG14F	BSBG14G	BSBG14H	BSBG14I	BSBG14J	BSBG14K	BSBG14L	BSBG14M	BSBG14N
Benchmarking Participants																
Ontario, Canada	0.89	43	0.68	0.73	0.66	0.65	0.65	0.55	0.64	0.73	0.72	0.59	0.66	0.61	0.64	0.66
Quebec, Canada	0.87	39	0.68	0.70	0.62	0.61	0.57	0.51	0.62	0.71	0.66	0.51	0.67	0.63	0.62	0.61
Moscow City, Russian Fed.	0.85	36	0.62	0.63	0.60	0.50	0.63	0.47	0.52	0.66	0.70	0.54	0.62	0.64	0.60	0.61
Gauteng, RSA (9)	0.83	33	0.55	0.59	0.53	0.55	0.56	0.37	0.56	0.67	0.65	0.53	0.63	0.62	0.53	0.62
Western Cape, RSA (9)	0.86	38	0.58	0.62	0.58	0.61	0.61	0.42	0.60	0.72	0.72	0.57	0.67	0.62	0.59	0.64
Abu Dhabi, UAE	0.93	54	0.71	0.72	0.66	0.68	0.73	0.69	0.74	0.79	0.80	0.74	0.79	0.75	0.73	0.76
Dubai, UAE	0.90	45	0.63	0.68	0.61	0.65	0.68	0.63	0.67	0.71	0.69	0.63	0.71	0.70	0.63	0.73

A dash (–) indicates comparable data not available.

Relationship Between the TIMSS 2019 *Student Bullying* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Australia	0.15	0.02	0.03	0.16	0.02	0.04
Bahrain	0.11	0.01	0.02	0.20	0.04	0.08
Chile	0.08	0.01	0.02	0.12	0.01	0.03
Chinese Taipei	-0.02	0.00	0.00	-0.01	0.00	0.00
Cyprus	0.09	0.01	0.01	0.10	0.01	0.02
Egypt	0.24	0.06	0.08	0.27	0.07	0.10
England	0.13	0.02	0.03	0.12	0.01	0.03
Finland	0.05	0.00	0.02	0.06	0.00	0.02
France	0.03	0.00	0.01	0.06	0.00	0.01
Georgia	0.04	0.00	0.01	0.04	0.00	0.02
Hong Kong SAR	0.03	0.00	0.01	0.04	0.00	0.01
Hungary	0.15	0.02	0.03	0.13	0.02	0.02
Iran, Islamic Rep. of	0.16	0.03	0.04	0.18	0.03	0.05
Ireland	0.08	0.01	0.02	0.09	0.01	0.02
Israel	-	-	-	-	-	-
Italy	0.07	0.01	0.02	0.09	0.01	0.01
Japan	-0.02	0.00	0.00	0.01	0.00	0.00
Jordan	0.20	0.04	0.06	0.27	0.07	0.11
Kazakhstan	-0.06	0.00	0.00	-0.10	0.01	0.00
Korea, Rep. of	-0.06	0.00	0.00	-0.03	0.00	0.00
Kuwait	0.07	0.00	0.02	0.12	0.01	0.05
Lebanon	0.11	0.01	0.02	0.14	0.02	0.03
Lithuania	0.03	0.00	0.01	0.02	0.00	0.01
Malaysia	0.12	0.02	0.01	0.12	0.02	0.02
Morocco	0.09	0.01	0.02	0.12	0.01	0.03
New Zealand	0.10	0.01	0.02	0.11	0.01	0.02
Norway (9)	0.06	0.00	0.02	0.05	0.00	0.02
Oman	0.17	0.03	0.06	0.21	0.04	0.08
Portugal	0.05	0.00	0.00	0.04	0.00	0.00
Qatar	0.12	0.01	0.06	0.20	0.04	0.09
Romania	0.08	0.01	0.02	0.07	0.01	0.02
Russian Federation	-0.01	0.00	0.00	-0.01	0.00	0.00
Saudi Arabia	0.14	0.02	0.04	0.21	0.04	0.08
Singapore	0.14	0.02	0.03	0.13	0.02	0.02
South Africa (9)	0.24	0.06	0.08	0.29	0.08	0.10
Sweden	0.10	0.01	0.02	0.08	0.01	0.02
Turkey	0.08	0.01	0.01	0.10	0.01	0.02
United Arab Emirates	0.20	0.04	0.09	0.26	0.07	0.13
United States	0.10	0.01	0.02	0.09	0.01	0.02
International Median	**0.09**	**0.01**	**0.02**	**0.11**	**0.01**	**0.02**

Relationship Between the TIMSS 2019 *Student Bullying* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Benchmarking Participants						
Ontario, Canada	0.07	0.00	0.01	0.09	0.01	0.01
Quebec, Canada	0.13	0.02	0.02	0.11	0.01	0.01
Moscow City, Russian Fed.	0.09	0.01	0.02	0.09	0.01	0.01
Gauteng, RSA (9)	0.26	0.07	0.07	0.31	0.10	0.10
Western Cape, RSA (9)	0.18	0.03	0.05	0.22	0.05	0.07
Abu Dhabi, UAE	0.27	0.07	0.13	0.31	0.10	0.18
Dubai, UAE	0.12	0.01	0.03	0.16	0.02	0.04

A dash (–) indicates comparable data not available.

Students Confident in Biology – Grade 8

About the Scale

The *Students Confident in Biology* scale was created based on students' responses to eight items listed below.

How much do you agree with these statements about biology?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BSBB25A	1) I usually do well in biology	○	○	○	○
BSBB25B	2) Biology is more difficult for me than for many of my classmates [R]	○	○	○	○
BSBB25C	3) Biology is not one of my strengths [R]	○	○	○	○
BSBB25D	4) I learn things quickly in biology	○	○	○	○
BSBB25E	5) I am good at working out difficult biology problems	○	○	○	○
BSBB25F	6) My teacher tells me I am good at biology	○	○	○	○
BSBB25G	7) Biology is harder for me than any other subject [R]	○	○	○	○
BSBB25H	8) Biology makes me confused [R]	○	○	○	○

Very Confident	Somewhat Confident	Not Confident

Scale Cut Scores **11.3** **8.8**

[R] Reverse coded

Item Parameters for the TIMSS 2019 *Students Confident in Biology* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBB25A	-0.51693	-1.44764	-0.43643	1.88407	0.87
BSBB25B R	-0.10746	-1.36535	-0.13132	1.49667	1.08
BSBB25C R	0.40263	-1.44249	0.06347	1.37902	0.99
BSBB25D	-0.20096	-1.67008	-0.11982	1.78990	0.87
BSBB25E	0.43707	-1.79371	0.01706	1.77665	1.01
BSBB25F	0.49259	-1.57640	-0.09570	1.67210	1.12
BSBB25G R	-0.26525	-1.05911	-0.23902	1.29813	1.08
BSBB25H R	-0.24169	-1.02437	-0.19301	1.21738	1.06

R Reverse coded

Scale Transformation Constants for the TIMSS 2019 *Students Confident in Biology* Scale – Grade 8

Scale Transformation Constants	
A = 8.851451	Transformed Scale Score = 8.851451 + 1.356973 • Logit Scale Score
B = 1.356973	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Students Confident in Biology* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	3.04975	
1	4.61732	
2	5.39034	
3	5.93525	
4	6.36886	
5	6.73771	
6	7.07124	
7	7.37840	
8	7.66831	
9	7.94767	
10	8.22161	
11	8.49411	
12	8.76836	8.8
13	9.04715	
14	9.33297	
15	9.62841	
16	9.93642	
17	10.25945	
18	10.60451	
19	10.97905	
20	11.39630	11.3
21	11.87989	
22	12.47348	
23	13.30261	
24	14.92563	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Confident in Biology Scale – Grade 8*

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item: BSBB25A	BSBB25B R	BSBB25C R	BSBB25D	BSBB25E	BSBB25F	BSBB25G R	BSBB25H R
Cyprus	0.89	57	0.79	0.71	0.75	0.80	0.78	0.73	0.74	0.75
Finland	0.86	51	0.81	0.71	0.77	0.80	0.71	0.67	0.65	0.55
France	0.88	55	0.80	0.66	0.76	0.83	0.76	0.60	0.71	0.80
Georgia	0.84	48	0.74	0.64	0.72	0.75	0.72	0.67	0.64	0.63
Hungary	0.89	56	0.80	0.69	0.81	0.80	0.74	0.74	0.74	0.67
Kazakhstan	0.87	53	0.76	0.67	0.78	0.77	0.77	0.69	0.69	0.67
Lebanon	0.80	42	0.69	0.59	0.51	0.76	0.74	0.71	0.61	0.53
Lithuania	0.87	54	0.81	0.71	0.76	0.79	0.75	0.60	0.74	0.66
Morocco	0.76	38	0.66	0.61	0.47	0.72	0.67	0.64	0.58	0.57
Portugal	0.84	48	0.74	0.61	0.68	0.80	0.74	0.62	0.63	0.70
Romania	0.80	43	0.73	0.62	0.66	0.73	0.58	0.62	0.60	0.67
Russian Federation	0.88	54	0.75	0.70	0.81	0.81	0.76	0.69	0.70	0.67
Sweden	0.87	53	0.78	0.72	0.76	0.80	0.74	0.66	0.69	0.70
Benchmarking Participants										
Moscow City, Russian Fed.	0.89	57	0.77	0.74	0.83	0.85	0.78	0.68	0.74	0.67

R Reverse coded

Relationship Between the TIMSS 2019 *Students Confident in Biology Scale and TIMSS 2019 Achievement – Grade 8*

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Cyprus	0.34	0.12	0.13
Finland	0.32	0.11	0.12
France	0.33	0.11	0.11
Georgia	0.30	0.09	0.08
Hungary	0.24	0.06	0.06
Kazakhstan	0.19	0.04	0.03
Lebanon	0.34	0.12	0.12
Lithuania	0.23	0.05	0.05
Morocco	0.34	0.11	0.10
Portugal	0.36	0.13	0.13
Romania	0.20	0.04	0.05
Russian Federation	0.15	0.02	0.02
Sweden	0.33	0.11	0.11
International Median	**0.32**	**0.11**	**0.10**
Benchmarking Participants			
Moscow City, Russian Fed.	0.11	0.01	0.01

Scale was created for TIMSS 2019 countries where science is taught as separate subjects.

Students Confident in Chemistry – Grade 8

About the Scale

The *Students Confident in Chemistry* scale was created based on students' responses to eight items listed below.

How much do you agree with these statements about chemistry?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BSBC35A	1) I usually do well in chemistry	○	○	○	○
BSBC35B	2) Chemistry is more difficult for me than for many of my classmates [R]	○	○	○	○
BSBC35C	3) Chemistry is not one of my strengths [R]	○	○	○	○
BSBC35D	4) I learn things quickly in chemistry	○	○	○	○
BSBC35E	5) I am good at working out difficult chemistry problems	○	○	○	○
BSBC35F	6) My teacher tells me I am good at chemistry	○	○	○	○
BSBC35G	7) Chemistry is harder for me than any other subject [R]	○	○	○	○
BSBC35H	8) Chemistry makes me confused [R]	○	○	○	○

Very Confident	Somewhat Confident	Not Confident

Scale Cut Scores 11.5 9.3

[R] Reverse coded

Item Parameters for the TIMSS 2019 *Students Confident in Chemistry* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBC35A	-0.45681	-1.43599	-0.23074	1.66673	0.87
BSBC35B [R]	-0.20723	-1.36687	-0.01365	1.38052	1.11
BSBC35C [R]	0.30893	-1.38095	0.15716	1.22379	1.01
BSBC35D	-0.09496	-1.62194	0.02911	1.59283	0.87
BSBC35E	0.41467	-1.58287	0.08541	1.49746	0.97
BSBC35F	0.33594	-1.48997	-0.01839	1.50836	1.08
BSBC35G [R]	-0.12895	-1.14551	-0.09743	1.24294	1.14
BSBC35H [R]	-0.17159	-1.06641	-0.06190	1.12831	1.19

[R] Reverse coded

Scale Transformation Constants for the TIMSS 2019 *Students Confident in Chemistry* Scale – Grade 8

Scale Transformation Constants	
A = 9.287425	Transformed Scale Score = 9.287425 + 1.273947 • Logit Scale Score
B = 1.273947	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Students Confident in Chemistry* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	3.89635	
1	5.37931	
2	6.11672	
3	6.63419	
4	7.04816	
5	7.40098	
6	7.71515	
7	8.00298	
8	8.27440	
9	8.53331	
10	8.78409	
11	9.03015	
12	9.27400	9.3
13	9.51933	
14	9.76754	
15	10.02173	
16	10.28517	
17	10.56193	
18	10.85539	
19	11.17623	
20	11.53629	11.5
21	11.95803	
22	12.48347	
23	13.22968	
24	14.72190	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Confident in Chemistry* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			BSBC35A	BSBC35B R	BSBC35C R	BSBC35D	BSBC35E	BSBC35F	BSBC35G R	BSBC35H R
Cyprus	0.89	57	0.77	0.74	0.74	0.80	0.77	0.74	0.73	0.75
Finland	0.90	59	0.85	0.75	0.82	0.84	0.80	0.74	0.72	0.56
France	0.89	57	0.83	0.68	0.80	0.84	0.79	0.63	0.67	0.78
Georgia	0.86	51	0.77	0.67	0.72	0.78	0.75	0.73	0.65	0.65
Hungary	0.90	58	0.84	0.67	0.80	0.83	0.76	0.77	0.73	0.67
Kazakhstan	0.88	54	0.80	0.67	0.78	0.80	0.80	0.76	0.63	0.64
Lebanon	0.80	42	0.70	0.61	0.51	0.75	0.73	0.72	0.59	0.55
Lithuania	0.90	59	0.85	0.74	0.78	0.83	0.82	0.68	0.77	0.65
Morocco	0.73	36	0.68	0.53	0.42	0.74	0.68	0.70	0.44	0.49
Portugal	0.89	57	0.79	0.66	0.77	0.84	0.81	0.71	0.71	0.72
Romania	0.84	47	0.78	0.56	0.59	0.79	0.75	0.75	0.59	0.63
Russian Federation	0.90	60	0.80	0.74	0.83	0.85	0.81	0.79	0.73	0.63
Sweden	0.88	54	0.79	0.68	0.76	0.82	0.76	0.68	0.68	0.69
Benchmarking Participants										
Moscow City, Russian Fed.	0.91	62	0.82	0.77	0.84	0.85	0.83	0.75	0.74	0.67

R Reverse coded

Relationship Between the TIMSS 2019 *Students Confident in Chemistry* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Cyprus	0.33	0.11	0.12
Finland	0.40	0.16	0.17
France	0.34	0.11	0.13
Georgia	0.34	0.11	0.12
Hungary	0.19	0.04	0.05
Kazakhstan	0.19	0.04	0.04
Lebanon	0.32	0.10	0.10
Lithuania	0.28	0.08	0.09
Morocco	0.34	0.12	0.10
Portugal	0.37	0.14	0.14
Romania	0.20	0.04	0.07
Russian Federation	0.24	0.06	0.06
Sweden	0.34	0.11	0.11
International Median	**0.33**	**0.11**	**0.10**
Benchmarking Participants			
Moscow City, Russian Fed.	0.24	0.06	0.06

Scale was created for TIMSS 2019 countries where science is taught as separate subjects.

Students Confident in Earth Science – Grade 8

About the Scale

The *Students Confident in Earth Science* scale was created based on students' responses to nine items listed below.

How much do you agree with these statements about earth science?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BSBE30A	1) I usually do well in earth science	○	○	○	○
BSBE30B	2) Earth science is more difficult for me than for many of my classmates [R]	○	○	○	○
BSBE30C	3) Earth science is not one of my strengths [R]	○	○	○	○
BSBE30D	4) I learn things quickly in earth science	○	○	○	○
BSBE30E	5) I am good at working out difficult earth science problems	○	○	○	○
BSBE30F	6) My teacher tells me I am good at earth science	○	○	○	○
BSBE30G	7) Earth science is harder for me than any other subject [R]	○	○	○	○
BSBE30H	8) Earth science makes me confused [R]	○	○	○	○

Very Confident		Somewhat Confident		Not Confident
Scale Cut Scores	11.4		8.9	

[R] Reverse coded

Item Parameters for the TIMSS 2019 *Students Confident in Earth Science* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBE30A	-0.46107	-1.41872	-0.43930	1.85802	0.87
BSBE30B R	-0.19875	-1.37648	-0.15004	1.52652	1.09
BSBE30C R	0.33686	-1.49366	0.07644	1.41722	0.96
BSBE30D	-0.12943	-1.65316	-0.17396	1.82712	0.88
BSBE30E	0.45259	-1.79586	0.02998	1.76588	1.01
BSBE30F	0.55212	-1.58204	-0.08004	1.66208	1.15
BSBE30G R	-0.24372	-1.06339	-0.29194	1.35533	1.10
BSBE30H R	-0.30860	-1.02496	-0.22507	1.25003	1.10

R Reverse coded

Scale Transformation Constants for the TIMSS 2019 *Students Confident in Earth Science* Scale – Grade 8

Scale Transformation Constants	
A = 8.904350	Transformed Scale Score = 8.904350 + 1.335162 • Logit Scale Score
B = 1.335162	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Students Confident in Earth Science* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	3.21094	
1	4.74712	
2	5.50216	
3	6.03282	
4	6.45453	
5	6.81430	
6	7.13915	
7	7.43625	
8	7.71978	
9	7.99591	
10	8.26570	
11	8.53520	
12	8.80751	8.9
13	9.08518	
14	9.37050	
15	9.66574	
16	9.97356	
17	10.29618	
18	10.64027	
19	11.01305	
20	11.42745	11.4
21	11.90664	
22	12.49348	
23	13.31156	
24	14.91022	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Confident in Earth Science Scale – Grade 8*

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			BSBE30A	BSBE30B R	BSBE30C R	BSBE30D	BSBE30E	BSBE30F	BSBE30G R	BSBE30H R
Cyprus	0.88	55	0.79	0.69	0.73	0.80	0.76	0.73	0.70	0.74
Finland	0.86	50	0.81	0.72	0.78	0.81	0.73	0.65	0.62	0.52
France	0.88	55	0.80	0.66	0.76	0.83	0.76	0.60	0.71	0.80
Georgia	0.83	45	0.72	0.65	0.73	0.74	0.68	0.67	0.58	0.59
Hungary	0.89	57	0.82	0.68	0.81	0.81	0.74	0.75	0.74	0.66
Kazakhstan	0.86	51	0.79	0.64	0.79	0.76	0.76	0.68	0.63	0.61
Lebanon	-	-	-	-	-	-	-	-	-	-
Lithuania	0.89	56	0.83	0.71	0.78	0.81	0.78	0.66	0.76	0.64
Morocco	0.74	36	0.66	0.56	0.43	0.72	0.68	0.68	0.49	0.51
Portugal	0.84	48	0.74	0.61	0.68	0.80	0.74	0.62	0.63	0.70
Romania	0.77	39	0.65	0.59	0.63	0.70	0.52	0.62	0.60	0.66
Russian Federation	0.87	52	0.75	0.70	0.78	0.81	0.75	0.69	0.67	0.61
Sweden	-	-	-	-	-	-	-	-	-	-
Benchmarking Participants										
Moscow City, Russian Fed.	0.89	57	0.78	0.74	0.82	0.84	0.77	0.70	0.70	0.65

R Reverse coded

Relationship Between the TIMSS 2019 *Students Confident in Earth Science Scale and TIMSS 2019 Achievement – Grade 8*

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Cyprus	0.31	0.10	0.10
Finland	0.32	0.10	0.11
France	0.33	0.11	0.11
Georgia	0.34	0.11	0.12
Hungary	0.24	0.06	0.06
Kazakhstan	0.20	0.04	0.04
Lebanon	-	-	-
Lithuania	0.21	0.05	0.05
Morocco	0.29	0.09	0.08
Portugal	0.36	0.13	0.13
Romania	0.23	0.05	0.06
Russian Federation	0.16	0.03	0.03
Sweden	-	-	-
International Median	**0.29**	**0.09**	**0.08**
Benchmarking Participants			
Moscow City, Russian Fed.	0.17	0.03	0.03

Scale was created for TIMSS 2019 countries where science is taught as separate subjects.
A dash (–) indicates comparable data not available.

Students Confident in Mathematics – Grade 8

About the Scale

The *Students Confident in Mathematics* scale was created based on students' responses to nine items listed below.

How much do you agree with these statements about mathematics?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BSBM19A T	1) I usually do well in mathematics	○	○	○	○
BSBM19B T	2) Mathematics is more difficult for me than for many of my classmates R	○	○	○	○
BSBM19C T	3) Mathematics is not one of my strengths R	○	○	○	○
BSBM19D T	4) I learn things quickly in mathematics	○	○	○	○
BSBM19E T	5) Mathematics makes me nervous R	○	○	○	○
BSBM19F T	6) I am good at working out difficult mathematics problems	○	○	○	○
BSBM19G T	7) My teacher tells me I am good at mathematics	○	○	○	○
BSBM19H T	8) Mathematics is harder for me than any other subject R	○	○	○	○
BSBM19I T	9) Mathematics makes me confused R	○	○	○	○

Very Confident | Somewhat Confident | Not Confident

Scale Cut Scores 12.1 9.5

R Reverse coded

T Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Students Confident in Mathematics* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBM19A	-0.50049	-1.11873	-0.29405	1.41278	0.91
BSBM19B R	-0.00401	-1.14715	0.08367	1.06348	1.02
BSBM19C R	0.30051	-0.87683	0.15214	0.72469	0.91
BSBM19D	-0.22037	-1.32473	-0.05161	1.37634	0.90
BSBM19E R	-0.10393	-0.97276	0.03057	0.94219	1.26
BSBM19F	0.30742	-1.30327	-0.11911	1.42238	0.95
BSBM19G	0.07317	-1.04695	-0.15298	1.19993	1.13
BSBM19H R	0.07380	-0.69985	-0.01041	0.71026	0.93
BSBM19I R	0.07390	-0.96460	0.10241	0.86219	0.97

R Reverse coded

Scale Transformation Constants for the TIMSS 2019 *Students Confident in Mathematics* Scale – Grade 8

Scale Transformation Constants	
A = 9.553292	Transformed Scale Score = 9.553292 + 1.562014 • Logit Scale Score
B = 1.562014	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Students Confident in Mathematics* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	3.27793	
1	5.04751	
2	5.90622	
3	6.49606	
4	6.95612	
5	7.34181	
6	7.67800	
7	7.97991	
8	8.25602	
9	8.51577	
10	8.76168	
11	8.99736	
12	9.22591	
13	9.45004	9.5
14	9.67230	
15	9.89514	
16	10.12116	
17	10.35323	
18	10.59469	
19	10.84951	
20	11.12082	
21	11.41779	
22	11.75008	
23	12.13375	12.1
24	12.59517	
25	13.19086	
26	14.06165	
27	15.85188	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Confident in Mathematics* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			BSBM19A	BSBM19B R	BSBM19C R	BSBM19D	BSBM19E R	BSBM19F	BSBM19G	BSBM19H R	BSBM19I R
Australia	0.90	56	0.78	0.76	0.84	0.80	0.58	0.80	0.53	0.79	0.78
Bahrain	0.85	45	0.58	0.64	0.69	0.70	0.63	0.66	0.59	0.76	0.73
Chile	0.87	49	0.75	0.61	0.72	0.76	0.54	0.75	0.64	0.73	0.75
Chinese Taipei	0.92	61	0.82	0.79	0.85	0.83	0.55	0.80	0.79	0.80	0.78
Cyprus	0.91	59	0.81	0.77	0.82	0.78	0.60	0.79	0.71	0.81	0.81
Egypt	0.75	33	0.31	0.65	0.55	0.49	0.68	0.47	0.44	0.73	0.70
England	0.88	51	0.74	0.73	0.82	0.75	0.58	0.74	0.44	0.80	0.77
Finland	0.91	58	0.84	0.76	0.86	0.82	0.63	0.79	0.67	0.77	0.66
France	0.91	59	0.84	0.72	0.84	0.84	0.57	0.76	0.67	0.80	0.85
Georgia	0.84	45	0.74	0.67	0.77	0.71	0.44	0.65	0.65	0.68	0.67
Hong Kong SAR	0.89	54	0.73	0.73	0.82	0.72	0.62	0.74	0.62	0.81	0.77
Hungary	0.91	58	0.84	0.70	0.82	0.83	0.57	0.79	0.77	0.78	0.75
Iran, Islamic Rep. of	0.87	49	0.76	0.55	0.68	0.75	0.68	0.73	0.70	0.70	0.74
Ireland	0.90	56	0.80	0.78	0.79	0.81	0.60	0.78	0.51	0.82	0.79
Israel	0.88	50	0.75	0.71	0.75	0.74	0.62	0.70	0.61	0.76	0.73
Italy	0.93	63	0.83	0.77	0.86	0.85	0.67	0.79	0.75	0.81	0.79
Japan	0.90	55	0.77	0.69	0.81	0.73	0.63	0.79	0.66	0.77	0.77
Jordan	0.79	36	0.34	0.67	0.67	0.55	0.64	0.54	0.50	0.72	0.71
Kazakhstan	0.89	54	0.79	0.73	0.79	0.81	0.54	0.76	0.73	0.71	0.72
Korea, Rep. of	0.90	56	0.85	0.77	0.78	0.75	0.38	0.83	0.78	0.77	0.72
Kuwait	0.79	37	0.45	0.62	0.61	0.62	0.61	0.64	0.55	0.68	0.69
Lebanon	0.83	42	0.69	0.59	0.57	0.70	0.60	0.68	0.66	0.68	0.62
Lithuania	0.90	57	0.83	0.73	0.82	0.81	0.66	0.79	0.66	0.80	0.68
Malaysia	0.81	39	0.62	0.53	0.67	0.58	0.62	0.66	0.57	0.66	0.71
Morocco	0.79	37	0.63	0.57	0.49	0.69	0.59	0.60	0.64	0.61	0.64
New Zealand	0.87	51	0.75	0.74	0.79	0.76	0.57	0.77	0.47	0.78	0.72
Norway (9)	0.92	61	0.81	0.78	0.87	0.84	0.69	0.82	0.63	0.81	0.76
Oman	0.79	37	0.42	0.70	0.64	0.56	0.68	0.44	0.45	0.74	0.71
Portugal	0.89	55	0.77	0.64	0.80	0.84	0.46	0.82	0.72	0.76	0.78
Qatar	0.82	40	0.61	0.64	0.66	0.63	0.60	0.61	0.55	0.73	0.67
Romania	0.84	46	0.73	0.69	0.66	0.76	0.27	0.74	0.68	0.73	0.69
Russian Federation	0.91	58	0.79	0.75	0.85	0.83	0.57	0.78	0.73	0.77	0.73
Saudi Arabia	0.81	39	0.43	0.67	0.60	0.65	0.64	0.60	0.56	0.74	0.70
Singapore	0.91	58	0.83	0.77	0.87	0.77	0.62	0.78	0.60	0.82	0.77
South Africa (9)	0.78	37	0.65	0.49	0.63	0.64	0.46	0.64	0.60	0.63	0.66
Sweden	0.91	58	0.81	0.77	0.83	0.82	0.54	0.81	0.63	0.82	0.78
Turkey	0.89	54	0.79	0.66	0.80	0.72	0.62	0.73	0.72	0.76	0.75
United Arab Emirates	0.84	43	0.64	0.67	0.70	0.64	0.62	0.63	0.55	0.74	0.69
United States	0.90	56	0.78	0.77	0.83	0.79	0.62	0.77	0.55	0.82	0.78

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Confident in Mathematics* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			BSBM19A	BSBM19B R	BSBM19C R	BSBM19D	BSBM19E R	BSBM19F	BSBM19G	BSBM19H R	BSBM19I R
Benchmarking Participants											
Ontario, Canada	0.91	60	0.81	0.81	0.86	0.81	0.64	0.78	0.52	0.85	0.81
Quebec, Canada	0.92	62	0.85	0.81	0.89	0.84	0.65	0.81	0.59	0.81	0.83
Moscow City, Russian Fed.	0.92	61	0.80	0.78	0.87	0.85	0.61	0.80	0.69	0.81	0.76
Gauteng, RSA (9)	0.83	43	0.70	0.57	0.69	0.68	0.52	0.67	0.63	0.70	0.71
Western Cape, RSA (9)	0.86	47	0.73	0.63	0.74	0.69	0.58	0.68	0.65	0.73	0.71
Abu Dhabi, UAE	0.80	38	0.54	0.66	0.68	0.55	0.62	0.55	0.47	0.73	0.68
Dubai, UAE	0.87	50	0.74	0.72	0.78	0.70	0.64	0.69	0.58	0.76	0.72

R Reverse coded

Relationship Between the TIMSS 2019 *Students Confident in Mathematics* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Australia	0.49	0.24	0.23
Bahrain	0.32	0.10	0.09
Chile	0.36	0.13	0.13
Chinese Taipei	0.49	0.24	0.23
Cyprus	0.52	0.27	0.26
Egypt	0.28	0.08	0.08
England	0.43	0.18	0.16
Finland	0.55	0.30	0.29
France	0.56	0.31	0.30
Georgia	0.47	0.22	0.20
Hong Kong SAR	0.32	0.10	0.11
Hungary	0.55	0.30	0.30
Iran, Islamic Rep. of	0.43	0.19	0.18
Ireland	0.43	0.18	0.18
Israel	0.39	0.15	0.15
Italy	0.50	0.25	0.25
Japan	0.46	0.21	0.19
Jordan	0.37	0.14	0.13
Kazakhstan	0.29	0.08	0.09
Korea, Rep. of	0.49	0.24	0.22
Kuwait	0.30	0.09	0.08
Lebanon	0.37	0.14	0.13
Lithuania	0.49	0.24	0.24
Malaysia	0.23	0.05	0.07
Morocco	0.34	0.11	0.11
New Zealand	0.46	0.21	0.20
Norway (9)	0.62	0.39	0.36
Oman	0.38	0.15	0.13
Portugal	0.52	0.27	0.27
Qatar	0.35	0.12	0.13
Romania	0.46	0.21	0.21
Russian Federation	0.43	0.18	0.18
Saudi Arabia	0.37	0.14	0.12
Singapore	0.40	0.16	0.15
South Africa (9)	0.26	0.07	0.09
Sweden	0.56	0.31	0.29
Turkey	0.46	0.21	0.22
United Arab Emirates	0.31	0.10	0.10
United States	0.42	0.18	0.18
International Median	**0.43**	**0.18**	**0.18**

Relationship Between the TIMSS 2019 *Students Confident in Mathematics* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Benchmarking Participants			
Ontario, Canada	0.56	0.32	0.31
Quebec, Canada	0.54	0.29	0.26
Moscow City, Russian Fed.	0.53	0.28	0.27
Gauteng, RSA (9)	0.33	0.11	0.12
Western Cape, RSA (9)	0.36	0.13	0.17
Abu Dhabi, UAE	0.34	0.11	0.12
Dubai, UAE	0.36	0.13	0.12

Students Confident in Physics – Grade 8

About the Scale

The *Students Confident in Physics* scale was created based on students' responses to eight items listed below.

How much do you agree with these statements about physics?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BSBP40A	1) I usually do well in physics	◯	◯	◯	◯
BSBP40B	2) Physics is more difficult for me than for many of my classmates [R]	◯	◯	◯	◯
BSBP40C	3) Physics is not one of my strengths [R]	◯	◯	◯	◯
BSBP40D	4) I learn things quickly in physics	◯	◯	◯	◯
BSBP40E	5) I am good at working out difficult physics problems	◯	◯	◯	◯
BSBP40F	6) My teacher tells me I am good at physics	◯	◯	◯	◯
BSBP40G	7) Physics is harder for me than any other subject [R]	◯	◯	◯	◯
BSBP40H	8) Physics makes me confused [R]	◯	◯	◯	◯

Very Confident	Somewhat Confident	Not Confident

Scale Cut Scores **11.7** **9.5**

[R] Reverse coded

Item Parameters for the TIMSS 2019 *Students Confident in Physics* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBP40A	-0.39746	-1.37025	-0.24044	1.61069	0.88
BSBP40B R	-0.21128	-1.31428	-0.01234	1.32662	1.10
BSBP40C R	0.28620	-1.30084	0.14053	1.16031	0.98
BSBP40D	-0.07362	-1.58237	-0.00919	1.59156	0.86
BSBP40E	0.39820	-1.51916	0.09858	1.42058	0.96
BSBP40F	0.29764	-1.44459	-0.00874	1.45333	1.07
BSBP40G R	-0.11946	-1.11132	-0.09117	1.20249	1.14
BSBP40H R	-0.18022	-1.01891	-0.05197	1.07088	1.20

R Reverse coded

Scale Transformation Constants for the TIMSS 2019 *Students Confident in Physics* Scale – Grade 8

Scale Transformation Constants	
A = 9.428686	Transformed Scale Score = 9.428686 + 1.324891 • Logit Scale Score
B = 1.324891	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Students Confident in Physics* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	3.90479	
1	5.44008	
2	6.20005	
3	6.73122	
4	7.15482	
5	7.51483	
6	7.83462	
7	8.12702	
8	8.40231	
9	8.66453	
10	8.91821	
11	9.16694	
12	9.41328	9.5
13	9.66135	
14	9.91242	
15	10.16987	
16	10.43713	
17	10.71844	
18	11.01730	
19	11.34485	
20	11.71341	11.7
21	12.14624	
22	12.68694	
23	13.45737	
24	15.00359	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Confident in Physics* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			BSBP40A	BSBP40B R	BSBP40C R	BSBP40D	BSBP40E	BSBP40F	BSBP40G R	BSBP40H R
Cyprus	0.91	61	0.81	0.74	0.75	0.83	0.82	0.76	0.75	0.75
Finland	0.89	58	0.85	0.76	0.81	0.84	0.81	0.75	0.70	0.53
France	0.89	57	0.83	0.68	0.80	0.84	0.79	0.63	0.67	0.78
Georgia	0.85	49	0.77	0.64	0.70	0.78	0.75	0.73	0.60	0.58
Hungary	0.89	57	0.83	0.68	0.79	0.83	0.76	0.76	0.72	0.66
Kazakhstan	0.87	52	0.79	0.65	0.79	0.79	0.79	0.74	0.61	0.60
Lebanon	0.78	41	0.73	0.54	0.49	0.77	0.77	0.74	0.50	0.46
Lithuania	0.90	58	0.85	0.73	0.76	0.84	0.82	0.72	0.76	0.63
Morocco	0.72	34	0.68	0.49	0.36	0.76	0.69	0.72	0.40	0.42
Portugal	0.89	57	0.79	0.66	0.77	0.84	0.81	0.71	0.71	0.72
Romania	0.81	44	0.78	0.50	0.56	0.77	0.73	0.74	0.54	0.59
Russian Federation	0.88	55	0.78	0.71	0.80	0.83	0.78	0.73	0.68	0.61
Sweden	0.88	54	0.80	0.70	0.78	0.82	0.77	0.68	0.67	0.68
Benchmarking Participants										
Moscow City, Russian Fed.	0.90	59	0.81	0.76	0.83	0.85	0.80	0.73	0.74	0.63

R Reverse coded

Relationship Between the TIMSS 2019 *Students Confident in Physics Scale* and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Cyprus	0.34	0.12	0.13
Finland	0.35	0.12	0.15
France	0.34	0.11	0.13
Georgia	0.29	0.08	0.10
Hungary	0.31	0.10	0.11
Kazakhstan	0.18	0.03	0.04
Lebanon	0.29	0.08	0.10
Lithuania	0.25	0.06	0.08
Morocco	0.32	0.11	0.10
Portugal	0.37	0.14	0.14
Romania	0.17	0.03	0.06
Russian Federation	0.22	0.05	0.05
Sweden	0.37	0.13	0.15
International Median	**0.31**	**0.10**	**0.10**
Benchmarking Participants			
Moscow City, Russian Fed.	0.28	0.08	0.08

Scale was created for TIMSS 2019 countries where science is taught as separate subjects.

Students Confident in Science – Grade 8

About the Scale

The *Students Confident in Science* scale was created based on students' responses to eight items listed below.

How much do you agree with these statements about science?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BSBS24A [T]	1) I usually do well in science	○	○	○	○
BSBS24B [T]	2) Science is more difficult for me than for many of my classmates [R]	○	○	○	○
BSBS24C [T]	3) Science is not one of my strengths [R]	○	○	○	○
BSBS24D [T]	4) I learn things quickly in science	○	○	○	○
BSBS24E [T]	5) I am good at working out difficult science problems	○	○	○	○
BSBS24F [T]	6) My teacher tells me I am good at science	○	○	○	○
BSBS24G [T]	7) Science is harder for me than any other subject [R]	○	○	○	○
BSBS24H [T]	8) Science makes me confused [R]	○	○	○	○

Very Confident	Somewhat Confident	Not Confident

Scale Cut Scores 11.3 9.2

[R] Reverse coded

[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Students Confident in Science* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBS24A	-0.46541	-1.14798	-0.28032	1.42830	0.92
BSBS24B R	0.03957	-1.24140	0.10527	1.13613	1.04
BSBS24C R	0.31928	-1.10283	0.16654	0.93629	1.02
BSBS24D	-0.30224	-1.38866	0.03271	1.35595	0.89
BSBS24E	0.16257	-1.43808	0.06778	1.37030	0.99
BSBS24F	0.14256	-1.20554	0.05071	1.15483	1.10
BSBS24G R	0.01606	-0.96248	-0.04795	1.01043	1.02
BSBS24H R	0.08761	-1.03184	0.12933	0.90251	1.05

R Reverse coded

Scale Transformation Constants for the TIMSS 2019 *Students Confident in Science* Scale – Grade 8

Scale Transformation Constants	
A = 9.091884	Transformed Scale Score = 9.091884 + 1.446981 • Logit Scale Score
B = 1.446981	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Students Confident in Science* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	3.24992	
1	4.91876	
2	5.74112	
3	6.31300	
4	6.76777	
5	7.15260	
6	7.49248	
7	7.80038	
8	8.08863	
9	8.36075	
10	8.62122	
11	8.87381	
12	9.12286	9.2
13	9.36803	
14	9.61581	
15	9.86849	
16	10.13014	
17	10.40580	
18	10.70197	
19	11.02414	
20	11.39185	11.3
21	11.82899	
22	12.38397	
23	13.18921	
24	14.84182	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Confident in Science* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			BSBS24A	BSBS24B R	BSBS24C R	BSBS24D	BSBS24E	BSBS24F	BSBS24G R	BSBS24H R
Australia	0.89	56	0.77	0.72	0.79	0.80	0.79	0.62	0.74	0.75
Bahrain	0.82	44	0.54	0.67	0.67	0.71	0.65	0.64	0.71	0.68
Chile	0.82	44	0.75	0.55	0.64	0.77	0.73	0.62	0.58	0.66
Chinese Taipei	0.92	63	0.82	0.78	0.82	0.83	0.80	0.78	0.77	0.75
Cyprus	-	-	-	-	-	-	-	-	-	-
Egypt	0.73	35	0.31	0.74	0.61	0.53	0.46	0.52	0.72	0.70
England	0.89	57	0.79	0.73	0.80	0.81	0.79	0.61	0.75	0.76
Finland	-	-	-	-	-	-	-	-	-	-
France	-	-	-	-	-	-	-	-	-	-
Georgia	-	-	-	-	-	-	-	-	-	-
Hong Kong SAR	0.85	49	0.74	0.69	0.75	0.73	0.71	0.56	0.71	0.68
Hungary	-	-	-	-	-	-	-	-	-	-
Iran, Islamic Rep. of	0.84	48	0.73	0.61	0.63	0.75	0.73	0.71	0.66	0.68
Ireland	0.90	60	0.81	0.77	0.79	0.83	0.79	0.65	0.78	0.78
Israel	0.85	50	0.76	0.70	0.65	0.78	0.76	0.67	0.66	0.65
Italy	0.86	52	0.78	0.69	0.74	0.79	0.67	0.67	0.68	0.71
Japan	0.89	57	0.78	0.74	0.82	0.76	0.77	0.67	0.76	0.73
Jordan	0.75	36	0.32	0.73	0.67	0.54	0.50	0.52	0.72	0.70
Kazakhstan	-	-	-	-	-	-	-	-	-	-
Korea, Rep. of	0.92	64	0.86	0.76	0.84	0.82	0.84	0.80	0.78	0.72
Kuwait	0.76	38	0.40	0.67	0.59	0.69	0.61	0.59	0.66	0.65
Lebanon	-	-	-	-	-	-	-	-	-	-
Lithuania	-	-	-	-	-	-	-	-	-	-
Malaysia	0.77	38	0.60	0.52	0.64	0.66	0.67	0.57	0.64	0.63
Morocco	-	-	-	-	-	-	-	-	-	-
New Zealand	0.86	50	0.75	0.68	0.74	0.77	0.75	0.61	0.69	0.67
Norway (9)	0.88	55	0.77	0.72	0.80	0.80	0.76	0.62	0.73	0.70
Oman	0.74	36	0.38	0.74	0.65	0.56	0.43	0.47	0.74	0.69
Portugal	-	-	-	-	-	-	-	-	-	-
Qatar	0.79	41	0.62	0.64	0.61	0.70	0.66	0.62	0.64	0.61
Romania	-	-	-	-	-	-	-	-	-	-
Russian Federation	-	-	-	-	-	-	-	-	-	-
Saudi Arabia	0.76	38	0.46	0.72	0.60	0.64	0.52	0.56	0.69	0.68
Singapore	0.91	60	0.82	0.78	0.84	0.79	0.79	0.66	0.78	0.74
South Africa (9)	0.76	37	0.68	0.52	0.59	0.70	0.68	0.57	0.55	0.58
Sweden	-	-	-	-	-	-	-	-	-	-
Turkey	0.85	49	0.75	0.68	0.74	0.70	0.69	0.65	0.69	0.70
United Arab Emirates	0.81	43	0.63	0.67	0.66	0.70	0.66	0.60	0.67	0.64
United States	0.87	52	0.73	0.72	0.74	0.79	0.74	0.58	0.74	0.73

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Confident in Science* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			BSBS24A	BSBS24B R	BSBS24C R	BSBS24D	BSBS24E	BSBS24F	BSBS24G R	BSBS24H R
Benchmarking Participants										
Ontario, Canada	0.89	57	0.79	0.76	0.81	0.79	0.78	0.59	0.77	0.75
Quebec, Canada	0.89	58	0.80	0.73	0.82	0.83	0.76	0.59	0.74	0.77
Moscow City, Russian Fed.	-	-	-	-	-	-	-	-	-	-
Gauteng, RSA (9)	0.82	45	0.72	0.60	0.68	0.73	0.71	0.59	0.63	0.67
Western Cape, RSA (9)	0.82	44	0.73	0.59	0.64	0.73	0.70	0.63	0.62	0.66
Abu Dhabi, UAE	0.77	38	0.58	0.63	0.58	0.67	0.64	0.58	0.63	0.60
Dubai, UAE	0.85	50	0.73	0.69	0.76	0.73	0.72	0.61	0.71	0.69

R Reverse coded

A dash (–) indicates comparable data not available.

Relationship Between the TIMSS 2019 *Students Confident in Science* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Australia	0.34	0.12	0.13
Bahrain	0.36	0.13	0.14
Chile	0.22	0.05	0.07
Chinese Taipei	0.35	0.13	0.19
Cyprus	-	-	-
Egypt	0.37	0.14	0.14
England	0.36	0.13	0.15
Finland	-	-	-
France	-	-	-
Georgia	-	-	-
Hong Kong SAR	0.29	0.09	0.13
Hungary	-	-	-
Iran, Islamic Rep. of	0.37	0.14	0.14
Ireland	0.43	0.19	0.18
Israel	0.39	0.15	0.17
Italy	0.32	0.10	0.10
Japan	0.37	0.14	0.15
Jordan	0.43	0.18	0.18
Kazakhstan	-	-	-
Korea, Rep. of	0.42	0.18	0.22
Kuwait	0.32	0.10	0.10
Lebanon	-	-	-
Lithuania	-	-	-
Malaysia	0.21	0.04	0.06
Morocco	-	-	-
New Zealand	0.33	0.11	0.13
Norway (9)	0.38	0.15	0.16
Oman	0.39	0.15	0.15
Portugal	-	-	-
Qatar	0.35	0.12	0.15
Romania	-	-	-
Russian Federation	-	-	-
Saudi Arabia	0.34	0.12	0.11
Singapore	0.28	0.08	0.09
South Africa (9)	0.23	0.05	0.07
Sweden	-	-	-
Turkey	0.46	0.21	0.22
United Arab Emirates	0.37	0.13	0.16
United States	0.32	0.10	0.11
International Median	**0.36**	**0.13**	**0.14**

Relationship Between the TIMSS 2019 *Students Confident in Science* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Benchmarking Participants			
Ontario, Canada	0.34	0.11	0.10
Quebec, Canada	0.37	0.14	0.13
Moscow City, Russian Fed.	-	-	-
Gauteng, RSA (9)	0.20	0.04	0.05
Western Cape, RSA (9)	0.18	0.03	0.05
Abu Dhabi, UAE	0.42	0.17	0.20
Dubai, UAE	0.31	0.10	0.11

A dash (–) indicates comparable data not available.

Students Like Learning Biology – Grade 8

About the Scale

The *Students Like Learning Biology* scale was created based on students' responses to nine items listed below.

How much do you agree with these statements about learning biology?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BSBB23A	1) I enjoy learning biology	○	○	○	○
BSBB23B	2) I wish I did not have to study biology [R]	○	○	○	○
BSBB23C	3) Biology is boring [R]	○	○	○	○
BSBB23D	4) I learn many interesting things in biology	○	○	○	○
BSBB23E	5) I like biology	○	○	○	○
BSBB23F	6) I look forward to learning biology in school	○	○	○	○
BSBB23G	7) Biology teaches me how things in the world work	○	○	○	○
BSBB23H	8) I like to conduct biology experiments	○	○	○	○
BSBB23I	9) Biology is one of my favorite subjects	○	○	○	○

Very Much Like | Somewhat Like | Do Not Like

Scale Cut Scores **11.0** **8.5**

[R] Reverse coded

Item Parameters for the TIMSS 2019 *Students Like Learning Biology* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBB23A	-0.16212	-1.47892	-0.46297	1.94189	0.75
BSBB23B [R]	0.20512	-1.24042	-0.29613	1.53655	1.45
BSBB23C [R]	0.20601	-1.47581	-0.19591	1.67172	1.21
BSBB23D	-0.77117	-1.31189	-0.49265	1.80454	0.88
BSBB23E	-0.08618	-1.39978	-0.31230	1.71208	0.64
BSBB23F	0.71997	-1.80718	0.05532	1.75186	0.90
BSBB23G	-0.76337	-1.31194	-0.65352	1.96546	1.21
BSBB23H	-0.19527	-0.99360	-0.40491	1.39851	1.48
BSBB23I	0.84701	-1.40607	0.05039	1.35568	0.85

[R] Reverse coded

Scale Transformation Constants for the TIMSS 2019 *Students Like Learning Biology* Scale – Grade 8

Scale Transformation Constants	
A = 8.641623	Transformed Scale Score = 8.641623 + 1.107618 • Logit Scale Score
B = 1.107618	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Students Like Learning Biology* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	3.77033	
1	5.00936	
2	5.61016	
3	6.03106	
4	6.36568	
5	6.65211	
6	6.90865	
7	7.14591	
8	7.37038	
9	7.58667	
10	7.79825	
11	8.00863	
12	8.21825	
13	8.43038	8.5
14	8.64675	
15	8.86844	
16	9.09669	
17	9.33234	
18	9.57605	
19	9.82964	
20	10.09524	
21	10.37666	
22	10.68028	
23	11.01688	11.0
24	11.40423	
25	11.88207	
26	12.54999	
27	13.86476	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Like Learning Biology* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			BSBB23A	BSBB23B R	BSBB23C R	BSBB23D	BSBB23E	BSBB23F	BSBB23G	BSBB23H	BSBB23I
Cyprus	0.92	62	0.87	0.68	0.73	0.81	0.91	0.85	0.67	0.63	0.86
Finland	0.93	63	0.86	0.70	0.75	0.82	0.91	0.87	0.70	0.69	0.83
France	0.91	59	0.87	0.70	0.76	0.79	0.90	0.83	0.64	0.54	0.84
Georgia	0.89	54	0.83	0.62	0.64	0.72	0.89	0.83	0.56	0.62	0.84
Hungary	0.91	60	0.88	0.67	0.76	0.79	0.91	0.86	0.63	0.57	0.85
Kazakhstan	0.90	56	0.83	0.49	0.72	0.79	0.88	0.83	0.64	0.68	0.82
Lebanon	0.86	50	0.81	0.41	0.44	0.77	0.88	0.61	0.73	0.75	0.83
Lithuania	0.92	61	0.87	0.78	0.80	0.77	0.90	0.79	0.64	0.56	0.84
Morocco	0.88	53	0.77	0.43	0.54	0.79	0.86	0.84	0.68	0.69	0.83
Portugal	0.91	59	0.86	0.73	0.73	0.80	0.90	0.84	0.70	0.50	0.81
Romania	0.88	54	0.86	0.38	0.59	0.78	0.89	0.85	0.68	0.60	0.83
Russian Federation	0.90	58	0.86	0.61	0.71	0.77	0.89	0.82	0.67	0.60	0.84
Sweden	0.92	61	0.87	0.68	0.77	0.83	0.91	0.88	0.65	0.57	0.81
Benchmarking Participants											
Moscow City, Russian Fed.	0.92	61	0.87	0.72	0.77	0.80	0.90	0.83	0.65	0.54	0.87

R Reverse coded

Relationship Between the TIMSS 2019 *Students Like Learning Biology* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale
	r	r^2	
Cyprus	0.22	0.05	0.04
Finland	0.22	0.05	0.03
France	0.18	0.03	0.03
Georgia	0.13	0.02	0.02
Hungary	0.09	0.01	0.01
Kazakhstan	0.09	0.01	0.00
Lebanon	0.28	0.08	0.07
Lithuania	0.12	0.01	0.01
Morocco	0.23	0.05	0.04
Portugal	0.16	0.03	0.02
Romania	0.08	0.01	0.01
Russian Federation	0.06	0.00	0.00
Sweden	0.17	0.03	0.02
International Median	**0.16**	**0.03**	**0.02**
Benchmarking Participants			
Moscow City, Russian Fed.	0.03	0.00	0.00

Scale was created for TIMSS 2019 countries where science is taught as separate subjects.

Students Like Learning Chemistry – Grade 8

About the Scale

The *Students Like Learning Chemistry* scale was created based on students' responses to nine items listed below.

How much do you agree with these statements about learning chemistry?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BSBC33A	1) I enjoy learning chemistry	○	○	○	○
BSBC33B	2) I wish I did not have to study chemistry [R]	○	○	○	○
BSBC33C	3) Chemistry is boring [R]	○	○	○	○
BSBC33D	4) I learn many interesting things in chemistry	○	○	○	○
BSBC33E	5) I like chemistry	○	○	○	○
BSBC33F	6) I look forward to learning chemistry in school	○	○	○	○
BSBC33G	7) Chemistry teaches me how things in the world work	○	○	○	○
BSBC33H	8) I like to conduct chemistry experiments	○	○	○	○
BSBC33I	9) Chemistry is one of my favorite subjects	○	○	○	○

Very Much Like	Somewhat Like	Do Not Like

Scale Cut Scores 11.1 8.8

[R] Reverse coded

Item Parameters for the TIMSS 2019 *Students Like Learning Chemistry* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBC33A	-0.14307	-1.46743	-0.25973	1.72716	0.73
BSBC33B [R]	0.23803	-1.27441	-0.19417	1.46858	1.66
BSBC33C [R]	0.14201	-1.45492	-0.16456	1.61948	1.36
BSBC33D	-0.61457	-1.44666	-0.41712	1.86378	0.86
BSBC33E	0.01742	-1.41770	-0.19178	1.60948	0.63
BSBC33F	0.68268	-1.77504	0.14928	1.62576	0.91
BSBC33G	-0.28919	-1.51200	-0.36951	1.88151	1.20
BSBC33H	-0.74644	-0.84521	-0.55612	1.40133	1.37
BSBC33I	0.71313	-1.44581	0.09000	1.35581	0.82

[R] Reverse coded

Scale Transformation Constants for the TIMSS 2019 *Students Like Learning Chemistry* Scale – Grade 8

Scale Transformation Constants	
A = 8.951511	Transformed Scale Score = 8.951511 + 1.041262 • Logit Scale Score
B = 1.041262	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Students Like Learning Chemistry* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	4.38181	
1	5.56044	
2	6.13177	
3	6.52970	
4	6.84408	
5	7.11089	
6	7.35187	
7	7.57186	
8	7.78125	
9	7.98339	
10	8.18235	
11	8.37833	
12	8.57458	
13	8.77259	8.8
14	8.97386	
15	9.17852	
16	9.38835	
17	9.60400	
18	9.82660	
19	10.05696	
20	10.29850	
21	10.55457	
22	10.83124	
23	11.13865	11.1
24	11.49497	
25	11.93332	
26	12.55122	
27	13.77684	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Like Learning Chemistry* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			BSBC33A	BSBC33B R	BSBC33C R	BSBC33D	BSBC33E	BSBC33F	BSBC33G	BSBC33H	BSBC33I
Cyprus	0.92	63	0.87	0.61	0.67	0.84	0.91	0.85	0.74	0.72	0.86
Finland	0.94	67	0.90	0.74	0.76	0.85	0.92	0.88	0.74	0.70	0.86
France	0.92	60	0.87	0.63	0.73	0.81	0.91	0.82	0.69	0.59	0.85
Georgia	0.91	59	0.86	0.57	0.62	0.80	0.90	0.84	0.68	0.67	0.86
Hungary	0.92	61	0.88	0.66	0.70	0.81	0.91	0.87	0.70	0.59	0.86
Kazakhstan	0.92	61	0.85	0.50	0.76	0.82	0.90	0.85	0.68	0.75	0.86
Lebanon	0.86	51	0.81	0.34	0.42	0.81	0.88	0.61	0.75	0.75	0.84
Lithuania	0.92	62	0.88	0.76	0.78	0.80	0.91	0.80	0.67	0.60	0.86
Morocco	0.88	55	0.77	0.38	0.45	0.81	0.88	0.86	0.76	0.78	0.83
Portugal	0.93	64	0.89	0.74	0.74	0.83	0.91	0.86	0.76	0.57	0.84
Romania	0.90	58	0.89	0.31	0.53	0.82	0.91	0.86	0.77	0.67	0.85
Russian Federation	0.91	60	0.88	0.61	0.67	0.81	0.90	0.85	0.69	0.65	0.86
Sweden	0.93	64	0.87	0.64	0.77	0.85	0.92	0.89	0.69	0.67	0.82
Benchmarking Participants											
Moscow City, Russian Fed.	0.93	64	0.89	0.68	0.73	0.83	0.91	0.86	0.72	0.62	0.88

R Reverse coded

Relationship Between the TIMSS 2019 *Students Like Learning Chemistry* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale
	r	r^2	
Cyprus	0.22	0.05	0.04
Finland	0.36	0.13	0.10
France	0.21	0.04	0.04
Georgia	0.20	0.04	0.04
Hungary	0.11	0.01	0.02
Kazakhstan	0.16	0.03	0.02
Lebanon	0.25	0.06	0.07
Lithuania	0.23	0.05	0.04
Morocco	0.25	0.06	0.05
Portugal	0.24	0.06	0.06
Romania	0.18	0.03	0.03
Russian Federation	0.17	0.03	0.03
Sweden	0.20	0.04	0.03
International Median	**0.21**	**0.04**	**0.04**
Benchmarking Participants			
Moscow City, Russian Fed.	0.18	0.03	0.03

Scale was created for TIMSS 2019 countries where science is taught as separate subjects.

Students Like Learning Earth Science – Grade 8

About the Scale

The *Students Like Learning Earth Science* scale was created based on students' responses to nine items listed below.

How much do you agree with these statements about learning earth science?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BSBE28A	1) I enjoy learning earth science	◯	◯	◯	◯
BSBE28B	2) I wish I did not have to study earth science [R]	◯	◯	◯	◯
BSBE28C	3) Earth science is boring [R]	◯	◯	◯	◯
BSBE28D	4) I learn many interesting things in earth science	◯	◯	◯	◯
BSBE28E	5) I like earth science	◯	◯	◯	◯
BSBE28F	6) I look forward to learning earth science in school	◯	◯	◯	◯
BSBE28G	7) Earth science teaches me how things in the world work	◯	◯	◯	◯
BSBE28H	8) I like to conduct earth science experiments	◯	◯	◯	◯
BSBE28I	9) Earth science is one of my favorite subjects	◯	◯	◯	◯

Very Much Like | Somewhat Like | Do Not Like

Scale Cut Scores **11.2** **8.8**

[R] Reverse coded

Item Parameters for the TIMSS 2019 *Students Like Learning Earth Science* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBE28A	-0.22224	-1.55381	-0.36522	1.91903	0.75
BSBE28B [R]	0.08027	-1.31109	-0.29421	1.60530	1.56
BSBE28C [R]	0.08661	-1.54312	-0.20119	1.74431	1.29
BSBE28D	-0.72581	-1.46248	-0.48855	1.95103	0.87
BSBE28E	-0.12416	-1.50925	-0.28124	1.79049	0.64
BSBE28F	0.66218	-1.87680	0.10256	1.77424	0.84
BSBE28G	-0.84329	-1.28030	-0.70858	1.98888	1.11
BSBE28H	0.27811	-0.98803	-0.30562	1.29365	1.52
BSBE28I	0.80833	-1.46329	0.03245	1.43084	0.85

[R] Reverse coded

Scale Transformation Constants for the TIMSS 2019 *Students Like Learning Earth Science* Scale – Grade 8

Scale Transformation Constants	
A = 8.903182	Transformed Scale Score = 8.903182 + 1.089155 • Logit Scale Score
B = 1.089155	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019
***Students Like Learning Earth Science* Scale – Grade 8**

Raw Score	Transformed Scale Score	Cutpoint
0	4.02015	
1	5.24902	
2	5.84858	
3	6.27182	
4	6.61042	
5	6.90196	
6	7.16441	
7	7.40916	
8	7.64056	
9	7.86305	
10	8.07990	
11	8.29368	
12	8.50651	
13	8.72026	8.8
14	8.93707	
15	9.15692	
16	9.38258	
17	9.61474	
18	9.85461	
19	10.10305	
20	10.36341	
21	10.63911	
22	10.93638	
23	11.26581	11.2
24	11.64644	
25	12.11289	
26	12.76721	
27	14.05738	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Like Learning Earth Science Scale – Grade 8*

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			BSBE28A	BSBE28B R	BSBE28C R	BSBE28D	BSBE28E	BSBE28F	BSBE28G	BSBE28H	BSBE28I
Cyprus	0.92	61	0.87	0.61	0.66	0.82	0.91	0.86	0.70	0.70	0.85
Finland	0.93	64	0.87	0.69	0.74	0.83	0.91	0.86	0.73	0.71	0.82
France	0.91	59	0.87	0.70	0.75	0.79	0.90	0.83	0.64	0.54	0.84
Georgia	0.87	52	0.80	0.42	0.49	0.75	0.87	0.81	0.70	0.69	0.81
Hungary	0.92	61	0.89	0.61	0.71	0.81	0.92	0.87	0.68	0.64	0.86
Kazakhstan	0.90	57	0.81	0.43	0.73	0.78	0.88	0.83	0.68	0.70	0.84
Lebanon	-	-	-	-	-	-	-	-	-	-	-
Lithuania	0.91	61	0.88	0.75	0.79	0.79	0.90	0.79	0.66	0.55	0.85
Morocco	0.88	55	0.79	0.40	0.49	0.81	0.87	0.85	0.72	0.75	0.83
Portugal	0.91	59	0.86	0.73	0.73	0.80	0.90	0.84	0.70	0.50	0.81
Romania	0.85	51	0.84	0.16	0.36	0.80	0.89	0.85	0.76	0.61	0.81
Russian Federation	0.90	57	0.84	0.57	0.67	0.79	0.89	0.83	0.72	0.62	0.83
Sweden	-	-	-	-	-	-	-	-	-	-	-
Benchmarking Participants											
Moscow City, Russian Fed.	0.91	60	0.86	0.66	0.73	0.81	0.89	0.84	0.69	0.62	0.85

R Reverse coded

Relationship Between the TIMSS 2019 *Students Like Learning Earth Science Scale and TIMSS 2019 Achievement – Grade 8*

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale
	r	r^2	
Cyprus	0.10	0.01	0.01
Finland	0.19	0.04	0.03
France	0.18	0.03	0.03
Georgia	0.08	0.01	0.01
Hungary	0.03	0.00	0.01
Kazakhstan	0.07	0.00	0.00
Lebanon	-	-	-
Lithuania	0.06	0.00	0.00
Morocco	0.19	0.04	0.03
Portugal	0.16	0.03	0.02
Romania	0.09	0.01	0.02
Russian Federation	0.01	0.00	0.00
Sweden	-	-	-
International Median	**0.09**	**0.01**	**0.01**
Benchmarking Participants			
Moscow City, Russian Fed.	0.02	0.00	0.00

Scale was created for TIMSS 2019 countries where science is taught as separate subjects.
A dash (–) indicates comparable data not available.

Students Like Learning Mathematics – Grade 8

About the Scale

The *Students Like Learning Mathematics* scale was created based on students' responses to nine items listed below.

How much do you agree with these statements about learning mathematics?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BSBM16A [T]	1) I enjoy learning mathematics	○	○	○	○
BSBM16B [T]	2) I wish I did not have to study mathematics [R]	○	○	○	○
BSBM16C [T]	3) Mathematics is boring [R]	○	○	○	○
BSBM16D [T]	4) I learn many interesting things in mathematics	○	○	○	○
BSBM16E [T]	5) I like mathematics	○	○	○	○
BSBM16F [T]	6) I like any schoolwork that involves numbers	○	○	○	○
BSBM16G [T]	7) I like to solve mathematics problems	○	○	○	○
BSBM16H [T]	8) I look forward to mathematics lessons	○	○	○	○
BSBM16I [T]	9) Mathematics is one of my favorite subjects	○	○	○	○

Very Much Like		Somewhat Like		Do Not Like
Scale Cut Scores	11.4		9.4	

[R] Reverse coded

[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Students Like Learning Mathematics* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBM16A	-0.52075	-1.52420	-0.44716	1.97136	0.80
BSBM16B [R]	-0.16493	-1.27666	0.02387	1.25279	1.68
BSBM16C [R]	0.08233	-1.70726	0.11045	1.59681	1.34
BSBM16D	-0.67768	-1.75407	-0.23907	1.99314	1.13
BSBM16E	-0.20808	-1.32239	-0.32723	1.64962	0.66
BSBM16F	0.24118	-1.93026	0.06299	1.86727	1.03
BSBM16G	0.07107	-1.61719	-0.15594	1.77313	0.85
BSBM16H	0.65566	-1.77947	-0.05548	1.83495	0.93
BSBM16I	0.52120	-1.07237	-0.04864	1.12101	0.81

[R] Reverse coded

Scale Transformation Constants for the TIMSS 2019 *Students Like Learning Mathematics* Scale – Grade 8

Scale Transformation Constants	
A = 9.42706	Transformed Scale Score = 9.42706 + 0.94246 • Logit Scale Score
B = 0.94246	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Students Like Learning Mathematics* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	5.09484	
1	6.19800	
2	6.74882	
3	7.13627	
4	7.44727	
5	7.71308	
6	7.95043	
7	8.16805	
8	8.37412	
9	8.57095	
10	8.76139	
11	8.94763	
12	9.13133	
13	9.31385	9.4
14	9.49640	
15	9.68014	
16	9.86638	
17	10.05658	
18	10.25252	
19	10.45644	
20	10.67004	
21	10.89898	
22	11.14859	
23	11.42832	11.4
24	11.75478	
25	12.15847	
26	12.72711	
27	13.84849	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Like Learning Mathematics* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			BSBM16A	BSBM16B R	BSBM16C R	BSBM16D	BSBM16E	BSBM16F	BSBM16G	BSBM16H	BSBM16I
Australia	0.94	67	0.89	0.72	0.75	0.73	0.91	0.81	0.86	0.83	0.86
Bahrain	0.94	66	0.85	0.66	0.74	0.79	0.91	0.79	0.85	0.84	0.87
Chile	0.93	64	0.85	0.64	0.74	0.71	0.90	0.80	0.87	0.84	0.85
Chinese Taipei	0.94	70	0.90	0.69	0.76	0.83	0.92	0.85	0.89	0.79	0.87
Cyprus	0.94	67	0.88	0.73	0.73	0.77	0.91	0.79	0.83	0.84	0.88
Egypt	0.88	53	0.74	0.41	0.49	0.74	0.86	0.71	0.86	0.82	0.82
England	0.93	65	0.86	0.72	0.75	0.72	0.90	0.79	0.84	0.82	0.84
Finland	0.94	67	0.87	0.72	0.77	0.80	0.92	0.79	0.81	0.83	0.85
France	0.93	63	0.86	0.71	0.78	0.71	0.90	0.73	0.74	0.82	0.85
Georgia	0.92	60	0.85	0.59	0.59	0.68	0.89	0.77	0.84	0.84	0.86
Hong Kong SAR	0.93	67	0.88	0.58	0.64	0.83	0.92	0.85	0.90	0.81	0.88
Hungary	0.93	65	0.88	0.64	0.71	0.74	0.91	0.83	0.83	0.83	0.85
Iran, Islamic Rep. of	0.93	64	0.84	0.71	0.69	0.74	0.89	0.77	0.84	0.82	0.88
Ireland	0.94	66	0.88	0.73	0.76	0.74	0.91	0.78	0.82	0.82	0.86
Israel	0.93	65	0.85	0.65	0.74	0.78	0.91	0.78	0.86	0.80	0.88
Italy	0.95	71	0.91	0.78	0.82	0.71	0.92	0.85	0.84	0.86	0.89
Japan	0.94	69	0.89	0.62	0.73	0.82	0.92	0.85	0.88	0.84	0.88
Jordan	0.91	60	0.82	0.43	0.54	0.78	0.89	0.80	0.85	0.85	0.87
Kazakhstan	0.92	62	0.81	0.58	0.73	0.74	0.88	0.77	0.87	0.84	0.85
Korea, Rep. of	0.94	69	0.89	0.59	0.74	0.82	0.92	0.88	0.91	0.76	0.89
Kuwait	0.92	62	0.83	0.45	0.63	0.79	0.90	0.80	0.84	0.85	0.87
Lebanon	0.89	53	0.83	0.51	0.52	0.71	0.88	0.69	0.82	0.64	0.86
Lithuania	0.93	64	0.86	0.70	0.74	0.70	0.90	0.80	0.87	0.80	0.83
Malaysia	0.90	57	0.82	0.65	0.71	0.69	0.87	0.73	0.78	0.71	0.81
Morocco	0.91	59	0.82	0.52	0.63	0.77	0.89	0.75	0.75	0.86	0.85
New Zealand	0.93	66	0.88	0.74	0.73	0.70	0.90	0.81	0.85	0.83	0.85
Norway (9)	0.94	69	0.88	0.69	0.77	0.79	0.91	0.83	0.89	0.86	0.86
Oman	0.88	53	0.80	0.35	0.55	0.77	0.87	0.69	0.82	0.71	0.85
Portugal	0.95	71	0.89	0.75	0.79	0.79	0.91	0.81	0.88	0.88	0.85
Qatar	0.92	63	0.84	0.54	0.58	0.78	0.91	0.82	0.87	0.84	0.88
Romania	0.92	61	0.85	0.52	0.63	0.75	0.89	0.72	0.88	0.84	0.86
Russian Federation	0.92	63	0.85	0.60	0.72	0.72	0.89	0.79	0.84	0.82	0.85
Saudi Arabia	0.93	64	0.82	0.55	0.75	0.78	0.89	0.80	0.85	0.86	0.87
Singapore	0.94	67	0.88	0.75	0.74	0.72	0.90	0.77	0.86	0.81	0.88
South Africa (9)	0.89	55	0.80	0.56	0.62	0.64	0.87	0.72	0.81	0.75	0.81
Sweden	0.94	70	0.88	0.69	0.81	0.78	0.92	0.87	0.86	0.84	0.86
Turkey	0.92	62	0.81	0.65	0.76	0.61	0.89	0.76	0.84	0.83	0.88
United Arab Emirates	0.92	62	0.83	0.60	0.63	0.76	0.89	0.78	0.84	0.83	0.87
United States	0.94	67	0.88	0.67	0.71	0.73	0.91	0.81	0.89	0.85	0.86

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Like Learning Mathematics* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			BSBM16A	BSBM16B R	BSBM16C R	BSBM16D	BSBM16E	BSBM16F	BSBM16G	BSBM16H	BSBM16I
Benchmarking Participants											
Ontario, Canada	0.94	68	0.89	0.71	0.76	0.71	0.91	0.81	0.87	0.87	0.87
Quebec, Canada	0.93	65	0.88	0.60	0.79	0.75	0.91	0.82	0.83	0.81	0.86
Moscow City, Russian Fed.	0.93	65	0.88	0.68	0.76	0.75	0.90	0.76	0.83	0.82	0.86
Gauteng, RSA (9)	0.91	59	0.83	0.64	0.68	0.67	0.88	0.73	0.82	0.78	0.83
Western Cape, RSA (9)	0.91	60	0.83	0.61	0.66	0.69	0.88	0.77	0.84	0.81	0.85
Abu Dhabi, UAE	0.90	57	0.82	0.51	0.53	0.72	0.88	0.78	0.83	0.80	0.85
Dubai, UAE	0.93	65	0.83	0.69	0.71	0.77	0.89	0.78	0.85	0.83	0.87

R Reverse coded

Relationship Between the TIMSS 2019 *Students Like Learning Mathematics* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Australia	0.39	0.15	0.12
Bahrain	0.20	0.04	0.04
Chile	0.21	0.05	0.04
Chinese Taipei	0.42	0.17	0.16
Cyprus	0.37	0.13	0.13
Egypt	0.17	0.03	0.04
England	0.24	0.06	0.05
Finland	0.44	0.19	0.16
France	0.35	0.12	0.11
Georgia	0.26	0.07	0.06
Hong Kong SAR	0.30	0.09	0.08
Hungary	0.38	0.14	0.14
Iran, Islamic Rep. of	0.25	0.06	0.07
Ireland	0.32	0.11	0.10
Israel	0.17	0.03	0.02
Italy	0.35	0.12	0.12
Japan	0.40	0.16	0.13
Jordan	0.16	0.02	0.03
Kazakhstan	0.16	0.03	0.03
Korea, Rep. of	0.39	0.15	0.13
Kuwait	0.17	0.03	0.03
Lebanon	0.22	0.05	0.05
Lithuania	0.26	0.07	0.07
Malaysia	0.21	0.04	0.05
Morocco	0.25	0.06	0.05
New Zealand	0.29	0.08	0.07
Norway (9)	0.42	0.18	0.13
Oman	0.27	0.07	0.08
Portugal	0.37	0.14	0.13
Qatar	0.25	0.06	0.05
Romania	0.30	0.09	0.09
Russian Federation	0.28	0.08	0.08
Saudi Arabia	0.14	0.02	0.02
Singapore	0.33	0.11	0.09
South Africa (9)	0.11	0.01	0.02
Sweden	0.34	0.12	0.10
Turkey	0.25	0.06	0.07
United Arab Emirates	0.23	0.05	0.05
United States	0.27	0.07	0.07
International Median	**0.27**	**0.07**	**0.07**

Relationship Between the TIMSS 2019 *Students Like Learning Mathematics* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Benchmarking Participants			
Ontario, Canada	0.40	0.16	0.14
Quebec, Canada	0.36	0.13	0.11
Moscow City, Russian Fed.	0.39	0.15	0.15
Gauteng, RSA (9)	0.09	0.01	0.01
Western Cape, RSA (9)	0.06	0.00	0.01
Abu Dhabi, UAE	0.21	0.05	0.07
Dubai, UAE	0.27	0.07	0.07

Students Like Learning Physics – Grade 8

About the Scale

The *Students Like Learning Physics* scale was created based on students' responses to nine items listed below.

How much do you agree with these statements about learning physics?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BSBP38A	1) I enjoy learning physics	◯	◯	◯	◯
BSBP38B	2) I wish I did not have to study physics [R]	◯	◯	◯	◯
BSBP38C	3) Physics is boring [R]	◯	◯	◯	◯
BSBP38D	4) I learn many interesting things in physics	◯	◯	◯	◯
BSBP38E	5) I like physics	◯	◯	◯	◯
BSBP38F	6) I look forward to learning physics in school	◯	◯	◯	◯
BSBP38G	7) Physics teaches me how things in the world work	◯	◯	◯	◯
BSBP38H	8) I like to conduct physics experiments	◯	◯	◯	◯
BSBP38I	9) Physics is one of my favorite subjects	◯	◯	◯	◯

Very Much Like | Somewhat Like | Do Not Like

Scale Cut Scores 11.3 9.0

[R] Reverse coded

Item Parameters for the TIMSS 2019 *Students Like Learning Physics* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBP38A	-0.09991	-1.44455	-0.24740	1.69195	0.76
BSBP38B [R]	0.21221	-1.24941	-0.18976	1.43917	1.65
BSBP38C [R]	0.16331	-1.45914	-0.11608	1.57522	1.41
BSBP38D	-0.60437	-1.42950	-0.41629	1.84579	0.88
BSBP38E	0.06539	-1.42600	-0.16967	1.59567	0.63
BSBP38F	0.66918	-1.74385	0.15965	1.58420	0.90
BSBP38G	-0.58101	-1.34701	-0.48738	1.83439	1.17
BSBP38H	-0.57636	-0.98606	-0.45642	1.44248	1.27
BSBP38I	0.75156	-1.40132	0.07443	1.32689	0.83

[R] Reverse coded

Scale Transformation Constants for the TIMSS 2019 *Students Like Learning Physics* Scale – Grade 8

Scale Transformation Constants	
A = 9.112372	Transformed Scale Score = 9.112372 + 1.052564 • Logit Scale Score
B = 1.052564	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019
***Students Like Learning Physics* Scale – Grade 8**

Raw Score	Transformed Scale Score	Cutpoint
0	4.50052	
1	5.69131	
2	6.27071	
3	6.67290	
4	6.99227	
5	7.26430	
6	7.50753	
7	7.73282	
8	7.94527	
9	8.14964	
10	8.34943	
11	8.54716	
12	8.74474	
13	8.94365	9.0
14	9.14548	
15	9.34993	
16	9.55919	
17	9.77382	
18	9.99507	
19	10.22382	
20	10.46372	
21	10.71828	
22	10.99372	
23	11.30039	11.3
24	11.65666	
25	12.09616	
26	12.71724	
27	13.95282	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Like Learning Physics* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			BSBP38A	BSBP38B R	BSBP38C R	BSBP38D	BSBP38E	BSBP38F	BSBP38G	BSBP38H	BSBP38I
Cyprus	0.93	65	0.89	0.64	0.68	0.85	0.92	0.87	0.74	0.76	0.88
Finland	0.94	67	0.89	0.72	0.74	0.84	0.92	0.88	0.74	0.75	0.85
France	0.92	60	0.87	0.63	0.73	0.81	0.91	0.82	0.69	0.59	0.85
Georgia	0.89	55	0.84	0.53	0.54	0.78	0.89	0.83	0.67	0.68	0.85
Hungary	0.92	62	0.88	0.64	0.70	0.80	0.91	0.86	0.68	0.72	0.85
Kazakhstan	0.91	59	0.83	0.44	0.74	0.79	0.89	0.83	0.70	0.77	0.84
Lebanon	0.86	52	0.83	0.21	0.31	0.82	0.90	0.67	0.78	0.79	0.84
Lithuania	0.93	63	0.88	0.78	0.76	0.80	0.91	0.78	0.68	0.67	0.86
Morocco	0.88	56	0.79	0.35	0.41	0.82	0.88	0.87	0.77	0.80	0.83
Portugal	0.93	64	0.89	0.74	0.74	0.83	0.91	0.86	0.76	0.57	0.84
Romania	0.89	55	0.86	0.28	0.45	0.81	0.90	0.85	0.76	0.70	0.83
Russian Federation	0.91	59	0.84	0.57	0.66	0.80	0.89	0.82	0.70	0.71	0.84
Sweden	0.93	65	0.88	0.65	0.75	0.86	0.92	0.90	0.72	0.70	0.82
Benchmarking Participants											
Moscow City, Russian Fed.	0.92	63	0.87	0.66	0.71	0.81	0.91	0.83	0.71	0.71	0.86

R Reverse coded

Relationship Between the TIMSS 2019 *Students Like Learning Physics* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale
	r	r^2	
Cyprus	0.27	0.07	0.06
Finland	0.31	0.10	0.08
France	0.21	0.04	0.04
Georgia	0.17	0.03	0.04
Hungary	0.19	0.04	0.03
Kazakhstan	0.15	0.02	0.02
Lebanon	0.20	0.04	0.05
Lithuania	0.17	0.03	0.03
Morocco	0.25	0.06	0.06
Portugal	0.24	0.06	0.06
Romania	0.15	0.02	0.03
Russian Federation	0.17	0.03	0.02
Sweden	0.25	0.06	0.05
International Median	**0.20**	**0.04**	**0.04**
Benchmarking Participants			
Moscow City, Russian Fed.	0.20	0.04	0.03

Scale was created for TIMSS 2019 countries where science is taught as separate subjects.

Students Like Learning Science – Grade 8

About the Scale

The *Students Like Learning Science* scale was created based on students' responses to nine items listed below.

How much do you agree with these statements about learning science?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BSBS22A [T]	1) I enjoy learning science	○	○	○	○
BSBS22B [T]	2) I wish I did not have to study science [R]	○	○	○	○
BSBS22C [T]	3) Science is boring [R]	○	○	○	○
BSBS22D [T]	4) I learn many interesting things in science	○	○	○	○
BSBS22E [T]	5) I like science	○	○	○	○
BSBS22F [T]	6) I look forward to learning science in school	○	○	○	○
BSBS22G [T]	7) Science teaches me how things in the world work	○	○	○	○
BSBS22H [T]	8) I like to conduct science experiments	○	○	○	○
BSBS22I [T]	9) Science is one of my favorite subjects	○	○	○	○

	Very Much Like		Somewhat Like		Do Not Like
Scale Cut Scores		**10.6**		**8.3**	

[R] Reverse coded

[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Students Like Learning Science* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBS22A	-0.18816	-1.12110	-0.51287	1.63397	0.77
BSBS22B [R]	0.51905	-1.21205	-0.09407	1.30612	1.56
BSBS22C [R]	0.45459	-1.38086	-0.07300	1.45386	1.35
BSBS22D	-0.51760	-1.05343	-0.51799	1.57142	0.88
BSBS22E	-0.00730	-1.17719	-0.29690	1.47409	0.64
BSBS22F	0.40936	-1.50803	-0.05340	1.56143	0.81
BSBS22G	-0.58381	-1.04270	-0.57925	1.62195	1.08
BSBS22H	-0.60801	-0.85093	-0.44646	1.29739	1.35
BSBS22I	0.52188	-1.21442	0.00714	1.20728	0.80

[R] Reverse coded

Scale Transformation Constants for the TIMSS 2019 *Students Like Learning Science* Scale – Grade 8

Scale Transformation Constants	
A = 8.489044	Transformed Scale Score = 8.489044 + 1.125793 • Logit Scale Score
B = 1.125793	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Students Like Learning Science* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	3.84650	
1	5.07484	
2	5.65807	
3	6.06022	
4	6.37614	
5	6.64332	
6	6.88421	
7	7.10319	
8	7.31154	
9	7.51243	
10	7.71018	
11	7.90450	
12	8.09896	
13	8.29500	8.3
14	8.49394	
15	8.69661	
16	8.90441	
17	9.11856	
18	9.34076	
19	9.57219	
20	9.81723	
21	10.07979	
22	10.36663	
23	10.68878	10.6
24	11.06414	
25	11.53300	
26	12.19630	
27	13.51778	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Like Learning Science* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			BSBS22A	BSBS22B R	BSBS22C R	BSBS22D	BSBS22E	BSBS22F	BSBS22G	BSBS22H	BSBS22I
Australia	0.93	64	0.88	0.70	0.76	0.80	0.92	0.90	0.74	0.57	0.87
Bahrain	0.90	58	0.81	0.54	0.62	0.82	0.89	0.85	0.74	0.62	0.86
Chile	0.90	57	0.85	0.50	0.66	0.80	0.90	0.85	0.71	0.62	0.84
Chinese Taipei	0.92	62	0.87	0.63	0.66	0.82	0.90	0.86	0.74	0.69	0.87
Cyprus	-	-	-	-	-	-	-	-	-	-	-
Egypt	0.83	47	0.66	0.33	0.41	0.75	0.83	0.82	0.73	0.70	0.79
England	0.93	63	0.88	0.75	0.75	0.81	0.91	0.89	0.70	0.53	0.86
Finland	-	-	-	-	-	-	-	-	-	-	-
France	-	-	-	-	-	-	-	-	-	-	-
Georgia	-	-	-	-	-	-	-	-	-	-	-
Hong Kong SAR	0.92	61	0.87	0.54	0.58	0.83	0.90	0.88	0.75	0.76	0.87
Hungary	-	-	-	-	-	-	-	-	-	-	-
Iran, Islamic Rep. of	0.89	56	0.82	0.63	0.62	0.73	0.87	0.81	0.67	0.63	0.86
Ireland	0.93	64	0.90	0.73	0.75	0.80	0.92	0.89	0.71	0.57	0.87
Israel	0.91	59	0.86	0.59	0.64	0.82	0.90	0.82	0.72	0.63	0.88
Italy	0.91	59	0.86	0.77	0.77	0.77	0.89	0.83	0.62	0.43	0.84
Japan	0.92	63	0.87	0.67	0.73	0.82	0.90	0.86	0.69	0.65	0.87
Jordan	0.86	55	0.78	0.28	0.38	0.82	0.88	0.86	0.79	0.76	0.84
Kazakhstan	-	-	-	-	-	-	-	-	-	-	-
Korea, Rep. of	0.93	64	0.89	0.67	0.75	0.81	0.91	0.84	0.72	0.69	0.90
Kuwait	0.89	56	0.79	0.39	0.53	0.82	0.89	0.87	0.73	0.66	0.86
Lebanon	-	-	-	-	-	-	-	-	-	-	-
Lithuania	-	-	-	-	-	-	-	-	-	-	-
Malaysia	0.88	52	0.80	0.58	0.63	0.76	0.85	0.73	0.68	0.63	0.78
Morocco	-	-	-	-	-	-	-	-	-	-	-
New Zealand	0.92	61	0.88	0.71	0.66	0.78	0.91	0.89	0.70	0.56	0.85
Norway (9)	0.92	60	0.88	0.70	0.75	0.81	0.91	0.87	0.66	0.50	0.83
Oman	0.84	49	0.73	0.28	0.48	0.77	0.85	0.78	0.74	0.68	0.81
Portugal	-	-	-	-	-	-	-	-	-	-	-
Qatar	0.89	57	0.83	0.39	0.45	0.82	0.90	0.86	0.78	0.70	0.87
Romania	-	-	-	-	-	-	-	-	-	-	-
Russian Federation	-	-	-	-	-	-	-	-	-	-	-
Saudi Arabia	0.89	55	0.80	0.44	0.58	0.82	0.87	0.85	0.69	0.68	0.84
Singapore	0.91	60	0.86	0.73	0.73	0.76	0.90	0.87	0.68	0.52	0.86
South Africa (9)	0.86	50	0.79	0.45	0.52	0.69	0.85	0.81	0.64	0.68	0.80
Sweden	-	-	-	-	-	-	-	-	-	-	-
Turkey	0.87	52	0.80	0.57	0.67	0.60	0.86	0.78	0.68	0.62	0.84
United Arab Emirates	0.90	56	0.81	0.49	0.55	0.80	0.89	0.85	0.75	0.68	0.85
United States	0.91	60	0.87	0.58	0.66	0.78	0.90	0.88	0.72	0.65	0.86

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Like Learning Science* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			BSBS22A	BSBS22B R	BSBS22C R	BSBS22D	BSBS22E	BSBS22F	BSBS22G	BSBS22H	BSBS22I
Benchmarking Participants											
Ontario, Canada	0.92	61	0.88	0.68	0.76	0.76	0.91	0.88	0.67	0.57	0.87
Quebec, Canada	0.91	60	0.88	0.58	0.76	0.80	0.91	0.87	0.63	0.57	0.86
Moscow City, Russian Fed.	-	-	-	-	-	-	-	-	-	-	-
Gauteng, RSA (9)	0.89	54	0.82	0.56	0.64	0.70	0.88	0.84	0.66	0.63	0.83
Western Cape, RSA (9)	0.89	54	0.84	0.52	0.59	0.71	0.88	0.85	0.64	0.65	0.84
Abu Dhabi, UAE	0.88	54	0.80	0.39	0.41	0.80	0.88	0.84	0.76	0.71	0.84
Dubai, UAE	0.91	58	0.82	0.63	0.71	0.76	0.89	0.84	0.72	0.60	0.84

R Reverse coded

A dash (–) indicates comparable data not available.

Relationship Between the TIMSS 2019 *Students Like Learning Science* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Australia	0.33	0.11	0.09
Bahrain	0.26	0.07	0.06
Chile	0.16	0.02	0.02
Chinese Taipei	0.32	0.10	0.09
Cyprus	-	-	-
Egypt	0.34	0.12	0.11
England	0.31	0.09	0.08
Finland	-	-	-
France	-	-	-
Georgia	-	-	-
Hong Kong SAR	0.26	0.07	0.06
Hungary	-	-	-
Iran, Islamic Rep. of	0.22	0.05	0.05
Ireland	0.34	0.11	0.09
Israel	0.22	0.05	0.04
Italy	0.20	0.04	0.03
Japan	0.31	0.10	0.09
Jordan	0.26	0.07	0.08
Kazakhstan	-	-	-
Korea, Rep. of	0.39	0.15	0.15
Kuwait	0.24	0.06	0.04
Lebanon	-	-	-
Lithuania	-	-	-
Malaysia	0.28	0.08	0.08
Morocco	-	-	-
New Zealand	0.26	0.07	0.06
Norway (9)	0.27	0.07	0.06
Oman	0.34	0.12	0.10
Portugal	-	-	-
Qatar	0.28	0.08	0.07
Romania	-	-	-
Russian Federation	-	-	-
Saudi Arabia	0.22	0.05	0.04
Singapore	0.30	0.09	0.08
South Africa (9)	0.22	0.05	0.05
Sweden	-	-	-
Turkey	0.24	0.06	0.05
United Arab Emirates	0.36	0.13	0.13
United States	0.21	0.05	0.04
International Median	**0.27**	**0.07**	**0.07**

Relationship Between the TIMSS 2019 *Students Like Learning Science* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Benchmarking Participants			
Ontario, Canada	0.24	0.06	0.04
Quebec, Canada	0.26	0.07	0.06
Moscow City, Russian Fed.	-	-	-
Gauteng, RSA (9)	0.13	0.02	0.02
Western Cape, RSA (9)	0.14	0.02	0.02
Abu Dhabi, UAE	0.42	0.17	0.17
Dubai, UAE	0.27	0.07	0.08

A dash (–) indicates comparable data not available.

Students Value Mathematics – Grade 8

About the Scale

The *Students Value Mathematics* scale was created based on students' responses to nine items listed below.

How much do you agree with these statements about mathematics?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BSBM20A T	1) I think learning mathematics will help me in my daily life	○	○	○	○
BSBM20B T	2) I need mathematics to learn other school subjects	○	○	○	○
BSBM20C T	3) I need to do well in mathematics to get into the university of my choice	○	○	○	○
BSBM20D T	4) I need to do well in mathematics to get the job I want	○	○	○	○
BSBM20E T	5) I would like a job that involves using mathematics	○	○	○	○
BSBM20F T	6) It is important to learn about mathematics to get ahead in the world	○	○	○	○
BSBM20G T	7) Learning mathematics will give me more job opportunities when I am an adult	○	○	○	○
BSBM20H T	8) My parents think that it is important that I do well in mathematics	○	○	○	○
BSBM20I T	9) It is important to do well in mathematics	○	○	○	○

Strongly Value | Somewhat Value | Do Not Value

Scale Cut Scores 10.3 7.8

T Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Students Value Mathematics* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBM20A	-0.07683	-0.97955	-0.48401	1.46356	1.08
BSBM20B	0.21444	-1.41743	-0.41427	1.83170	1.14
BSBM20C	-0.21179	-1.03293	-0.24434	1.27727	0.96
BSBM20D	0.00026	-1.16359	-0.12444	1.28803	0.89
BSBM20E	1.46285	-1.34979	-0.01893	1.36872	1.28
BSBM20F	0.08173	-1.23148	-0.32418	1.55566	0.88
BSBM20G	-0.27682	-1.09839	-0.44615	1.54454	0.86
BSBM20H	-0.61122	-1.07802	-0.36012	1.43814	1.19
BSBM20I	-0.58262	-0.91732	-0.52093	1.43825	0.89

Scale Transformation Constants for the TIMSS 2019 *Students Value Mathematics* Scale – Grade 8

Scale Transformation Constants	
A = 7.964227	Transformed Scale Score = 7.964227 + 1.193777 • Logit Scale Score
B = 1.193777	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Students Value Mathematics* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	3.03682	
1	4.34723	
2	4.96949	
3	5.39638	
4	5.72855	
5	6.01079	
6	6.25944	
7	6.48714	
8	6.70110	
9	6.90631	
10	7.10653	
11	7.30474	
12	7.50425	
13	7.70526	7.8
14	7.91135	
15	8.12441	
16	8.34615	
17	8.57829	
18	8.82228	
19	9.08085	
20	9.35676	
21	9.65438	
22	9.98068	
23	10.34734	10.3
24	10.77308	
25	11.30190	
26	12.04145	
27	13.48294	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Value Mathematics* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			BSBM20A	BSBM20B	BSBM20C	BSBM20D	BSBM20E	BSBM20F	BSBM20G	BSBM20H	BSBM20I
Australia	0.90	58	0.76	0.74	0.78	0.80	0.66	0.83	0.82	0.63	0.81
Bahrain	0.90	56	0.74	0.73	0.78	0.77	0.64	0.80	0.81	0.63	0.77
Chile	0.89	54	0.71	0.69	0.78	0.79	0.58	0.81	0.81	0.67	0.75
Chinese Taipei	0.91	57	0.68	0.68	0.79	0.82	0.72	0.81	0.84	0.65	0.79
Cyprus	0.88	52	0.70	0.64	0.79	0.81	0.67	0.77	0.73	0.59	0.77
Egypt	0.87	49	0.65	0.69	0.74	0.75	0.61	0.73	0.73	0.66	0.71
England	0.88	53	0.69	0.70	0.72	0.77	0.63	0.80	0.79	0.63	0.76
Finland	0.91	58	0.72	0.70	0.80	0.82	0.68	0.83	0.81	0.69	0.81
France	0.88	53	0.71	0.68	0.74	0.78	0.60	0.79	0.77	0.64	0.80
Georgia	0.89	54	0.72	0.63	0.76	0.78	0.65	0.81	0.80	0.67	0.79
Hong Kong SAR	0.92	61	0.70	0.75	0.81	0.84	0.70	0.84	0.87	0.67	0.81
Hungary	0.88	52	0.72	0.68	0.69	0.78	0.68	0.77	0.76	0.67	0.76
Iran, Islamic Rep. of	0.89	53	0.68	0.65	0.77	0.80	0.73	0.80	0.78	0.56	0.76
Ireland	0.88	52	0.70	0.70	0.74	0.78	0.63	0.80	0.76	0.59	0.77
Israel	0.88	52	0.70	0.68	0.73	0.79	0.63	0.82	0.78	0.54	0.75
Italy	0.88	52	0.75	0.67	0.69	0.76	0.66	0.80	0.80	0.55	0.74
Japan	0.87	49	0.66	0.68	0.70	0.76	0.60	0.81	0.79	0.59	0.66
Jordan	0.89	54	0.67	0.72	0.79	0.78	0.57	0.80	0.79	0.71	0.77
Kazakhstan	0.90	56	0.72	0.65	0.79	0.82	0.73	0.77	0.83	0.66	0.77
Korea, Rep. of	0.90	55	0.65	0.70	0.79	0.82	0.60	0.80	0.83	0.65	0.82
Kuwait	0.89	53	0.70	0.69	0.79	0.78	0.62	0.80	0.79	0.60	0.76
Lebanon	0.88	51	0.70	0.69	0.75	0.75	0.62	0.79	0.77	0.62	0.73
Lithuania	0.88	51	0.74	0.65	0.73	0.79	0.67	0.72	0.77	0.56	0.78
Malaysia	0.87	50	0.64	0.60	0.75	0.76	0.57	0.75	0.76	0.70	0.77
Morocco	0.89	53	0.69	0.71	0.78	0.79	0.65	0.77	0.77	0.62	0.78
New Zealand	0.90	57	0.75	0.72	0.77	0.79	0.65	0.82	0.80	0.67	0.80
Norway (9)	0.90	57	0.73	0.73	0.79	0.80	0.68	0.82	0.81	0.63	0.76
Oman	0.87	49	0.66	0.69	0.73	0.73	0.58	0.76	0.76	0.66	0.74
Portugal	0.89	55	0.70	0.66	0.79	0.81	0.66	0.82	0.78	0.63	0.79
Qatar	0.91	59	0.74	0.74	0.81	0.81	0.67	0.83	0.84	0.67	0.80
Romania	0.90	56	0.74	0.71	0.74	0.79	0.69	0.78	0.80	0.69	0.79
Russian Federation	0.90	56	0.72	0.70	0.78	0.82	0.70	0.79	0.81	0.61	0.77
Saudi Arabia	0.87	50	0.70	0.66	0.68	0.75	0.60	0.81	0.79	0.61	0.75
Singapore	0.86	49	0.67	0.64	0.73	0.76	0.60	0.79	0.78	0.58	0.73
South Africa (9)	0.84	46	0.67	0.57	0.75	0.76	0.55	0.73	0.72	0.58	0.70
Sweden	0.87	50	0.64	0.65	0.79	0.80	0.68	0.75	0.76	0.59	0.68
Turkey	0.88	51	0.62	0.64	0.76	0.79	0.67	0.81	0.79	0.58	0.74
United Arab Emirates	0.90	55	0.70	0.71	0.78	0.78	0.65	0.82	0.80	0.66	0.76
United States	0.89	54	0.74	0.70	0.75	0.78	0.63	0.82	0.81	0.63	0.76

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Value Mathematics* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			BSBM20A	BSBM20B	BSBM20C	BSBM20D	BSBM20E	BSBM20F	BSBM20G	BSBM20H	BSBM20I
Benchmarking Participants											
Ontario, Canada	0.87	50	0.72	0.67	0.74	0.75	0.64	0.80	0.77	0.55	0.73
Quebec, Canada	0.87	51	0.73	0.66	0.75	0.77	0.64	0.73	0.77	0.62	0.76
Moscow City, Russian Fed.	0.90	55	0.73	0.64	0.80	0.83	0.74	0.78	0.81	0.55	0.75
Gauteng, RSA (9)	0.83	46	0.66	0.60	0.74	0.78	0.58	0.75	0.74	0.52	0.68
Western Cape, RSA (9)	0.85	48	0.67	0.63	0.74	0.77	0.60	0.77	0.74	0.53	0.73
Abu Dhabi, UAE	0.89	54	0.68	0.71	0.77	0.78	0.63	0.80	0.79	0.66	0.76
Dubai, UAE	0.88	53	0.69	0.65	0.77	0.75	0.66	0.81	0.79	0.64	0.73

Relationship Between the TIMSS 2019 *Students Value Mathematics* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Australia	0.22	0.05	0.05
Bahrain	0.12	0.02	0.01
Chile	0.07	0.01	0.01
Chinese Taipei	0.36	0.13	0.12
Cyprus	0.25	0.06	0.06
Egypt	0.12	0.02	0.02
England	0.07	0.01	0.01
Finland	0.33	0.11	0.09
France	0.16	0.02	0.03
Georgia	0.14	0.02	0.02
Hong Kong SAR	0.23	0.05	0.05
Hungary	0.24	0.06	0.04
Iran, Islamic Rep. of	0.12	0.02	0.01
Ireland	0.20	0.04	0.04
Israel	0.09	0.01	0.01
Italy	0.16	0.02	0.02
Japan	0.25	0.06	0.05
Jordan	0.15	0.02	0.03
Kazakhstan	0.05	0.00	0.00
Korea, Rep. of	0.41	0.17	0.16
Kuwait	0.13	0.02	0.02
Lebanon	0.21	0.04	0.04
Lithuania	0.13	0.02	0.01
Malaysia	0.24	0.06	0.05
Morocco	0.21	0.04	0.03
New Zealand	0.13	0.02	0.01
Norway (9)	0.23	0.05	0.05
Oman	0.20	0.04	0.04
Portugal	0.30	0.09	0.07
Qatar	0.16	0.03	0.03
Romania	0.17	0.03	0.03
Russian Federation	0.15	0.02	0.02
Saudi Arabia	0.09	0.01	0.01
Singapore	0.14	0.02	0.02
South Africa (9)	0.17	0.03	0.03
Sweden	0.13	0.02	0.02
Turkey	0.23	0.05	0.05
United Arab Emirates	0.18	0.03	0.04
United States	0.13	0.02	0.02
International Median	**0.16**	**0.03**	**0.03**

Relationship Between the TIMSS 2019 *Students Value Mathematics* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Benchmarking Participants			
Ontario, Canada	0.22	0.05	0.05
Quebec, Canada	0.20	0.04	0.04
Moscow City, Russian Fed.	0.27	0.07	0.06
Gauteng, RSA (9)	0.06	0.00	0.00
Western Cape, RSA (9)	0.04	0.00	0.00
Abu Dhabi, UAE	0.21	0.05	0.05
Dubai, UAE	0.17	0.03	0.03

Students Value Science – Grade 8

About the Scale

The *Students Value Science* scale was created based on students' responses to nine items listed below.

How much do you agree with these statements about science?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BSBS25A T	1) I think learning science will help me in my daily life	○	○	○	○
BSBS25B T	2) I need science to learn other school subjects	○	○	○	○
BSBS25C T	3) I need to do well in science to get into the university of my choice	○	○	○	○
BSBS25D T	4) I need to do well in science to get the job I want	○	○	○	○
BSBS25E T	5) I would like a job that involves using science	○	○	○	○
BSBS25F T	6) It is important to learn about science to get ahead in the world	○	○	○	○
BSBS25G T	7) Learning science will give me more job opportunities when I am an adult	○	○	○	○
BSBS25H T	8) My parents think that it is important that I do well in science	○	○	○	○
BSBS25I T	9) It is important to do well in science	○	○	○	○

Strongly Value | Somewhat Value | Do Not Value

Scale Cut Scores 10.6 8.5

T Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Students Value Science* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BSBS25A	-0.54895	-1.70374	-0.44149	2.14523	1.15
BSBS25B	0.21327	-2.19431	-0.06708	2.26139	1.21
BSBS25C	-0.00218	-1.78502	-0.01869	1.80371	0.91
BSBS25D	0.28699	-1.77451	0.08479	1.68972	0.87
BSBS25E	1.01281	-1.78405	0.20415	1.57990	1.16
BSBS25F	0.03247	-1.87630	-0.18356	2.05986	0.95
BSBS25G	-0.09115	-1.79872	-0.19220	1.99092	0.92
BSBS25H	-0.29541	-1.89534	-0.18736	2.08270	1.29
BSBS25I	-0.60785	-1.66485	-0.36766	2.03251	1.02

Scale Transformation Constants for the TIMSS 2019 *Students Value Science* Scale – Grade 8

Scale Transformation Constants	
A = 8.556049	Transformed Scale Score = 8.556049 + 0.883354 • Logit Scale Score
B = 0.883354	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Students Value Science* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	4.28340	
1	5.31763	
2	5.83503	
3	6.20452	
4	6.50215	
5	6.75863	
6	6.99286	
7	7.21075	
8	7.41821	
9	7.61946	
10	7.81751	
11	8.01452	
12	8.21205	
13	8.41123	8.5
14	8.61280	
15	8.81725	
16	9.02481	
17	9.23574	
18	9.45051	
19	9.66950	
20	9.89532	
21	10.13091	
22	10.38161	
23	10.65629	10.6
24	10.97071	
25	11.35416	
26	11.88785	
27	12.93670	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Value Science* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			BSBS25A	BSBS25B	BSBS25C	BSBS25D	BSBS25E	BSBS25F	BSBS25G	BSBS25H	BSBS25I
Australia	0.94	67	0.82	0.79	0.87	0.86	0.81	0.84	0.84	0.70	0.81
Bahrain	0.92	60	0.74	0.73	0.83	0.83	0.76	0.80	0.82	0.70	0.76
Chile	0.92	62	0.76	0.75	0.85	0.85	0.79	0.82	0.85	0.69	0.72
Chinese Taipei	0.93	65	0.72	0.76	0.85	0.87	0.81	0.83	0.87	0.72	0.81
Cyprus	0.92	63	0.76	0.75	0.85	0.85	0.75	0.82	0.81	0.72	0.81
Egypt	0.90	54	0.64	0.72	0.80	0.79	0.73	0.76	0.75	0.70	0.74
England	0.93	63	0.79	0.77	0.83	0.82	0.76	0.83	0.84	0.71	0.80
Finland	0.94	66	0.77	0.78	0.86	0.86	0.77	0.86	0.83	0.76	0.84
France	0.92	62	0.76	0.75	0.81	0.84	0.74	0.80	0.82	0.73	0.82
Georgia	0.91	58	0.68	0.69	0.78	0.81	0.73	0.81	0.82	0.75	0.79
Hong Kong SAR	0.94	69	0.71	0.80	0.88	0.89	0.83	0.87	0.89	0.74	0.83
Hungary	0.92	61	0.70	0.73	0.81	0.84	0.76	0.80	0.80	0.73	0.81
Iran, Islamic Rep. of	0.92	60	0.72	0.71	0.84	0.84	0.81	0.83	0.83	0.64	0.75
Ireland	0.93	63	0.78	0.75	0.84	0.84	0.78	0.82	0.83	0.69	0.81
Israel	0.94	66	0.81	0.83	0.85	0.86	0.80	0.88	0.84	0.64	0.78
Italy	0.91	58	0.76	0.69	0.79	0.81	0.78	0.82	0.82	0.62	0.74
Japan	0.90	56	0.70	0.74	0.79	0.82	0.66	0.84	0.82	0.65	0.65
Jordan	0.92	61	0.71	0.78	0.82	0.82	0.73	0.83	0.81	0.73	0.77
Kazakhstan	0.92	62	0.67	0.71	0.84	0.86	0.80	0.81	0.86	0.77	0.77
Korea, Rep. of	0.92	62	0.74	0.77	0.85	0.85	0.70	0.79	0.86	0.68	0.84
Kuwait	0.91	59	0.72	0.72	0.84	0.83	0.76	0.81	0.82	0.66	0.73
Lebanon	0.92	60	0.73	0.72	0.82	0.82	0.75	0.81	0.82	0.71	0.80
Lithuania	0.90	57	0.72	0.71	0.81	0.84	0.74	0.73	0.78	0.68	0.80
Malaysia	0.91	57	0.70	0.68	0.81	0.82	0.72	0.73	0.80	0.75	0.77
Morocco	0.92	60	0.70	0.76	0.81	0.82	0.76	0.81	0.80	0.72	0.76
New Zealand	0.93	65	0.81	0.80	0.83	0.84	0.77	0.83	0.83	0.73	0.82
Norway (9)	0.93	64	0.79	0.79	0.85	0.85	0.79	0.82	0.84	0.68	0.79
Oman	0.89	54	0.69	0.72	0.80	0.78	0.68	0.77	0.78	0.69	0.71
Portugal	0.92	62	0.72	0.69	0.83	0.85	0.79	0.85	0.84	0.71	0.80
Qatar	0.94	67	0.77	0.77	0.87	0.86	0.79	0.86	0.86	0.75	0.82
Romania	0.93	65	0.73	0.76	0.83	0.84	0.79	0.82	0.84	0.79	0.84
Russian Federation	0.92	60	0.71	0.70	0.83	0.84	0.78	0.80	0.82	0.68	0.78
Saudi Arabia	0.92	61	0.75	0.76	0.83	0.83	0.76	0.83	0.83	0.69	0.75
Singapore	0.91	60	0.74	0.63	0.83	0.83	0.76	0.82	0.84	0.68	0.79
South Africa (9)	0.91	58	0.73	0.71	0.83	0.83	0.77	0.76	0.80	0.68	0.76
Sweden	0.92	59	0.70	0.72	0.85	0.85	0.81	0.79	0.79	0.68	0.72
Turkey	0.90	56	0.64	0.68	0.81	0.83	0.77	0.81	0.81	0.65	0.71
United Arab Emirates	0.92	61	0.74	0.73	0.83	0.83	0.77	0.82	0.82	0.72	0.77
United States	0.92	62	0.77	0.74	0.83	0.83	0.76	0.84	0.84	0.69	0.76

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Students Value Science* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	BSBS25A	BSBS25B	BSBS25C	BSBS25D	BSBS25E	BSBS25F	BSBS25G	BSBS25H	BSBS25I
Benchmarking Participants											
Ontario, Canada	0.92	61	0.78	0.72	0.84	0.83	0.78	0.82	0.81	0.69	0.76
Quebec, Canada	0.92	61	0.79	0.74	0.85	0.83	0.74	0.76	0.83	0.70	0.79
Moscow City, Russian Fed.	0.92	60	0.71	0.70	0.84	0.85	0.81	0.81	0.81	0.66	0.78
Gauteng, RSA (9)	0.93	63	0.76	0.74	0.86	0.86	0.81	0.77	0.83	0.72	0.78
Western Cape, RSA (9)	0.92	62	0.73	0.73	0.85	0.85	0.80	0.79	0.81	0.71	0.78
Abu Dhabi, UAE	0.92	61	0.73	0.74	0.83	0.82	0.76	0.82	0.81	0.73	0.78
Dubai, UAE	0.91	60	0.74	0.72	0.84	0.84	0.80	0.81	0.82	0.67	0.73

Component Loadings for Each Item: BSBS25A–BSBS25I

Relationship Between the TIMSS 2019 *Students Value Science* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Australia	0.28	0.08	0.07
Bahrain	0.22	0.05	0.04
Chile	0.05	0.00	0.00
Chinese Taipei	0.30	0.09	0.07
Cyprus	0.28	0.08	0.06
Egypt	0.23	0.05	0.06
England	0.20	0.04	0.04
Finland	0.33	0.11	0.09
France	0.24	0.06	0.06
Georgia	0.11	0.01	0.01
Hong Kong SAR	0.18	0.03	0.03
Hungary	0.15	0.02	0.02
Iran, Islamic Rep. of	0.15	0.02	0.02
Ireland	0.30	0.09	0.07
Israel	0.15	0.02	0.02
Italy	0.15	0.02	0.02
Japan	0.28	0.08	0.06
Jordan	0.17	0.03	0.03
Kazakhstan	0.07	0.00	0.00
Korea, Rep. of	0.40	0.16	0.15
Kuwait	0.15	0.02	0.02
Lebanon	0.25	0.06	0.05
Lithuania	0.12	0.02	0.01
Malaysia	0.36	0.13	0.12
Morocco	0.15	0.02	0.02
New Zealand	0.18	0.03	0.03
Norway (9)	0.16	0.03	0.02
Oman	0.26	0.07	0.06
Portugal	0.23	0.05	0.04
Qatar	0.20	0.04	0.04
Romania	0.12	0.01	0.01
Russian Federation	0.02	0.00	0.00
Saudi Arabia	0.11	0.01	0.01
Singapore	0.27	0.07	0.07
South Africa (9)	0.05	0.00	0.02
Sweden	0.14	0.02	0.02
Turkey	0.13	0.02	0.02
United Arab Emirates	0.25	0.06	0.06
United States	0.13	0.02	0.02
International Median	**0.18**	**0.03**	**0.03**

Relationship Between the TIMSS 2019 *Students Value Science* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Benchmarking Participants			
Ontario, Canada	0.21	0.04	0.04
Quebec, Canada	0.28	0.08	0.07
Moscow City, Russian Fed.	0.02	0.00	0.00
Gauteng, RSA (9)	-0.06	0.00	0.02
Western Cape, RSA (9)	-0.01	0.00	0.01
Abu Dhabi, UAE	0.29	0.09	0.08
Dubai, UAE	0.19	0.04	0.04

Scales Based on Principals' Reports

Instruction Affected by Mathematics Resource Shortages – Grade 8

About the Scale

The *Instruction Affected by Mathematics Resource Shortages* scale was created based on principals' responses to thirteen items listed below.

How much is your school's capacity to provide instruction affected by a shortage or inadequacy of the following?

		Not at all	A little	Some	A lot
	A. General School Resources				
BCBG13AA T	1) Instructional materials (e.g., textbooks)	○	○	○	○
BCBG13AB T	2) Supplies (e.g., papers, pencils, materials)	○	○	○	○
BCBG13AC T	3) School buildings and grounds	○	○	○	○
BCBG13AD T	4) Heating/cooling and lighting systems	○	○	○	○
BCBG13AE T	5) Instructional space (e.g., classrooms)	○	○	○	○
BCBG13AF T	6) Technologically competent staff	○	○	○	○
BCBG13AG T	7) Audio-visual resources for delivery of instruction (e.g., interactive white boards, digital projectors)	○	○	○	○
BCBG13AH T	8) Computer technology for teaching and learning (e.g., computers or tablets for student use)	○	○	○	○
	B. Resources for Mathematics Instruction				
BCBG13BA T	1) Teachers with a specialization in mathematics	○	○	○	○
BCBG13BB T	2) Computer software/applications for mathematics instruction	○	○	○	○
BCBG13BC T	3) Library resources relevant to mathematics instruction	○	○	○	○
BCBG13BD T	4) Calculators for mathematics instruction	○	○	○	○
BCBG13BE T	5) Concrete objects or materials to help students understand quantities or procedures	○	○	○	○

Not Affected	Somewhat Affected	Affected A Lot

Scale Cut Scores 11.0 7.5

T Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Instruction Affected by Mathematics Resource Shortages* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BCBG13AA	-0.08523	-0.29202	0.01436	0.27766	0.91
BCBG13AB	-0.31466	-0.30062	0.24458	0.05604	0.91
BCBG13AC	0.24178	-0.75472	0.05376	0.70096	1.06
BCBG13AD	-0.07941	-0.62905	0.03772	0.59133	1.08
BCBG13AE	0.29495	-0.47671	-0.14092	0.61763	0.99
BCBG13AF	0.08339	-1.41713	-0.02038	1.43751	0.98
BCBG13AG	0.05752	-1.15114	0.03387	1.11727	0.92
BCBG13AH	0.22037	-1.21651	0.06667	1.14984	1.06
BCBG13BA	0.12749	0.03139	-0.10299	0.07160	1.10
BCBG13BB	0.00156	-1.56949	0.12923	1.44026	1.01
BCBG13BC	-0.14676	-1.50055	0.10438	1.39617	1.17
BCBG13BD	-0.38067	-0.76280	0.07944	0.68336	1.15
BCBG13BE	-0.02033	-1.41394	-0.03525	1.44919	0.90

Scale Transformation Constants for the TIMSS 2019 *Instruction Affected by Mathematics Resource Shortages* Scale – Grade 8

Scale Transformation Constants	
A = 9.262831	Transformed Scale Score = 9.262831 + 1.171716 • Logit Scale Score
B = 1.171716	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Instruction Affected by Mathematics Resource Shortages* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	4.23486	
1	5.57102	
2	6.21558	
3	6.65270	
4	6.98745	
5	7.26025	
6	7.49129	7.5
7	7.69254	
8	7.87145	
9	8.03304	
10	8.18097	
11	8.31802	
12	8.44578	
13	8.56742	
14	8.68343	
15	8.79492	
16	8.90286	
17	9.00816	
18	9.11164	
19	9.21409	
20	9.31621	
21	9.41874	
22	9.52239	
23	9.62794	
24	9.73618	
25	9.84801	
26	9.96435	
27	10.08629	
28	10.21426	
29	10.35137	
30	10.49907	
31	10.66003	
32	10.83775	
33	11.03708	11.0
34	11.26511	
35	11.53380	
36	11.86315	
37	12.29366	
38	12.92960	
39	14.25596	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instruction Affected by Mathematics Resource Shortages* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item												
			BCBG13AA	BCBG13AB	BCBG13AC	BCBG13AD	BCBG13AE	BCBG13AF	BCBG13AG	BCBG13AH	BCBG13BA	BCBG13BB	BCBG13BC	BCBG13BD	BCBG13BE
Australia	0.94	58	0.84	0.85	0.70	0.71	0.71	0.75	0.77	0.81	0.69	0.74	0.74	0.81	0.75
Bahrain	0.97	73	0.92	0.89	0.89	0.87	0.91	0.80	0.93	0.76	0.87	0.80	0.74	0.78	0.90
Chile	0.92	53	0.69	0.72	0.71	0.55	0.76	0.79	0.79	0.81	0.66	0.73	0.73	0.72	0.76
Chinese Taipei	0.90	46	0.55	0.62	0.71	0.70	0.71	0.76	0.79	0.79	0.50	0.67	0.67	0.55	0.68
Cyprus	0.92	53	0.84	0.77	0.70	0.79	0.71	0.79	0.72	0.64	0.76	0.69	0.75	0.63	0.68
Egypt	0.89	43	0.66	0.69	0.70	0.55	0.69	0.74	0.55	0.69	0.63	0.66	0.61	0.68	0.70
England	0.91	48	0.72	0.62	0.72	0.63	0.70	0.66	0.78	0.69	0.66	0.69	0.70	0.66	0.74
Finland	0.90	46	0.75	0.64	0.61	0.63	0.72	0.69	0.71	0.69	0.62	0.74	0.46	0.68	0.78
France	0.86	40	0.68	0.56	0.61	0.56	0.68	0.58	0.77	0.68	0.49	0.74	0.74	0.45	0.62
Georgia	0.89	45	0.67	0.62	0.58	0.62	0.67	0.67	0.80	0.78	0.69	0.77	0.74	0.18	0.71
Hong Kong SAR	0.96	70	0.86	0.88	0.78	0.91	0.76	0.78	0.82	0.83	0.84	0.83	0.84	0.86	0.85
Hungary	0.89	44	0.68	0.57	0.65	0.68	0.73	0.55	0.72	0.73	0.56	0.66	0.67	0.59	0.75
Iran, Islamic Rep. of	0.89	44	0.68	0.75	0.67	0.79	0.77	0.74	0.72	0.50	0.69	0.65	0.46	0.53	0.58
Ireland	0.85	40	0.48	0.47	0.73	0.61	0.65	0.72	0.76	0.68	0.54	0.64	0.42	0.66	0.77
Israel	0.90	46	0.71	0.72	0.58	0.71	0.66	0.68	0.70	0.60	0.71	0.69	0.61	0.74	0.73
Italy	0.85	37	0.46	0.48	0.59	0.47	0.59	0.61	0.66	0.73	0.60	0.66	0.57	0.63	0.72
Japan	0.90	48	0.69	0.73	0.69	0.59	0.81	0.71	0.73	0.73	0.71	0.62	0.71	0.39	0.78
Jordan	0.88	43	0.76	0.80	0.77	0.54	0.79	0.68	0.51	0.71	0.76	0.59	0.41	0.46	0.60
Kazakhstan	0.93	55	0.78	0.68	0.67	0.76	0.75	0.80	0.81	0.80	0.78	0.77	0.76	0.55	0.73
Korea, Rep. of	0.95	64	0.81	0.83	0.79	0.87	0.85	0.74	0.81	0.81	0.84	0.77	0.77	0.72	0.75
Kuwait	0.95	61	0.86	0.83	0.71	0.80	0.83	0.74	0.82	0.78	0.80	0.73	0.61	0.75	0.86
Lebanon	0.94	59	0.84	0.85	0.85	0.70	0.87	0.82	0.81	0.39	0.78	0.72	0.65	0.84	0.73
Lithuania	0.88	43	0.74	0.62	0.49	0.54	0.53	0.58	0.73	0.73	0.52	0.68	0.75	0.71	0.77
Malaysia	0.91	47	0.74	0.77	0.72	0.63	0.71	0.72	0.59	0.49	0.74	0.67	0.66	0.76	0.67
Morocco	0.79	32	0.20	0.24	-0.06	0.42	-0.12	0.74	0.73	0.76	-0.11	0.80	0.79	0.67	0.70
New Zealand	0.95	64	0.85	0.84	0.76	0.80	0.70	0.76	0.86	0.75	0.74	0.88	0.83	0.80	0.84
Norway (9)	0.82	33	0.59	0.41	0.65	0.66	0.76	0.73	0.71	0.51	0.31	0.62	0.55	0.36	0.38
Oman	0.95	61	0.79	0.79	0.83	0.81	0.80	0.84	0.79	0.81	0.83	0.78	0.72	0.58	0.74
Portugal	0.90	47	0.58	0.63	0.72	0.63	0.77	0.71	0.79	0.67	0.60	0.69	0.70	0.70	0.66
Qatar	0.98	85	0.95	0.92	0.91	0.95	0.95	0.95	0.94	0.95	0.93	0.85	0.91	0.89	0.91
Romania	0.89	45	0.79	0.79	0.73	0.74	0.70	0.70	0.57	0.60	0.66	0.50	0.67	0.50	0.69
Russian Federation	0.91	49	0.76	0.66	0.59	0.72	0.61	0.73	0.77	0.71	0.66	0.74	0.75	0.63	0.77
Saudi Arabia	0.92	51	0.72	0.80	0.59	0.80	0.78	0.77	0.78	0.74	0.69	0.66	0.45	0.72	0.69
Singapore	0.98	82	0.95	0.94	0.91	0.87	0.91	0.86	0.93	0.94	0.93	0.89	0.82	0.91	0.90
South Africa (9)	0.84	35	0.58	0.41	0.67	0.64	0.52	0.69	0.69	0.57	0.32	0.63	0.53	0.65	0.65
Sweden	0.86	40	0.78	0.74	0.61	0.60	0.61	0.66	0.71	0.74	0.56	0.63	0.27	0.52	0.63
Turkey	0.93	53	0.84	0.82	0.73	0.80	0.80	0.57	0.77	0.67	0.73	0.73	0.64	0.55	0.76
United Arab Emirates	0.97	76	0.88	0.87	0.89	0.92	0.90	0.91	0.91	0.89	0.87	0.82	0.78	0.81	0.85
United States	0.94	58	0.79	0.82	0.68	0.71	0.67	0.76	0.82	0.78	0.74	0.82	0.72	0.76	0.83

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instruction Affected by Mathematics Resource Shortages* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item												
			BCBG13AA	BCBG13AB	BCBG13AC	BCBG13AD	BCBG13AE	BCBG13AF	BCBG13AG	BCBG13AH	BCBG13BA	BCBG13BB	BCBG13BC	BCBG13BD	BCBG13BE
Benchmarking Participants															
Ontario, Canada	0.90	46	0.75	0.72	0.64	0.63	0.63	0.65	0.79	0.76	0.51	0.69	0.69	0.63	0.71
Quebec, Canada	0.93	54	0.80	0.81	0.74	0.62	0.78	0.71	0.79	0.61	0.77	0.69	0.75	0.74	0.74
Moscow City, Russian Fed.	0.95	66	0.82	0.78	0.77	0.81	0.82	0.84	0.84	0.83	0.82	0.85	0.90	0.59	0.85
Gauteng, RSA (9)	0.87	39	0.73	0.71	0.77	0.50	0.75	0.73	0.58	0.27	0.62	0.44	0.33	0.74	0.66
Western Cape, RSA (9)	0.90	45	0.74	0.66	0.70	0.69	0.60	0.73	0.78	0.67	0.67	0.68	0.46	0.67	0.61
Abu Dhabi, UAE	0.98	78	0.89	0.86	0.92	0.92	0.92	0.93	0.93	0.87	0.92	0.85	0.74	0.87	0.84
Dubai, UAE	0.98	80	0.90	0.93	0.85	0.94	0.88	0.90	0.92	0.92	0.91	0.85	0.86	0.84	0.90

Relationship Between the TIMSS 2019 *Instruction Affected by Mathematics Resource Shortages* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Australia	0.28	0.08	0.05
Bahrain	0.05	0.00	0.01
Chile	0.24	0.06	0.06
Chinese Taipei	0.06	0.00	0.00
Cyprus	0.00	0.00	0.01
Egypt	0.08	0.01	0.02
England	0.11	0.01	0.01
Finland	0.04	0.00	0.00
France	0.10	0.01	0.02
Georgia	0.14	0.02	0.01
Hong Kong SAR	0.09	0.01	0.04
Hungary	-0.02	0.00	0.00
Iran, Islamic Rep. of	0.09	0.01	0.03
Ireland	0.14	0.02	0.01
Israel	0.14	0.02	0.01
Italy	0.07	0.01	0.00
Japan	0.08	0.01	0.01
Jordan	-0.12	0.01	0.02
Kazakhstan	0.05	0.00	0.00
Korea, Rep. of	0.03	0.00	0.00
Kuwait	0.03	0.00	0.02
Lebanon	0.20	0.04	0.08
Lithuania	-0.02	0.00	0.00
Malaysia	-0.09	0.01	0.03
Morocco	-0.06	0.00	0.04
New Zealand	0.20	0.04	0.02
Norway (9)	0.02	0.00	0.00
Oman	0.10	0.01	0.02
Portugal	0.11	0.01	0.01
Qatar	0.28	0.08	0.10
Romania	0.18	0.03	0.03
Russian Federation	0.12	0.01	0.01
Saudi Arabia	-0.04	0.00	0.01
Singapore	0.05	0.00	0.00
South Africa (9)	0.21	0.04	0.09
Sweden	0.03	0.00	0.01
Turkey	0.17	0.03	0.03
United Arab Emirates	0.21	0.04	0.05
United States	0.11	0.01	0.02
International Median	**0.09**	**0.01**	**0.01**

Relationship Between the TIMSS 2019 *Instruction Affected by Mathematics Resource Shortages* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Benchmarking Participants			
Ontario, Canada	0.12	0.01	0.02
Quebec, Canada	0.01	0.00	0.00
Moscow City, Russian Fed.	-0.02	0.00	0.00
Gauteng, RSA (9)	0.30	0.09	0.13
Western Cape, RSA (9)	0.47	0.22	0.21
Abu Dhabi, UAE	0.19	0.04	0.04
Dubai, UAE	0.19	0.04	0.05

Instruction Affected by Science Resource Shortages – Grade 8

About the Scale

The *Instruction Affected by Science Resource Shortages* scale was created based on students' responses to thirteen items listed below.

How much is your school's capacity to provide instruction affected by a shortage or inadequacy of the following?

		Not at all	A little	Some	A lot
	A. General School Resources				
BCBG13AA [T]	1) Instructional materials (e.g., textbooks)	○	○	○	○
BCBG13AB [T]	2) Supplies (e.g., papers, pencils, materials)	○	○	○	○
BCBG13AC [T]	3) School buildings and grounds	○	○	○	○
BCBG13AD [T]	4) Heating/cooling and lighting systems	○	○	○	○
BCBG13AE [T]	5) Instructional space (e.g., classrooms)	○	○	○	○
BCBG13AF [T]	6) Technologically competent staff	○	○	○	○
BCBG13AG [T]	7) Audio-visual resources for delivery of instruction (e.g., interactive white boards, digital projectors)	○	○	○	○
BCBG13AH [T]	8) Computer technology for teaching and learning (e.g., computers or tablets for student use)	○	○	○	○
	B. Resources for Science Instruction				
BCBG13CA [T]	1) Teachers with a specialization in science	○	○	○	○
BCBG13CB [T]	2) Computer software/applications for science instruction	○	○	○	○
BCBG13CC [T]	3) Library resources relevant to science instruction	○	○	○	○
BCBG13CD [T]	4) Calculators for science instruction	○	○	○	○
BCBG13CE [T]	5) Science equipment and materials for experiments	○	○	○	○

Not Affected		Somewhat Affected		Affected A Lot
Scale Cut Scores	11.1		7.5	

[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Instruction Affected by Science Resource Shortages* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BCBG13AA	-0.11528	-0.33350	0.01422	0.31928	0.92
BCBG13AB	-0.35173	-0.34117	0.24349	0.09768	0.92
BCBG13AC	0.22227	-0.79974	0.05530	0.74444	1.07
BCBG13AD	-0.10888	-0.67368	0.03847	0.63521	1.12
BCBG13AE	0.27674	-0.51943	-0.13970	0.65913	1.00
BCBG13AF	0.05952	-1.47321	-0.01797	1.49118	1.01
BCBG13AG	0.03294	-1.20191	0.03559	1.16632	0.95
BCBG13AH	0.20110	-1.26755	0.06834	1.19921	1.09
BCBG13CA	0.07561	0.07128	-0.07062	-0.00066	1.08
BCBG13CB	0.04668	-1.47047	0.01664	1.45383	0.97
BCBG13CC	-0.14890	-1.46091	0.00973	1.45118	1.05
BCBG13CD	-0.50898	-0.84828	0.02025	0.82803	1.27
BCBG13CE	0.31891	-0.97265	-0.04312	1.01577	0.90

Scale Transformation Constants for the TIMSS 2019 *Instruction Affected by Science Resource Shortages* Scale – Grade 8

Scale Transformation Constants	
A = 9.315269	Transformed Scale Score = 9.315269 + 1.229797 • Logit Scale Score
B = 1.229797	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Instruction Affected by Science Resource Shortages* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	4.07624	
1	5.47004	
2	6.13851	
3	6.59064	
4	6.93605	
5	7.21741	
6	7.45575	7.5
7	7.66379	
8	7.84903	
9	8.01669	
10	8.17049	
11	8.31328	
12	8.44672	
13	8.57396	
14	8.69549	
15	8.81243	
16	8.92582	
17	9.03657	
18	9.14554	
19	9.25352	
20	9.36128	
21	9.46956	
22	9.57915	
23	9.69082	
24	9.80545	
25	9.92393	
26	10.04727	
27	10.17656	
28	10.31231	
29	10.45776	
30	10.61443	
31	10.78505	
32	10.97329	
33	11.18424	11.1
34	11.42529	
35	11.70901	
36	12.05633	
37	12.50978	
38	13.17876	
39	14.57239	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instruction Affected by Science Resource Shortages* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item												
			BCBG13AA	BCBG13AB	BCBG13AC	BCBG13AD	BCBG13AE	BCBG13AF	BCBG13AG	BCBG13AH	BCBG13CA	BCBG13CB	BCBG13CC	BCBG13CD	BCBG13CE
Australia	0.94	59	0.85	0.83	0.70	0.71	0.73	0.74	0.76	0.79	0.74	0.78	0.76	0.76	0.79
Bahrain	0.97	73	0.92	0.89	0.89	0.88	0.92	0.79	0.93	0.75	0.88	0.88	0.80	0.65	0.88
Chile	0.93	54	0.69	0.71	0.73	0.58	0.79	0.78	0.78	0.79	0.70	0.75	0.78	0.73	0.73
Chinese Taipei	0.90	48	0.57	0.65	0.71	0.71	0.70	0.76	0.78	0.78	0.57	0.72	0.72	0.52	0.76
Cyprus	0.93	54	0.83	0.76	0.74	0.80	0.74	0.78	0.71	0.61	0.75	0.72	0.70	0.63	0.76
Egypt	0.89	44	0.65	0.67	0.72	0.56	0.70	0.72	0.54	0.71	0.68	0.70	0.63	0.59	0.76
England	0.91	50	0.72	0.61	0.74	0.62	0.70	0.63	0.80	0.72	0.63	0.76	0.79	0.69	0.78
Finland	0.90	45	0.73	0.62	0.64	0.66	0.73	0.69	0.73	0.69	0.54	0.73	0.54	0.67	0.76
France	0.87	40	0.67	0.56	0.63	0.60	0.69	0.58	0.76	0.68	0.48	0.70	0.69	0.48	0.69
Georgia	0.91	48	0.64	0.57	0.57	0.59	0.67	0.70	0.78	0.80	0.71	0.82	0.77	0.47	0.76
Hong Kong SAR	0.97	71	0.86	0.88	0.79	0.90	0.77	0.79	0.82	0.84	0.88	0.85	0.81	0.88	0.87
Hungary	0.89	43	0.69	0.62	0.68	0.66	0.75	0.52	0.70	0.72	0.54	0.70	0.63	0.58	0.71
Iran, Islamic Rep. of	0.89	44	0.70	0.76	0.67	0.78	0.79	0.73	0.69	0.48	0.69	0.67	0.57	0.31	0.65
Ireland	0.88	42	0.49	0.50	0.73	0.62	0.66	0.70	0.74	0.65	0.55	0.69	0.58	0.67	0.79
Israel	0.89	43	0.71	0.74	0.61	0.70	0.70	0.66	0.67	0.56	0.59	0.66	0.58	0.66	0.71
Italy	0.84	35	0.42	0.49	0.62	0.46	0.61	0.60	0.65	0.72	0.57	0.70	0.64	0.51	0.62
Japan	0.91	49	0.71	0.73	0.68	0.61	0.79	0.70	0.70	0.73	0.72	0.71	0.73	0.50	0.70
Jordan	0.90	46	0.74	0.77	0.79	0.52	0.78	0.65	0.53	0.74	0.73	0.65	0.55	0.48	0.79
Kazakhstan	0.94	58	0.76	0.67	0.69	0.76	0.74	0.80	0.82	0.79	0.83	0.80	0.80	0.60	0.80
Korea, Rep. of	0.96	66	0.82	0.84	0.79	0.87	0.86	0.74	0.81	0.80	0.83	0.79	0.80	0.68	0.86
Kuwait	0.95	62	0.86	0.84	0.72	0.81	0.84	0.73	0.81	0.78	0.87	0.72	0.74	0.60	0.87
Lebanon	0.94	61	0.85	0.85	0.84	0.67	0.86	0.82	0.82	0.36	0.84	0.79	0.75	0.77	0.76
Lithuania	0.88	42	0.70	0.59	0.52	0.55	0.54	0.64	0.76	0.75	0.62	0.69	0.78	0.58	0.63
Malaysia	0.91	48	0.76	0.77	0.73	0.61	0.72	0.71	0.56	0.50	0.78	0.71	0.72	0.64	0.74
Morocco	0.82	33	0.38	0.35	0.13	0.51	0.08	0.73	0.72	0.70	0.06	0.77	0.81	0.70	0.67
New Zealand	0.96	66	0.85	0.83	0.75	0.81	0.70	0.77	0.86	0.75	0.80	0.87	0.85	0.84	0.87
Norway (9)	0.84	36	0.63	0.43	0.63	0.62	0.74	0.73	0.71	0.52	0.28	0.66	0.58	0.40	0.63
Oman	0.95	63	0.79	0.80	0.81	0.81	0.82	0.82	0.80	0.82	0.85	0.81	0.79	0.55	0.84
Portugal	0.91	47	0.58	0.63	0.74	0.67	0.80	0.71	0.78	0.67	0.62	0.65	0.69	0.66	0.71
Qatar	0.99	86	0.95	0.93	0.91	0.94	0.95	0.95	0.94	0.95	0.93	0.86	0.91	0.89	0.91
Romania	0.90	45	0.77	0.75	0.73	0.72	0.69	0.69	0.61	0.63	0.62	0.62	0.65	0.62	0.59
Russian Federation	0.91	50	0.77	0.66	0.60	0.71	0.61	0.73	0.77	0.72	0.68	0.77	0.80	0.60	0.75
Saudi Arabia	0.92	53	0.74	0.80	0.57	0.80	0.78	0.77	0.77	0.75	0.75	0.70	0.54	0.64	0.75
Singapore	0.99	85	0.95	0.95	0.91	0.87	0.91	0.86	0.94	0.94	0.94	0.94	0.87	0.95	0.96
South Africa (9)	0.85	36	0.55	0.38	0.65	0.62	0.49	0.69	0.70	0.59	0.41	0.66	0.60	0.67	0.69
Sweden	0.87	40	0.77	0.74	0.60	0.57	0.59	0.63	0.72	0.72	0.58	0.64	0.40	0.51	0.67
Turkey	0.93	54	0.84	0.81	0.73	0.81	0.81	0.57	0.78	0.66	0.77	0.72	0.63	0.54	0.79
United Arab Emirates	0.97	76	0.88	0.87	0.89	0.92	0.90	0.91	0.91	0.90	0.87	0.85	0.81	0.76	0.88
United States	0.94	57	0.78	0.82	0.68	0.70	0.68	0.76	0.81	0.77	0.72	0.78	0.74	0.73	0.79

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Instruction Affected by Science Resource Shortages* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item												
			BCBG13AA	BCBG13AB	BCBG13AC	BCBG13AD	BCBG13AE	BCBG13AF	BCBG13AG	BCBG13AH	BCBG13CA	BCBG13CB	BCBG13CC	BCBG13CD	BCBG13CE
Benchmarking Participants															
Ontario, Canada	0.88	43	0.73	0.70	0.61	0.59	0.61	0.69	0.74	0.77	0.48	0.68	0.73	0.53	0.57
Quebec, Canada	0.93	56	0.80	0.83	0.76	0.61	0.81	0.70	0.80	0.61	0.84	0.62	0.72	0.71	0.85
Moscow City, Russian Fed.	0.95	66	0.82	0.78	0.77	0.80	0.82	0.84	0.85	0.84	0.83	0.84	0.87	0.67	0.84
Gauteng, RSA (9)	0.87	39	0.69	0.67	0.74	0.53	0.74	0.74	0.59	0.31	0.66	0.50	0.45	0.65	0.72
Western Cape, RSA (9)	0.89	43	0.76	0.68	0.71	0.68	0.62	0.71	0.77	0.65	0.68	0.56	0.42	0.59	0.64
Abu Dhabi, UAE	0.98	80	0.89	0.85	0.92	0.93	0.92	0.93	0.93	0.87	0.92	0.86	0.81	0.83	0.94
Dubai, UAE	0.98	79	0.89	0.92	0.85	0.94	0.88	0.90	0.92	0.93	0.92	0.87	0.87	0.80	0.86

Relationship Between the TIMSS 2019 *Instruction Affected by Science Resource Shortages* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Australia	0.25	0.06	0.03
Bahrain	-0.01	0.00	0.03
Chile	0.21	0.04	0.06
Chinese Taipei	0.06	0.00	0.00
Cyprus	0.01	0.00	0.01
Egypt	0.09	0.01	0.01
England	0.13	0.02	0.01
Finland	0.05	0.00	0.00
France	0.09	0.01	0.02
Georgia	0.06	0.00	0.00
Hong Kong SAR	0.09	0.01	0.03
Hungary	-0.02	0.00	0.00
Iran, Islamic Rep. of	0.11	0.01	0.04
Ireland	0.12	0.01	0.01
Israel	0.12	0.02	0.01
Italy	0.03	0.00	0.00
Japan	0.08	0.01	0.01
Jordan	-0.12	0.01	0.01
Kazakhstan	0.07	0.00	0.01
Korea, Rep. of	0.02	0.00	0.00
Kuwait	-0.01	0.00	0.02
Lebanon	0.17	0.03	0.10
Lithuania	-0.02	0.00	0.00
Malaysia	-0.08	0.01	0.01
Morocco	-0.03	0.00	0.03
New Zealand	0.21	0.04	0.03
Norway (9)	0.01	0.00	0.00
Oman	0.05	0.00	0.01
Portugal	0.10	0.01	0.01
Qatar	0.20	0.04	0.05
Romania	0.16	0.03	0.02
Russian Federation	0.08	0.01	0.02
Saudi Arabia	-0.05	0.00	0.01
Singapore	0.04	0.00	0.00
South Africa (9)	0.21	0.05	0.09
Sweden	0.04	0.00	0.00
Turkey	0.18	0.03	0.02
United Arab Emirates	0.20	0.04	0.05
United States	0.12	0.01	0.02
International Median	**0.08**	**0.01**	**0.01**

Relationship Between the TIMSS 2019 *Instruction Affected by Science Resource Shortages* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Benchmarking Participants			
Ontario, Canada	0.08	0.01	0.02
Quebec, Canada	-0.02	0.00	0.01
Moscow City, Russian Fed.	0.00	0.00	0.00
Gauteng, RSA (9)	0.32	0.10	0.15
Western Cape, RSA (9)	0.47	0.22	0.24
Abu Dhabi, UAE	0.16	0.03	0.04
Dubai, UAE	0.19	0.04	0.05

School Discipline – Grade 8

About the Scale

The *School Discipline* scale was created based on principals' responses to eleven items listed below.

To what degree is each of the following a problem among eighth grade students in your school?

		Not a problem	Minor problem	Moderate problem	Severe Problem
BCBG16A [T]	1) Arriving late at school	○	○	○	○
BCBG16B [T]	2) Absenteeism (i.e., unjustified absences)	○	○	○	○
BCBG16C [T]	3) Classroom disturbance	○	○	○	○
BCBG16D [T]	4) Cheating	○	○	○	○
BCBG16E [T]	5) Profanity	○	○	○	○
BCBG16F [T]	6) Vandalism	○	○	○	○
BCBG16G [T]	7) Theft	○	○	○	○
BCBG16H [T]	8) Intimidation or verbal abuse among students (including texting, emailing, etc.)	○	○	○	○
BCBG16I [T]	9) Physical injury to other students	○	○	○	○
BCBG16J [T]	10) Intimidation or verbal abuse of teachers or staff (including texting, emailing, etc.)	○	○	○	○
BCBG16K [T]	11) Physical injury to teachers or staff	○	○	○	○

Hardly Any Problems | Minor Problems | Moderate to Severe Problems

Scale Cut Scores **10.8** **8.0**

[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *School Discipline* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BCBG16A	0.32951	-2.48171	-0.42419	2.90590	1.27
BCBG16B	0.89186	-1.91504	-0.57924	2.49428	1.28
BCBG16C	0.79991	-2.31615	-0.33558	2.65173	0.97
BCBG16D	-0.17728	-1.38113	-0.75069	2.13182	1.09
BCBG16E	0.54617	-1.88683	-0.34275	2.22958	0.92
BCBG16F	-0.05891	-0.80027	-0.74541	1.54568	0.83
BCBG16G	-0.59715	-0.18031	-1.07997	1.26028	0.82
BCBG16H	0.45456	-1.89207	-0.59214	2.48421	1.02
BCBG16I	-0.30789	-0.85864	-0.95601	1.81465	0.80
BCBG16J	-0.64994	-0.36374	-0.79318	1.15692	0.82
BCBG16K	-1.23084	1.01064	-0.65683	-0.35381	0.88

Scale Transformation Constants for the TIMSS 2019 *School Discipline* Scale – Grade 8

Scale Transformation Constants	
A = 8.418512	Transformed Scale Score = 8.418512 + 0.982377 • Logit Scale Score
B = 0.982377	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *School Discipline* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	4.10007	
1	5.19151	
2	5.69640	
3	6.02711	
4	6.27380	
5	6.47245	
6	6.64117	
7	6.79025	
8	6.92622	
9	7.05373	
10	7.17625	
11	7.29658	
12	7.41582	
13	7.53896	
14	7.66553	
15	7.79736	
16	7.93621	8.0
17	8.08398	
18	8.24184	
19	8.41207	
20	8.59633	
21	8.79625	
22	9.01493	
23	9.25300	
24	9.51280	
25	9.79607	
26	10.10482	
27	10.43939	
28	10.80136	10.8
29	11.19421	
30	11.62884	
31	12.13532	
32	12.80218	
33	14.02858	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *School Discipline* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item										
			BCBG16A	BCBG16B	BCBG16C	BCBG16D	BCBG16E	BCBG16F	BCBG16G	BCBG16H	BCBG16I	BCBG16J	BCBG16K
Australia	0.92	57	0.71	0.74	0.82	0.65	0.85	0.78	0.61	0.76	0.86	0.83	0.66
Bahrain	0.96	73	0.76	0.76	0.79	0.86	0.87	0.88	0.90	0.87	0.90	0.90	0.88
Chile	0.90	52	0.56	0.64	0.77	0.76	0.80	0.73	0.76	0.72	0.80	0.72	0.60
Chinese Taipei	0.88	48	0.68	0.73	0.60	0.75	0.69	0.72	0.80	0.68	0.75	0.67	0.46
Cyprus	0.94	61	0.58	0.68	0.77	0.82	0.83	0.85	0.80	0.83	0.88	0.83	0.69
Egypt	0.96	70	0.64	0.71	0.82	0.81	0.84	0.91	0.88	0.89	0.92	0.88	0.86
England	0.82	36	0.67	0.65	0.68	0.46	0.71	0.68	0.44	0.50	0.65	0.62	0.47
Finland	0.85	41	0.59	0.65	0.67	0.61	0.70	0.66	0.50	0.71	0.70	0.67	0.54
France	0.88	46	0.64	0.75	0.78	0.52	0.81	0.58	0.60	0.72	0.72	0.75	0.49
Georgia	0.95	69	0.68	0.80	0.78	0.62	0.85	0.88	0.90	0.90	0.91	0.90	0.88
Hong Kong SAR	0.81	37	0.65	0.53	0.60	0.72	0.74	0.66	0.68	0.58	0.63	0.41	0.37
Hungary	0.89	47	0.69	0.70	0.81	0.69	0.82	0.78	0.49	0.70	0.71	0.71	0.14
Iran, Islamic Rep. of	0.90	52	0.64	0.61	0.75	0.73	0.82	0.78	0.75	0.78	0.71	0.73	0.60
Ireland	0.88	45	0.71	0.73	0.76	0.63	0.79	0.72	0.59	0.74	0.60	0.74	0.11
Israel	0.92	56	0.63	0.65	0.70	0.71	0.79	0.78	0.79	0.74	0.77	0.84	0.80
Italy	0.91	52	0.57	0.74	0.64	0.64	0.68	0.80	0.74	0.75	0.84	0.79	0.71
Japan	0.92	63	0.62	0.15	0.86	0.80	0.86	0.92	0.79	0.67	0.88	0.88	0.93
Jordan	0.95	68	0.55	0.70	0.81	0.84	0.88	0.89	0.86	0.87	0.89	0.88	0.79
Kazakhstan	0.98	82	0.74	0.91	0.89	0.72	0.89	0.95	0.96	0.96	0.96	0.96	0.96
Korea, Rep. of	0.95	67	0.77	0.86	0.75	0.76	0.83	0.85	0.81	0.83	0.87	0.85	0.78
Kuwait	0.95	67	0.60	0.72	0.79	0.87	0.84	0.89	0.86	0.85	0.83	0.85	0.83
Lebanon	0.97	77	0.83	0.86	0.81	0.91	0.87	0.92	0.93	0.80	0.88	0.90	0.92
Lithuania	0.83	38	0.62	0.61	0.65	0.62	0.73	0.66	0.49	0.60	0.74	0.63	0.22
Malaysia	0.92	58	0.59	0.64	0.76	0.85	0.82	0.82	0.84	0.81	0.84	0.75	0.59
Morocco	0.96	72	0.59	0.72	0.87	0.87	0.89	0.89	0.92	0.85	0.91	0.89	0.86
New Zealand	0.90	52	0.71	0.74	0.79	0.59	0.78	0.79	0.75	0.71	0.78	0.75	0.42
Norway (9)	0.88	46	0.72	0.76	0.67	0.68	0.69	0.64	0.69	0.75	0.65	0.70	0.49
Oman	0.97	80	0.80	0.82	0.83	0.91	0.90	0.93	0.94	0.91	0.92	0.92	0.91
Portugal	0.92	56	0.61	0.68	0.76	0.70	0.70	0.84	0.79	0.76	0.82	0.84	0.67
Qatar	0.97	75	0.77	0.86	0.86	0.92	0.87	0.88	0.83	0.89	0.87	0.89	0.87
Romania	0.94	65	0.76	0.67	0.80	0.81	0.76	0.87	0.90	0.75	0.80	0.88	0.83
Russian Federation	0.81	36	0.52	0.48	0.61	0.56	0.67	0.60	0.68	0.68	0.68	0.63	0.39
Saudi Arabia	0.96	72	0.53	0.68	0.80	0.89	0.90	0.92	0.93	0.87	0.92	0.91	0.87
Singapore	0.86	43	0.64	0.68	0.70	0.64	0.75	0.71	0.74	0.56	0.73	0.68	0.30
South Africa (9)	0.91	53	0.67	0.70	0.77	0.71	0.77	0.77	0.75	0.80	0.80	0.69	0.56
Sweden	0.85	40	0.63	0.69	0.63	0.52	0.73	0.64	0.55	0.67	0.69	0.63	0.49
Turkey	0.96	72	0.69	0.75	0.86	0.86	0.81	0.87	0.89	0.88	0.92	0.91	0.88
United Arab Emirates	0.92	57	0.62	0.66	0.73	0.80	0.77	0.82	0.75	0.80	0.77	0.78	0.76
United States	0.89	49	0.70	0.67	0.76	0.45	0.78	0.69	0.71	0.71	0.80	0.79	0.57

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *School Discipline* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item										
			BCBG16A	BCBG16B	BCBG16C	BCBG16D	BCBG16E	BCBG16F	BCBG16G	BCBG16H	BCBG16I	BCBG16J	BCBG16K
Benchmarking Participants													
Ontario, Canada	0.88	48	0.67	0.57	0.73	0.59	0.81	0.75	0.75	0.75	0.71	0.78	0.44
Quebec, Canada	0.87	45	0.63	0.72	0.75	0.51	0.83	0.67	0.57	0.74	0.67	0.77	0.34
Moscow City, Russian Fed.	0.79	35	0.59	0.54	0.67	0.61	0.73	0.56	0.52	0.65	0.56	0.44	-
Gauteng, RSA (9)	0.92	54	0.74	0.74	0.75	0.69	0.74	0.80	0.79	0.81	0.77	0.60	0.62
Western Cape, RSA (9)	0.93	58	0.77	0.80	0.78	0.76	0.83	0.80	0.81	0.78	0.80	0.71	0.50
Abu Dhabi, UAE	0.92	58	0.70	0.69	0.80	0.83	0.76	0.84	0.63	0.83	0.76	0.79	0.71
Dubai, UAE	0.85	45	0.45	0.59	0.59	0.77	0.71	0.83	0.78	0.63	0.64	0.55	0.74

Relationship Between the TIMSS 2019 *School Discipline* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Australia	0.35	0.12	0.12	0.32	0.10	0.10
Bahrain	0.08	0.01	0.01	0.07	0.01	0.02
Chile	0.26	0.07	0.06	0.25	0.06	0.06
Chinese Taipei	0.12	0.01	0.01	0.10	0.01	0.01
Cyprus	0.24	0.06	0.08	0.21	0.04	0.07
Egypt	0.06	0.00	0.01	0.06	0.00	0.01
England	0.22	0.05	0.02	0.22	0.05	0.02
Finland	0.11	0.01	0.01	0.12	0.01	0.02
France	0.28	0.08	0.08	0.28	0.08	0.08
Georgia	0.13	0.02	0.02	0.12	0.01	0.02
Hong Kong SAR	0.24	0.06	0.02	0.17	0.03	0.00
Hungary	0.28	0.08	0.06	0.27	0.07	0.06
Iran, Islamic Rep. of	0.23	0.05	0.04	0.22	0.05	0.03
Ireland	0.24	0.06	0.04	0.24	0.06	0.03
Israel	0.22	0.05	0.05	0.20	0.04	0.04
Italy	0.08	0.01	0.01	0.08	0.01	0.01
Japan	0.08	0.01	0.00	0.07	0.01	0.00
Jordan	0.09	0.01	0.01	0.10	0.01	0.02
Kazakhstan	0.10	0.01	0.01	0.07	0.00	0.00
Korea, Rep. of	0.01	0.00	0.00	0.02	0.00	0.00
Kuwait	0.15	0.02	0.03	0.15	0.02	0.03
Lebanon	0.04	0.00	0.01	0.06	0.00	0.01
Lithuania	0.09	0.01	0.00	0.08	0.01	0.00
Malaysia	0.20	0.04	0.03	0.17	0.03	0.03
Morocco	0.01	0.00	0.01	-0.02	0.00	0.01
New Zealand	0.28	0.08	0.05	0.25	0.06	0.04
Norway (9)	0.11	0.01	0.00	0.12	0.02	0.01
Oman	0.12	0.01	0.02	0.10	0.01	0.01
Portugal	0.11	0.01	0.01	0.09	0.01	0.01
Qatar	0.09	0.01	0.01	0.04	0.00	0.00
Romania	0.18	0.03	0.03	0.18	0.03	0.03
Russian Federation	0.03	0.00	0.00	0.02	0.00	0.00
Saudi Arabia	0.10	0.01	0.01	0.12	0.01	0.01
Singapore	0.15	0.02	0.02	0.14	0.02	0.02
South Africa (9)	0.17	0.03	0.04	0.15	0.02	0.03
Sweden	0.16	0.03	0.02	0.16	0.03	0.02
Turkey	0.14	0.02	0.03	0.13	0.02	0.02
United Arab Emirates	0.26	0.07	0.06	0.28	0.08	0.07
United States	0.28	0.08	0.07	0.28	0.08	0.07
International Median	**0.14**	**0.02**	**0.02**	**0.13**	**0.02**	**0.02**

Relationship Between the TIMSS 2019 *School Discipline* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		*r*	r^2	
Benchmarking Participants						
Ontario, Canada	0.20	0.04	0.02	0.15	0.02	0.02
Quebec, Canada	0.15	0.02	0.01	0.12	0.01	0.01
Moscow City, Russian Fed.	0.02	0.00	0.00	0.00	0.00	0.00
Gauteng, RSA (9)	0.39	0.15	0.17	0.39	0.16	0.17
Western Cape, RSA (9)	0.58	0.34	0.31	0.58	0.33	0.31
Abu Dhabi, UAE	0.23	0.05	0.04	0.26	0.07	0.06
Dubai, UAE	0.23	0.05	0.04	0.23	0.05	0.03

School Emphasis on Academic Success—Principals' Reports – Grade 8

About the Scale

The *School Emphasis on Academic Success* scale was created based on principals' responses to eleven items listed below.[1]

How would you characterize each of the following within your school?

		Very high	High	Medium	Low	Very low
BCBG14A T	1) Teachers' understanding of the school's curricular goals	○	○	○	○	○
BCBG14B T	2) Teachers' degree of success in implementing the school's curriculum	○	○	○	○	○
BCBG14C T	3) Teachers' expectations for student achievement	○	○	○	○	○
BCBG14D T	4) Teachers' ability to inspire students	○	○	○	○	○
BCBG14E T	5) Parental involvement in school activities	○	○	○	○	○
BCBG14F T	6) Parental commitment to ensure that students are ready to learn	○	○	○	○	○
BCBG14G T	7) Parental expectations for student achievement	○	○	○	○	○
BCBG14H T	8) Parental support for student achievement	○	○	○	○	○
BCBG14I T	9) Students' desire to do well in school	○	○	○	○	○
BCBG14J T	10) Students' ability to reach school's academic goals	○	○	○	○	○
BCBG14K T	11) Students' respect for classmates who excel academically	○	○	○	○	○

Very High Emphasis | High Emphasis | Medium Emphasis

Scale Cut Scores 13.1 9.6

[1] For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Low" and "Very low" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

T Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *School Emphasis on Academic Success—Principals' Reports* Scale – Grade 8

Item	delta	tau_1	tau_2	tau_3	Infit
BCBG14A	-1.68808	-3.41252	0.01816	3.39436	1.19
BCBG14B	-1.24407	-3.72289	0.13877	3.58412	1.00
BCBG14C	-0.83743	-3.38294	-0.01936	3.40230	0.98
BCBG14D	-0.60365	-3.45458	0.05816	3.39642	1.03
BCBG14E	1.62892	-2.42516	0.09836	2.32680	1.17
BCBG14F	1.44778	-2.60585	0.04277	2.56308	0.82
BCBG14G	-0.09004	-2.61114	-0.18026	2.79140	1.02
BCBG14H	1.20810	-2.67184	0.05169	2.62015	0.85
BCBG14I	0.05948	-3.47526	0.16895	3.30631	0.86
BCBG14J	0.43096	-3.76989	0.24784	3.52205	0.85
BCBG14K	-0.31197	-2.97209	-0.24200	3.21409	1.20

Scale Transformation Constants for the TIMSS 2019 *School Emphasis on Academic Success—Principals' Reports* Scale – Grade 8

Scale Transformation Constants	
A = 9.585866	Transformed Scale Score = 9.585866 + 1.044298 • Logit Scale Score
B = 1.044298	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019
***School Emphasis on Academic Success—Principals' Reports* Scale – Grade 8**

Raw Score	Transformed Scale Score	Cutpoint
1	3.55446	
2	4.30371	
3	4.88060	
4	5.37607	
5	5.82698	
6	6.24903	
7	6.65002	
8	7.03177	
9	7.39368	
10	7.73586	
11	8.06046	
12	8.37165	
13	8.67341	
14	8.96990	
15	9.26475	
16	9.56054	9.6
17	9.85898	
18	10.15969	
19	10.46161	
20	10.76309	
21	11.06274	
22	11.35999	
23	11.65536	
24	11.95038	
25	12.24755	
26	12.55044	
27	12.86324	
28	13.19357	13.1
29	13.55081	
30	13.95351	
31	14.43815	
32	15.09677	
33	16.36222	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *School Emphasis on Academic Success—Principals' Reports* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item										
			BCBG14A	BCBG14B	BCBG14C	BCBG14D	BCBG14E	BCBG14F	BCBG14G	BCBG14H	BCBG14I	BCBG14J	BCBG14K
Australia	0.94	65	0.68	0.76	0.81	0.73	0.79	0.85	0.83	0.87	0.87	0.81	0.83
Bahrain	0.94	63	0.72	0.76	0.77	0.85	0.69	0.87	0.84	0.88	0.81	0.85	0.69
Chile	0.95	66	0.75	0.79	0.81	0.81	0.77	0.83	0.83	0.86	0.86	0.84	0.77
Chinese Taipei	0.93	61	0.64	0.77	0.77	0.78	0.77	0.81	0.85	0.82	0.84	0.76	0.73
Cyprus	0.89	50	0.39	0.73	0.70	0.53	0.59	0.78	0.83	0.82	0.77	0.82	0.66
Egypt	0.89	49	0.46	0.64	0.63	0.67	0.71	0.82	0.78	0.81	0.74	0.74	0.59
England	0.93	58	0.64	0.70	0.75	0.71	0.70	0.85	0.80	0.82	0.85	0.83	0.71
Finland	0.87	44	0.67	0.62	0.63	0.64	0.65	0.75	0.71	0.72	0.66	0.68	0.57
France	0.86	42	0.39	0.54	0.56	0.54	0.69	0.82	0.65	0.80	0.71	0.74	0.56
Georgia	0.91	51	0.76	0.76	0.74	0.77	0.73	0.78	0.62	0.72	0.74	0.65	0.56
Hong Kong SAR	0.92	55	0.62	0.71	0.78	0.78	0.67	0.83	0.78	0.82	0.77	0.79	0.54
Hungary	0.91	54	0.47	0.72	0.63	0.77	0.73	0.87	0.69	0.86	0.81	0.76	0.67
Iran, Islamic Rep. of	0.89	49	0.71	0.69	0.69	0.69	0.69	0.77	0.59	0.76	0.75	0.76	0.55
Ireland	0.92	59	0.65	0.70	0.82	0.67	0.76	0.83	0.80	0.80	0.85	0.84	0.68
Israel	0.88	46	0.54	0.62	0.69	0.60	0.60	0.72	0.77	0.79	0.71	0.72	0.66
Italy	0.84	40	0.53	0.64	0.61	0.56	0.65	0.74	0.56	0.72	0.71	0.65	0.51
Japan	0.91	52	0.55	0.57	0.73	0.71	0.67	0.85	0.72	0.83	0.78	0.81	0.65
Jordan	0.91	53	0.66	0.74	0.63	0.72	0.70	0.80	0.77	0.78	0.78	0.76	0.64
Kazakhstan	0.89	50	0.63	0.68	0.68	0.79	0.67	0.79	0.68	0.78	0.73	0.62	0.69
Korea, Rep. of	0.91	54	0.62	0.64	0.69	0.53	0.78	0.83	0.78	0.81	0.77	0.81	0.72
Kuwait	0.91	53	0.61	0.69	0.71	0.72	0.68	0.80	0.79	0.75	0.80	0.75	0.68
Lebanon	0.87	44	0.58	0.54	0.66	0.59	0.67	0.73	0.68	0.76	0.65	0.78	0.56
Lithuania	0.85	41	0.54	0.62	0.61	0.68	0.70	0.77	0.58	0.77	0.62	0.57	0.48
Malaysia	0.91	51	0.54	0.61	0.60	0.65	0.77	0.80	0.73	0.80	0.83	0.81	0.69
Morocco	0.88	47	0.66	0.63	0.68	0.70	0.63	0.80	0.73	0.73	0.67	0.72	0.55
New Zealand	0.92	56	0.68	0.80	0.77	0.73	0.65	0.77	0.79	0.81	0.81	0.76	0.63
Norway (9)	0.89	48	0.49	0.60	0.74	0.59	0.77	0.72	0.76	0.71	0.78	0.77	0.61
Oman	0.89	49	0.66	0.68	0.61	0.64	0.67	0.83	0.74	0.77	0.68	0.77	0.61
Portugal	0.91	52	0.59	0.74	0.75	0.57	0.68	0.81	0.74	0.81	0.77	0.79	0.63
Qatar	0.92	56	0.67	0.76	0.78	0.85	0.79	0.77	0.72	0.70	0.74	0.81	0.62
Romania	0.88	46	0.51	0.64	0.63	0.71	0.73	0.76	0.58	0.73	0.75	0.81	0.56
Russian Federation	0.84	39	0.56	0.67	0.61	0.60	0.75	0.68	0.49	0.73	0.67	0.59	0.42
Saudi Arabia	0.92	53	0.64	0.67	0.76	0.73	0.72	0.75	0.75	0.78	0.77	0.79	0.66
Singapore	0.94	61	0.70	0.73	0.81	0.75	0.70	0.84	0.82	0.84	0.82	0.84	0.73
South Africa (9)	0.88	47	0.52	0.68	0.63	0.63	0.77	0.81	0.62	0.81	0.69	0.69	0.67
Sweden	0.89	49	0.61	0.60	0.74	0.68	0.74	0.82	0.74	0.79	0.74	0.75	0.46
Turkey	0.90	51	0.66	0.60	0.74	0.63	0.80	0.81	0.62	0.83	0.75	0.80	0.56
United Arab Emirates	0.94	65	0.75	0.78	0.85	0.83	0.76	0.85	0.75	0.83	0.85	0.84	0.74
United States	0.93	60	0.68	0.77	0.77	0.80	0.80	0.84	0.82	0.82	0.80	0.72	0.69

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *School Emphasis on Academic Success—Principals' Reports* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item										
			BCBG14A	BCBG14B	BCBG14C	BCBG14D	BCBG14E	BCBG14F	BCBG14G	BCBG14H	BCBG14I	BCBG14J	BCBG14K
Benchmarking Participants													
Ontario, Canada	0.90	53	0.54	0.68	0.76	0.61	0.63	0.77	0.77	0.83	0.82	0.76	0.74
Quebec, Canada	0.91	53	0.68	0.69	0.65	0.75	0.62	0.79	0.75	0.80	0.83	0.77	0.68
Moscow City, Russian Fed.	0.86	42	0.63	0.68	0.60	0.67	0.70	0.78	0.48	0.69	0.68	0.60	0.57
Gauteng, RSA (9)	0.90	50	0.64	0.74	0.63	0.63	0.72	0.77	0.63	0.78	0.73	0.76	0.71
Western Cape, RSA (9)	0.91	54	0.58	0.68	0.63	0.68	0.79	0.84	0.63	0.83	0.83	0.82	0.74
Abu Dhabi, UAE	0.95	65	0.72	0.74	0.83	0.82	0.80	0.85	0.74	0.86	0.87	0.85	0.78
Dubai, UAE	0.94	62	0.73	0.76	0.85	0.80	0.70	0.82	0.75	0.83	0.84	0.82	0.71

Relationship Between the TIMSS 2019 *School Emphasis on Academic Success—Principals' Reports* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Australia	0.45	0.20	0.17	0.39	0.15	0.11
Bahrain	0.14	0.02	0.02	0.32	0.10	0.10
Chile	0.35	0.12	0.13	0.34	0.11	0.12
Chinese Taipei	0.26	0.07	0.05	0.22	0.05	0.04
Cyprus	0.21	0.05	0.04	0.18	0.03	0.03
Egypt	0.19	0.03	0.03	0.20	0.04	0.03
England	0.29	0.08	0.18	0.26	0.07	0.15
Finland	0.14	0.02	0.01	0.13	0.02	0.01
France	0.26	0.07	0.05	0.25	0.06	0.04
Georgia	0.12	0.01	0.02	0.05	0.00	0.00
Hong Kong SAR	0.41	0.17	0.15	0.35	0.12	0.12
Hungary	0.44	0.20	0.12	0.42	0.17	0.10
Iran, Islamic Rep. of	0.34	0.12	0.10	0.34	0.12	0.10
Ireland	0.29	0.08	0.08	0.26	0.07	0.07
Israel	0.20	0.04	0.04	0.22	0.05	0.04
Italy	0.13	0.02	0.00	0.13	0.02	0.00
Japan	0.23	0.05	0.06	0.20	0.04	0.04
Jordan	0.24	0.06	0.05	0.25	0.06	0.05
Kazakhstan	0.05	0.00	0.00	0.02	0.00	0.00
Korea, Rep. of	0.15	0.02	0.02	0.11	0.01	0.02
Kuwait	0.28	0.08	0.06	0.29	0.09	0.08
Lebanon	0.33	0.11	0.09	0.36	0.13	0.11
Lithuania	0.15	0.02	0.02	0.14	0.02	0.01
Malaysia	0.30	0.09	0.07	0.30	0.09	0.07
Morocco	0.17	0.03	0.03	0.13	0.02	0.02
New Zealand	0.31	0.10	0.07	0.29	0.08	0.06
Norway (9)	0.14	0.02	0.01	0.11	0.01	0.01
Oman	0.18	0.03	0.03	0.16	0.03	0.03
Portugal	0.32	0.10	0.05	0.28	0.08	0.04
Qatar	0.24	0.06	0.08	0.19	0.04	0.05
Romania	0.26	0.07	0.05	0.24	0.06	0.04
Russian Federation	0.19	0.04	0.03	0.17	0.03	0.02
Saudi Arabia	0.16	0.02	0.02	0.19	0.03	0.03
Singapore	0.37	0.14	0.14	0.36	0.13	0.13
South Africa (9)	0.27	0.07	0.04	0.27	0.07	0.04
Sweden	0.25	0.06	0.05	0.24	0.06	0.05
Turkey	0.28	0.08	0.07	0.26	0.07	0.06
United Arab Emirates	0.36	0.13	0.11	0.37	0.14	0.12
United States	0.31	0.10	0.09	0.29	0.08	0.08
International Median	**0.26**	**0.07**	**0.05**	**0.25**	**0.06**	**0.04**

Relationship Between the TIMSS 2019 *School Emphasis on Academic Success—Principals' Reports* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Benchmarking Participants						
Ontario, Canada	0.24	0.06	0.05	0.17	0.03	0.03
Quebec, Canada	0.31	0.09	0.10	0.27	0.07	0.08
Moscow City, Russian Fed.	0.06	0.00	0.01	0.06	0.00	0.01
Gauteng, RSA (9)	0.29	0.08	0.04	0.29	0.09	0.04
Western Cape, RSA (9)	0.55	0.30	0.25	0.54	0.29	0.25
Abu Dhabi, UAE	0.37	0.13	0.13	0.40	0.16	0.16
Dubai, UAE	0.30	0.09	0.07	0.27	0.07	0.05

Scales Based on Teachers' Reports

Classroom Teaching Limited by Students Not Ready for Instruction – Grade 8

About the Scale

The *Classroom Teaching Limited by Students Not Ready for Instruction* scale was created based on teachers' responses to eight items listed below.

In your view, to what extent do the following limit how you teach this class?

		Not at all	Some	A lot
BTBG13A [T]	1) Students lacking prerequisite knowledge or skills	◯	◯	◯
BTBG13B [T]	2) Students suffering from lack of basic nutrition	◯	◯	◯
BTBG13C [T]	3) Students suffering from not enough sleep	◯	◯	◯
BTBG13D	4) Students absent from class	◯	◯	◯
BTBG13E [T]	5) Disruptive students	◯	◯	◯
BTBG13F [T]	6) Uninterested students	◯	◯	◯
BTBG13G [T]	7) Students with mental, emotional, or psychological impairment	◯	◯	◯
BTBG13H	8) Students with difficulties understanding the language of instruction	◯	◯	◯

	Very Little		Some		A Lot
Scale Cut Scores		11.2		7.2	

[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Classroom Teaching Limited by Students Not Ready for Instruction* Scale – Grade 8

Item	delta	tau_1	tau_2	Infit
BTBG13A	1.09871	-1.88960	1.88960	1.02
BTBG13B	-1.09149	-1.20078	1.20078	1.11
BTBG13C	0.14688	-1.69582	1.69582	0.99
BTBG13D	0.14205	-1.54079	1.54079	0.98
BTBG13E	0.27077	-1.26887	1.26887	0.95
BTBG13F	0.84736	-1.78301	1.78301	0.91
BTBG13G	-0.61321	-1.43379	1.43379	0.99
BTBG13H	-0.80107	-1.14438	1.14438	1.08

Scale Transformation Constants for the TIMSS 2019 *Classroom Teaching Limited by Students Not Ready for Instruction* Scale – Grade 8

Scale Transformation Constants	
A = 9.253899	Transformed Scale Score = 9.253899 + 1.261053 • Logit Scale Score
B = 1.261053	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Classroom Teaching Limited by Students Not Ready for Instruction* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	3.62369	
1	5.18385	
2	6.01647	
3	6.64377	
4	7.18242	7.2
5	7.67828	
6	8.15603	
7	8.63238	
8	9.11668	
9	9.61910	
10	10.14509	
11	10.69865	
12	11.28828	11.2
13	11.93193	
14	12.67245	
15	13.61824	
16	15.29510	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Classroom Teaching Limited by Students Not Ready for Instruction* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			BTBG13A	BTBG13B	BTBG13C	BTBG13D	BTBG13E	BTBG13F	BTBG13G	BTBG13H
Australia	0.83	46	0.69	0.60	0.65	0.67	0.75	0.76	0.66	0.61
Bahrain	0.76	38	0.63	0.60	0.64	0.39	0.71	0.73	0.58	0.55
Chile	0.84	47	0.67	0.72	0.73	0.73	0.71	0.76	0.64	0.53
Chinese Taipei	0.80	43	0.59	0.56	0.63	0.70	0.69	0.62	0.71	0.71
Cyprus	0.80	43	0.57	0.58	0.69	0.68	0.65	0.69	0.70	0.69
Egypt	0.68	31	0.49	0.42	0.35	0.55	0.72	0.71	0.60	0.55
England	0.82	45	0.58	0.70	0.76	0.64	0.72	0.72	0.62	0.63
Finland	0.78	40	0.72	0.44	0.68	0.62	0.63	0.73	0.70	0.46
France	0.75	36	0.58	0.56	0.61	0.70	0.57	0.59	0.65	0.53
Georgia	0.71	33	0.56	0.43	0.43	0.56	0.66	0.69	0.66	0.57
Hong Kong SAR	0.71	34	0.65	0.33	0.51	0.62	0.65	0.69	0.69	0.40
Hungary	0.78	40	0.62	0.55	0.66	0.58	0.69	0.71	0.74	0.47
Iran, Islamic Rep. of	0.76	38	0.50	0.59	0.60	0.59	0.69	0.61	0.75	0.55
Ireland	0.81	43	0.54	0.63	0.69	0.65	0.69	0.74	0.66	0.61
Israel	0.86	51	0.63	0.64	0.72	0.78	0.72	0.74	0.71	0.72
Italy	0.80	42	0.70	0.65	0.66	0.66	0.70	0.64	0.54	0.65
Japan	0.78	40	0.68	0.21	0.63	0.66	0.61	0.71	0.65	0.72
Jordan	0.71	34	0.57	0.40	0.56	0.59	0.67	0.73	0.47	0.60
Kazakhstan	0.83	46	0.62	0.53	0.71	0.70	0.74	0.71	0.76	0.62
Korea, Rep. of	0.77	39	0.37	0.53	0.66	0.62	0.67	0.64	0.73	0.68
Kuwait	0.68	32	0.43	0.33	0.60	0.60	0.69	0.74	0.55	0.47
Lebanon	0.75	37	0.57	0.49	0.60	0.64	0.64	0.70	0.59	0.60
Lithuania	0.85	49	0.63	0.68	0.73	0.76	0.74	0.73	0.72	0.60
Malaysia	0.84	48	0.67	0.58	0.63	0.73	0.78	0.82	0.69	0.57
Morocco	0.69	31	0.45	0.52	0.51	0.56	0.66	0.65	0.55	0.55
New Zealand	0.79	41	0.57	0.63	0.64	0.66	0.69	0.73	0.71	0.47
Norway (9)	0.74	36	0.55	0.63	0.71	0.67	0.64	0.65	0.47	0.41
Oman	0.75	37	0.53	0.54	0.56	0.62	0.61	0.64	0.65	0.67
Portugal	0.77	39	0.56	0.60	0.65	0.68	0.62	0.76	0.55	0.56
Qatar	0.78	39	0.62	0.56	0.70	0.61	0.70	0.74	0.50	0.56
Romania	0.78	40	0.59	0.57	0.55	0.67	0.71	0.67	0.65	0.60
Russian Federation	0.84	48	0.61	0.63	0.69	0.78	0.75	0.71	0.75	0.58
Saudi Arabia	0.78	40	0.55	0.60	0.64	0.67	0.70	0.66	0.61	0.61
Singapore	0.80	42	0.60	0.53	0.55	0.72	0.69	0.70	0.70	0.64
South Africa (9)	0.75	37	0.47	0.54	0.58	0.62	0.71	0.74	0.63	0.54
Sweden	0.75	36	0.65	0.37	0.64	0.57	0.69	0.68	0.62	0.54
Turkey	0.83	46	0.62	0.69	0.66	0.68	0.68	0.71	0.74	0.63
United Arab Emirates	0.84	48	0.64	0.54	0.70	0.67	0.77	0.78	0.69	0.71
United States	0.80	42	0.61	0.66	0.65	0.64	0.73	0.72	0.70	0.48

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Classroom Teaching Limited by Students Not Ready for Instruction* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			BTBG13A	BTBG13B	BTBG13C	BTBG13D	BTBG13E	BTBG13F	BTBG13G	BTBG13H
Benchmarking Participants										
Ontario, Canada	0.81	45	0.61	0.67	0.74	0.73	0.70	0.76	0.69	0.36
Quebec, Canada	0.80	42	0.61	0.68	0.70	0.71	0.59	0.74	0.68	0.42
Moscow City, Russian Fed.	0.87	52	0.65	0.63	0.72	0.78	0.76	0.73	0.77	0.71
Gauteng, RSA (9)	0.79	41	0.46	0.66	0.58	0.63	0.69	0.74	0.63	0.66
Western Cape, RSA (9)	0.82	45	0.58	0.67	0.58	0.64	0.77	0.79	0.69	0.58
Abu Dhabi, UAE	0.84	47	0.70	0.40	0.67	0.68	0.77	0.77	0.69	0.75
Dubai, UAE	0.84	48	0.52	0.61	0.71	0.68	0.77	0.79	0.71	0.71

Relationship Between the TIMSS 2019 *Classroom Teaching Limited by Students Not Ready for Instruction* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Australia	0.49	0.24	0.20	0.36	0.13	0.10
Bahrain	0.05	0.00	0.00	0.21	0.04	0.04
Chile	0.26	0.07	0.05	0.21	0.04	0.04
Chinese Taipei	0.13	0.02	0.02	0.08	0.01	0.01
Cyprus	0.16	0.02	0.02	0.10	0.01	0.02
Egypt	0.11	0.01	0.01	0.05	0.00	0.00
England	0.34	0.12	0.12	0.23	0.05	0.08
Finland	0.25	0.06	0.04	0.22	0.05	0.04
France	0.24	0.06	0.04	0.19	0.04	0.04
Georgia	0.01	0.00	0.00	0.11	0.01	0.01
Hong Kong SAR	0.45	0.20	0.14	0.26	0.07	0.06
Hungary	0.34	0.11	0.10	0.29	0.09	0.07
Iran, Islamic Rep. of	0.21	0.04	0.02	0.05	0.00	0.00
Ireland	0.36	0.13	0.09	0.24	0.06	0.04
Israel	0.34	0.11	0.11	0.33	0.11	0.11
Italy	0.05	0.00	0.00	0.03	0.00	0.01
Japan	0.09	0.01	0.00	0.10	0.01	0.01
Jordan	0.18	0.03	0.03	0.24	0.06	0.05
Kazakhstan	-0.01	0.00	0.01	0.01	0.00	0.01
Korea, Rep. of	0.00	0.00	0.00	0.06	0.00	0.01
Kuwait	0.22	0.05	0.04	0.03	0.00	0.00
Lebanon	0.08	0.01	0.01	0.15	0.02	0.01
Lithuania	0.21	0.04	0.03	0.13	0.02	0.02
Malaysia	0.39	0.15	0.12	0.39	0.15	0.11
Morocco	0.17	0.03	0.04	0.18	0.03	0.03
New Zealand	0.41	0.17	0.13	0.39	0.15	0.09
Norway (9)	0.17	0.03	0.02	0.15	0.02	0.02
Oman	0.15	0.02	0.02	0.08	0.01	0.01
Portugal	0.25	0.06	0.02	0.20	0.04	0.04
Qatar	0.27	0.07	0.04	0.17	0.03	0.02
Romania	0.29	0.08	0.08	0.23	0.05	0.03
Russian Federation	0.16	0.03	0.03	0.13	0.02	0.01
Saudi Arabia	0.10	0.01	0.01	0.15	0.02	0.02
Singapore	0.42	0.17	0.14	0.32	0.10	0.07
South Africa (9)	0.21	0.04	0.03	0.16	0.03	0.02
Sweden	0.21	0.05	0.03	0.18	0.03	0.02
Turkey	0.29	0.08	0.08	0.24	0.06	0.04
United Arab Emirates	0.37	0.14	0.13	0.40	0.16	0.13
United States	0.41	0.17	0.13	0.26	0.07	0.06
International Median	**0.21**	**0.05**	**0.03**	**0.18**	**0.03**	**0.03**

Relationship Between the TIMSS 2019 *Classroom Teaching Limited by Students Not Ready for Instruction* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Benchmarking Participants						
Ontario, Canada	0.26	0.07	0.08	0.24	0.06	0.05
Quebec, Canada	0.33	0.11	0.09	0.35	0.13	0.12
Moscow City, Russian Fed.	0.17	0.03	0.04	0.16	0.03	0.02
Gauteng, RSA (9)	0.23	0.05	0.09	0.17	0.03	0.03
Western Cape, RSA (9)	0.49	0.24	0.16	0.42	0.18	0.10
Abu Dhabi, UAE	0.39	0.15	0.13	0.43	0.18	0.15
Dubai, UAE	0.34	0.11	0.11	0.35	0.12	0.10

Safe and Orderly School – Grade 8

About the Scale

The *Safe and Orderly School* scale was created based on teachers' responses to eight items listed below.

Thinking about your current school, indicate the extent to which you agree or disagree with each of the following statements.

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
BTBG07A [T]	1) This school is located in a safe neighborhood	○	○	○	○
BTBG07B [T]	2) I feel safe at this school	○	○	○	○
BTBG07C [T]	3) This school's security policies and practices are sufficient	○	○	○	○
BTBG07D [T]	4) The students behave in an orderly manner	○	○	○	○
BTBG07E [T]	5) The students are respectful of the teachers	○	○	○	○
BTBG07F [T]	6) The students respect school property	○	○	○	○
BTBG07G [T]	7) This school has clear rules about student conduct	○	○	○	○
BTBG07H [T]	8) This school's rules are enforced in a fair and consistent manner	○	○	○	○

Very Safe and Orderly	Somewhat Safe and Orderly	Less than Safe and Orderly

Scale Cut Scores **10.5** **7.3**

[1] For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Disagree a little" and "Disagree a lot" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Safe and Orderly School* Scale – Grade 8

Item	delta	tau_1	tau_2	Infit
BTBG07A	-0.93180	-1.21284	1.21284	1.26
BTBG07B	-1.50733	-1.31584	1.31584	0.98
BTBG07C	-0.70617	-1.53302	1.53302	1.04
BTBG07D	1.04692	-1.96681	1.96681	0.91
BTBG07E	0.77423	-1.93544	1.93544	0.91
BTBG07F	1.59701	-1.88167	1.88167	0.94
BTBG07G	-0.48472	-1.53624	1.53624	1.05
BTBG07H	0.21186	-1.57604	1.57604	1.03

Scale Transformation Constants for the TIMSS 2019 *Safe and Orderly School* Scale – Grade 8

Scale Transformation Constants	
A = 8.929660	Transformed Scale Score = 8.929660 + 0.950459 • Logit Scale Score
B = 0.950459	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Safe and Orderly School* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	4.42632	
1	5.63149	
2	6.28831	
3	6.79059	
4	7.22558	7.3
5	7.62717	
6	8.01327	
7	8.39631	
8	8.78337	
9	9.18523	
10	9.61153	
11	10.07263	
12	10.58171	10.5
13	11.14932	
14	11.78998	
15	12.56459	
16	13.87035	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Safe and Orderly School* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			BTBG07A	BTBG07B	BTBG07C	BTBG07D	BTBG07E	BTBG07F	BTBG07G	BTBG07H
Australia	0.89	58	0.59	0.71	0.74	0.85	0.85	0.85	0.70	0.74
Bahrain	0.87	53	0.41	0.53	0.68	0.85	0.83	0.82	0.77	0.79
Chile	0.90	59	0.66	0.72	0.75	0.80	0.83	0.83	0.73	0.78
Chinese Taipei	0.88	54	0.69	0.73	0.72	0.76	0.78	0.75	0.70	0.72
Cyprus	0.91	61	0.63	0.81	0.79	0.83	0.84	0.81	0.73	0.78
Egypt	0.86	50	0.64	0.63	0.71	0.74	0.73	0.77	0.68	0.75
England	0.88	56	0.42	0.69	0.75	0.85	0.84	0.84	0.69	0.80
Finland	0.85	50	0.54	0.69	0.65	0.79	0.80	0.74	0.66	0.77
France	0.89	57	0.75	0.76	0.73	0.82	0.82	0.76	0.70	0.70
Georgia	0.78	40	0.35	0.58	0.54	0.73	0.74	0.71	0.65	0.68
Hong Kong SAR	0.85	49	0.51	0.62	0.71	0.76	0.77	0.74	0.73	0.73
Hungary	0.86	50	0.57	0.64	0.74	0.78	0.79	0.75	0.63	0.74
Iran, Islamic Rep. of	0.88	54	0.69	0.68	0.71	0.79	0.72	0.76	0.78	0.74
Ireland	0.88	56	0.56	0.70	0.68	0.85	0.85	0.79	0.75	0.78
Israel	0.87	52	0.48	0.62	0.62	0.82	0.82	0.80	0.77	0.78
Italy	0.83	46	0.69	0.73	0.65	0.69	0.63	0.69	0.64	0.72
Japan	0.83	46	0.66	0.76	0.61	0.72	0.67	0.68	0.62	0.72
Jordan	0.90	58	0.67	0.76	0.77	0.79	0.77	0.82	0.74	0.75
Kazakhstan	0.84	48	0.54	0.64	0.67	0.71	0.75	0.69	0.75	0.77
Korea, Rep. of	0.90	58	0.72	0.80	0.82	0.78	0.72	0.73	0.75	0.75
Kuwait	0.84	48	0.42	0.60	0.65	0.78	0.73	0.81	0.72	0.73
Lebanon	0.85	50	0.50	0.52	0.61	0.81	0.79	0.81	0.77	0.74
Lithuania	0.87	52	0.56	0.76	0.78	0.76	0.75	0.72	0.67	0.73
Malaysia	0.88	55	0.73	0.80	0.76	0.78	0.77	0.70	0.69	0.69
Morocco	0.89	58	0.68	0.77	0.80	0.79	0.77	0.77	0.73	0.75
New Zealand	0.88	54	0.56	0.67	0.66	0.84	0.82	0.82	0.71	0.73
Norway (9)	0.83	46	0.45	0.64	0.63	0.78	0.76	0.75	0.66	0.70
Oman	0.80	42	0.54	0.58	0.60	0.77	0.74	0.66	0.64	0.65
Portugal	0.87	53	0.61	0.71	0.74	0.77	0.79	0.76	0.72	0.71
Qatar	0.88	54	0.46	0.69	0.67	0.80	0.79	0.76	0.83	0.82
Romania	0.86	51	0.48	0.62	0.69	0.79	0.82	0.81	0.66	0.76
Russian Federation	0.87	52	0.60	0.68	0.69	0.80	0.81	0.81	0.59	0.76
Saudi Arabia	0.86	51	0.45	0.66	0.71	0.76	0.73	0.77	0.79	0.79
Singapore	0.88	56	0.59	0.65	0.70	0.82	0.83	0.82	0.72	0.79
South Africa (9)	0.87	53	0.67	0.76	0.77	0.78	0.79	0.78	0.59	0.65
Sweden	0.82	45	0.59	0.64	0.70	0.77	0.75	0.65	0.56	0.69
Turkey	0.90	58	0.73	0.75	0.78	0.80	0.74	0.78	0.76	0.74
United Arab Emirates	0.89	56	0.50	0.61	0.65	0.84	0.84	0.84	0.81	0.84
United States	0.87	54	0.50	0.64	0.65	0.86	0.84	0.83	0.70	0.78

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Safe and Orderly School* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			BTBG07A	BTBG07B	BTBG07C	BTBG07D	BTBG07E	BTBG07F	BTBG07G	BTBG07H
Benchmarking Participants										
Ontario, Canada	0.89	58	0.61	0.69	0.69	0.87	0.84	0.86	0.75	0.74
Quebec, Canada	0.87	54	0.56	0.72	0.80	0.80	0.73	0.77	0.75	0.71
Moscow City, Russian Fed.	0.87	52	0.58	0.64	0.68	0.81	0.79	0.79	0.68	0.79
Gauteng, RSA (9)	0.88	55	0.64	0.77	0.81	0.81	0.78	0.83	0.55	0.71
Western Cape, RSA (9)	0.89	57	0.75	0.83	0.74	0.81	0.76	0.80	0.63	0.68
Abu Dhabi, UAE	0.90	59	0.52	0.59	0.67	0.86	0.87	0.88	0.80	0.85
Dubai, UAE	0.87	54	0.49	0.60	0.60	0.84	0.85	0.79	0.79	0.83

Relationship Between the TIMSS 2019 *Safe and Orderly School* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Australia	0.26	0.07	0.05	0.23	0.05	0.03
Bahrain	0.05	0.00	0.01	0.16	0.03	0.02
Chile	0.27	0.07	0.10	0.24	0.06	0.05
Chinese Taipei	0.07	0.00	0.00	0.08	0.01	0.01
Cyprus	0.19	0.04	0.03	0.12	0.02	0.01
Egypt	0.15	0.02	0.02	0.08	0.01	0.01
England	0.14	0.02	0.02	0.09	0.01	0.01
Finland	0.11	0.01	0.01	0.11	0.01	0.01
France	0.18	0.03	0.04	0.21	0.04	0.04
Georgia	0.05	0.00	0.00	0.05	0.00	0.00
Hong Kong SAR	0.17	0.03	0.02	0.08	0.01	0.01
Hungary	0.22	0.05	0.04	0.20	0.04	0.03
Iran, Islamic Rep. of	0.13	0.02	0.01	0.13	0.02	0.02
Ireland	0.26	0.07	0.06	0.22	0.05	0.04
Israel	0.21	0.04	0.06	0.09	0.01	0.02
Italy	0.10	0.01	0.02	0.07	0.01	0.01
Japan	0.08	0.01	0.01	0.04	0.00	0.00
Jordan	0.18	0.03	0.03	0.22	0.05	0.02
Kazakhstan	-0.01	0.00	0.00	0.05	0.00	0.01
Korea, Rep. of	0.04	0.00	0.00	0.05	0.00	0.00
Kuwait	0.16	0.03	0.00	0.15	0.02	0.00
Lebanon	0.24	0.06	0.06	0.21	0.04	0.05
Lithuania	0.11	0.01	0.00	0.06	0.00	0.00
Malaysia	0.15	0.02	0.01	0.22	0.05	0.03
Morocco	0.12	0.01	0.01	0.06	0.00	0.00
New Zealand	0.27	0.07	0.06	0.20	0.04	0.03
Norway (9)	0.12	0.01	0.01	0.12	0.01	0.01
Oman	0.08	0.01	0.00	0.14	0.02	0.02
Portugal	0.16	0.02	0.02	0.12	0.01	0.01
Qatar	0.14	0.02	0.01	0.06	0.00	0.00
Romania	0.10	0.01	0.01	0.15	0.02	0.01
Russian Federation	0.01	0.00	0.00	0.08	0.01	0.01
Saudi Arabia	0.09	0.01	0.01	0.17	0.03	0.02
Singapore	0.12	0.01	0.00	0.14	0.02	0.01
South Africa (9)	0.21	0.05	0.04	0.13	0.02	0.02
Sweden	0.19	0.04	0.03	0.18	0.03	0.03
Turkey	0.11	0.01	0.01	0.18	0.03	0.04
United Arab Emirates	0.31	0.09	0.07	0.33	0.11	0.08
United States	0.20	0.04	0.05	0.25	0.06	0.06
International Median	**0.14**	**0.02**	**0.01**	**0.13**	**0.02**	**0.01**

Relationship Between the TIMSS 2019 *Safe and Orderly School* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		*r*	r^2	
Benchmarking Participants						
Ontario, Canada	0.20	0.04	0.04	0.19	0.03	0.04
Quebec, Canada	0.18	0.03	0.04	-0.06	0.00	0.02
Moscow City, Russian Fed.	0.13	0.02	0.02	0.09	0.01	0.01
Gauteng, RSA (9)	0.38	0.15	0.12	0.33	0.11	0.08
Western Cape, RSA (9)	0.51	0.26	0.21	0.57	0.33	0.32
Abu Dhabi, UAE	0.43	0.19	0.16	0.39	0.15	0.12
Dubai, UAE	0.19	0.04	0.03	0.20	0.04	0.01

Teachers' Emphasis on Science Investigation – Grade 8

About the Scale

The *Teachers' Emphasis on Science Investigation* scale was created based on teachers' responses to eight items listed below.

In teaching science to the students in this class, how often do you ask them to do the following?

		Every or almost every lesson	About half the lessons	Some lessons	Never
BTBS15B T	1) Observe natural phenomena and describe what they see	○	○	○	○
BTBS15C T	2) Watch me demonstrate an experiment or investigation	○	○	○	○
BTBS15D T	3) Design or plan experiments or investigations	○	○	○	○
BTBS15E T	4) Conduct experiments or investigations	○	○	○	○
BTBS15F T	5) Present data from experiments or investigations	○	○	○	○
BTBS15G T	6) Interpret data from experiments or investigations	○	○	○	○
BTBS15H T	7) Use evidence from experiments or investigations to support conclusions	○	○	○	○
BTBS15L T	8) Do field work outside the class	○	○	○	○

About Half the Lessons or More | **Less than Half the Lessons**

Scale Cut Scores **11.3**

[1] For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Some lessons" and "Never" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

[T] Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Teachers' Emphasis on Science Investigation* Scale – Grade 8

Item	delta	tau_1	tau_2	Infit
BTBS15B	-0.94861	-0.86697	0.86697	1.36
BTBS15C	-0.56557	-0.45648	0.45648	1.25
BTBS15D	0.34598	-0.71308	0.71308	0.89
BTBS15E	0.02202	-0.89215	0.89215	0.88
BTBS15F	0.14912	-0.77092	0.77092	0.75
BTBS15G	-0.20789	-0.79691	0.79691	0.76
BTBS15H	-0.49705	-0.80328	0.80328	0.92
BTBS15L	1.70200	-0.19205	0.19205	1.18

Scale Transformation Constants for the TIMSS 2019 *Teachers' Emphasis on Science Investigation* Scale – Grade 8

Scale Transformation Constants	
A = 11.333349	Transformed Scale Score = 11.333349 + 1.2708400 • Logit Scale Score
B = 1.2708400	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Teachers' Emphasis on Science Investigation* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	6.51309	
1	8.03870	
2	8.81776	
3	9.37904	
4	9.83696	
5	10.24200	
6	10.61377	
7	10.96886	
8	11.31862	11.3
9	11.67481	
10	12.04435	
11	12.43721	
12	12.86171	
13	13.33488	
14	13.88949	
15	14.62251	
16	16.03256	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Teachers' Emphasis on Science Investigation* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			BTBS15B	BTBS15C	BTBS15D	BTBS15E	BTBS15F	BTBS15G	BTBS15H	BTBS15L
Australia	0.83	47	0.49	0.65	0.63	0.79	0.80	0.84	0.75	0.40
Bahrain	0.88	56	0.50	0.62	0.81	0.89	0.87	0.87	0.84	0.48
Chile	0.88	55	0.49	0.76	0.85	0.87	0.82	0.82	0.75	0.49
Chinese Taipei	0.83	47	0.51	0.58	0.71	0.79	0.82	0.84	0.76	0.26
Cyprus	0.86	56	0.54	0.59	0.81	0.88	0.87	0.90	0.86	0.27
Egypt	0.87	53	0.61	0.42	0.77	0.75	0.84	0.88	0.84	0.60
England	0.76	39	0.43	0.58	0.72	0.69	0.73	0.73	0.70	0.18
Finland	0.85	50	0.54	0.59	0.58	0.88	0.87	0.88	0.86	0.05
France	0.77	40	0.44	0.42	0.72	0.68	0.79	0.77	0.69	0.36
Georgia	0.85	54	0.43	0.60	0.80	0.88	0.90	0.88	0.75	0.47
Hong Kong SAR	0.86	52	0.54	0.69	0.68	0.73	0.84	0.85	0.81	0.52
Hungary	0.83	50	0.60	0.65	0.75	0.76	0.82	0.79	0.67	0.54
Iran, Islamic Rep. of	0.85	51	0.50	0.40	0.77	0.80	0.87	0.86	0.75	0.60
Ireland	0.81	46	0.33	0.29	0.77	0.81	0.88	0.87	0.84	0.14
Israel	0.88	57	0.68	0.58	0.87	0.83	0.88	0.83	0.75	0.54
Italy	0.83	51	0.44	0.55	0.85	0.83	0.86	0.85	0.76	0.33
Japan	0.81	44	0.74	0.66	0.30	0.82	0.84	0.80	0.63	0.07
Jordan	0.83	47	0.54	0.52	0.77	0.79	0.81	0.72	0.70	0.56
Kazakhstan	0.86	52	0.49	0.61	0.81	0.83	0.84	0.84	0.72	0.52
Korea, Rep. of	0.87	54	0.59	0.71	0.78	0.84	0.84	0.81	0.72	0.49
Kuwait	0.84	50	0.28	0.44	0.81	0.87	0.87	0.90	0.80	0.33
Lebanon	0.81	44	0.61	0.56	0.73	0.77	0.76	0.75	0.60	0.45
Lithuania	0.87	57	0.59	0.63	0.84	0.90	0.89	0.80	0.76	0.53
Malaysia	0.89	58	0.67	0.67	0.77	0.84	0.86	0.90	0.76	0.57
Morocco	0.78	40	0.47	0.40	0.71	0.73	0.79	0.77	0.69	0.37
New Zealand	0.80	44	0.47	0.61	0.68	0.68	0.79	0.82	0.80	0.22
Norway (9)	0.76	39	0.54	0.46	0.72	0.73	0.74	0.73	0.63	0.32
Oman	0.82	45	0.48	0.52	0.72	0.85	0.83	0.83	0.64	0.33
Portugal	0.84	51	0.58	0.66	0.81	0.78	0.82	0.79	0.73	0.50
Qatar	0.86	51	0.50	0.67	0.75	0.78	0.85	0.85	0.73	0.51
Romania	0.84	53	0.42	0.63	0.82	0.87	0.88	0.86	0.80	0.27
Russian Federation	0.90	61	0.61	0.66	0.85	0.89	0.90	0.87	0.76	0.63
Saudi Arabia	0.86	53	0.48	0.47	0.81	0.83	0.89	0.89	0.84	0.44
Singapore	0.78	43	0.47	0.57	0.58	0.72	0.83	0.79	0.70	0.45
South Africa (9)	0.90	59	0.61	0.71	0.80	0.85	0.87	0.85	0.83	0.53
Sweden	0.85	51	0.44	0.53	0.78	0.85	0.89	0.85	0.78	0.41
Turkey	0.91	61	0.62	0.64	0.82	0.89	0.88	0.87	0.80	0.67
United Arab Emirates	0.89	57	0.58	0.64	0.84	0.84	0.87	0.86	0.76	0.54
United States	0.84	48	0.57	0.46	0.71	0.82	0.86	0.85	0.76	0.33

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Teachers' Emphasis on Science Investigation* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			BTBS15B	BTBS15C	BTBS15D	BTBS15E	BTBS15F	BTBS15G	BTBS15H	BTBS15L
Benchmarking Participants										
Ontario, Canada	0.90	61	0.46	0.73	0.88	0.88	0.90	0.91	0.75	0.64
Quebec, Canada	0.62	-	-	-	-	-	-	-	-	-
Moscow City, Russian Fed.	0.87	56	0.56	0.58	0.83	0.88	0.88	0.88	0.77	0.48
Gauteng, RSA (9)	0.91	62	0.70	0.75	0.80	0.87	0.86	0.84	0.85	0.58
Western Cape, RSA (9)	0.90	60	0.63	0.67	0.81	0.87	0.88	0.88	0.87	0.49
Abu Dhabi, UAE	0.90	61	0.60	0.70	0.84	0.86	0.89	0.90	0.80	0.54
Dubai, UAE	0.87	54	0.56	0.59	0.85	0.76	0.86	0.85	0.79	0.52

A dash (–) indicates comparable data not available.

Relationship Between the TIMSS 2019 *Teachers' Emphasis on Science Investigation* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Australia	0.06	0.00	0.00
Bahrain	-0.01	0.00	0.00
Chile	-0.06	0.00	0.00
Chinese Taipei	0.01	0.00	0.00
Cyprus	-0.10	0.01	0.00
Egypt	0.12	0.01	0.01
England	-0.06	0.00	0.01
Finland	0.04	0.00	0.00
France	0.07	0.00	0.00
Georgia	-0.01	0.00	0.00
Hong Kong SAR	0.12	0.01	0.01
Hungary	0.04	0.00	0.00
Iran, Islamic Rep. of	0.04	0.00	0.00
Ireland	0.02	0.00	0.00
Israel	0.00	0.00	0.00
Italy	-0.06	0.00	0.00
Japan	0.02	0.00	0.00
Jordan	0.04	0.00	0.00
Kazakhstan	0.00	0.00	0.00
Korea, Rep. of	0.03	0.00	0.00
Kuwait	-0.05	0.00	0.01
Lebanon	-0.03	0.00	0.00
Lithuania	-0.01	0.00	0.00
Malaysia	0.09	0.01	0.01
Morocco	0.02	0.00	0.00
New Zealand	0.05	0.00	0.00
Norway (9)	0.04	0.00	0.00
Oman	0.03	0.00	0.00
Portugal	0.01	0.00	0.00
Qatar	0.01	0.00	0.00
Romania	0.03	0.00	0.00
Russian Federation	0.05	0.00	0.00
Saudi Arabia	0.12	0.01	0.00
Singapore	0.08	0.01	0.00
South Africa (9)	-0.03	0.00	0.00
Sweden	-0.01	0.00	0.00
Turkey	0.05	0.00	0.00
United Arab Emirates	0.11	0.01	0.01
United States	0.13	0.02	0.00
International Median	**0.03**	**0.00**	**0.00**

Relationship Between the TIMSS 2019 *Teachers' Emphasis on Science Investigation* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2	
Benchmarking Participants			
Ontario, Canada	0.01	0.00	0.00
Quebec, Canada	-	-	-
Moscow City, Russian Fed.	0.06	0.00	0.00
Gauteng, RSA (9)	-0.08	0.01	0.00
Western Cape, RSA (9)	-0.01	0.00	0.01
Abu Dhabi, UAE	0.24	0.06	0.04
Dubai, UAE	-0.08	0.01	0.00

A dash (–) indicates comparable data not available.

Teachers’ Job Satisfaction – Grade 8

About the Scale

The *Teachers’ Job Satisfaction* scale was created based on students’ responses to five items listed below.[1]

How often do you feel the following way about being a teacher?

		Very often	Often	Sometimes	Never or almost never
BTBG08A T	1) I am content with my profession as a teacher	◯	◯	◯	◯
BTBG08B T	2) I find my work full of meaning and purpose	◯	◯	◯	◯
BTBG08C T	3) I am enthusiastic about my job	◯	◯	◯	◯
BTBG08D T	4) My work inspires me	◯	◯	◯	◯
BTBG08E T	5) I am proud of the work I do	◯	◯	◯	◯

Very Satisfied | Somewhat Satisfied | Less than Satisfied

Scale Cut Scores **10.2** **6.8**

[1] For the purpose of scaling, categories in which there were very few respondents were combined. The categories “Sometimes” and “Never or almost never” were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

T Trend item—item was included in the same scale in TIMSS 2015 and was used for linking the TIMSS 2015 and TIMSS 2019 scales.

Item Parameters for the TIMSS 2019 *Teachers' Job Satisfaction* Scale – Grade 8

Item	delta	tau_1	tau_2	Infit
BTBG08A	0.21265	-2.26347	2.26347	1.09
BTBG08B	-0.61570	-2.17906	2.17906	1.04
BTBG08C	0.10008	-2.26113	2.26113	0.93
BTBG08D	0.65284	-2.20301	2.20301	0.95
BTBG08E	-0.34987	-1.96126	1.96126	1.13

Scale Transformation Constants for the TIMSS 2019 *Teachers' Job Satisfaction* Scale – Grade 8

Scale Transformation Constants	
A = 8.497121	Transformed Scale Score = 8.497121 + 0.691344 • Logit Scale Score
B = 0.691344	

Equivalence Table of the Raw Score and Transformed Scale Scores for the TIMSS 2019 *Teachers' Job Satisfaction* Scale – Grade 8

Raw Score	Transformed Scale Score	Cutpoint
0	5.28118	
1	6.19622	
2	6.74486	6.8
3	7.21969	
4	7.74231	
5	8.48286	
6	9.22836	
7	9.76930	
8	10.25874	10.2
9	10.82007	
10	11.74559	

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Teachers' Job Satisfaction* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item				
			BTBG08A	BTBG08B	BTBG08C	BTBG08D	BTBG08E
Australia	0.91	74	0.86	0.89	0.91	0.90	0.75
Bahrain	0.89	69	0.82	0.78	0.89	0.86	0.82
Chile	0.88	68	0.84	0.80	0.80	0.89	0.79
Chinese Taipei	0.93	78	0.89	0.91	0.90	0.84	0.87
Cyprus	0.93	78	0.87	0.92	0.89	0.89	0.83
Egypt	0.83	60	0.80	0.73	0.82	0.74	0.79
England	0.89	69	0.84	0.87	0.88	0.84	0.71
Finland	0.92	76	0.88	0.84	0.91	0.87	0.86
France	0.89	70	0.82	0.77	0.89	0.86	0.82
Georgia	0.78	54	0.71	0.71	0.69	0.75	0.80
Hong Kong SAR	0.91	75	0.87	0.87	0.87	0.85	0.87
Hungary	0.91	73	0.85	0.84	0.86	0.87	0.85
Iran, Islamic Rep. of	0.88	68	0.82	0.81	0.86	0.84	0.77
Ireland	0.91	74	0.84	0.88	0.90	0.86	0.81
Israel	0.90	72	0.83	0.84	0.87	0.87	0.83
Italy	0.88	68	0.81	0.78	0.87	0.84	0.83
Japan	0.90	72	0.85	0.86	0.83	0.88	0.82
Jordan	0.91	73	0.80	0.86	0.89	0.88	0.85
Kazakhstan	0.86	65	0.75	0.81	0.80	0.84	0.82
Korea, Rep. of	0.94	80	0.88	0.89	0.88	0.92	0.91
Kuwait	0.80	57	0.73	0.69	0.84	0.81	0.70
Lebanon	0.81	56	0.72	0.68	0.83	0.80	0.71
Lithuania	0.90	72	0.85	0.74	0.89	0.89	0.86
Malaysia	0.94	80	0.90	0.90	0.90	0.90	0.89
Morocco	0.89	69	0.83	0.80	0.88	0.85	0.80
New Zealand	0.91	73	0.83	0.87	0.88	0.88	0.81
Norway (9)	0.89	69	0.78	0.82	0.82	0.88	0.84
Oman	0.90	72	0.87	0.79	0.88	0.87	0.84
Portugal	0.90	72	0.86	0.85	0.88	0.89	0.74
Qatar	0.89	69	0.85	0.75	0.87	0.87	0.82
Romania	0.91	75	0.84	0.87	0.88	0.88	0.85
Russian Federation	0.89	69	0.83	0.74	0.86	0.88	0.85
Saudi Arabia	0.84	65	0.81	0.78	0.85	0.86	0.71
Singapore	0.95	83	0.89	0.90	0.93	0.92	0.91
South Africa (9)	0.89	69	0.78	0.85	0.85	0.86	0.82
Sweden	0.88	68	0.82	0.77	0.87	0.88	0.79
Turkey	0.92	77	0.85	0.85	0.92	0.91	0.86
United Arab Emirates	0.90	72	0.78	0.85	0.90	0.88	0.83
United States	0.93	78	0.85	0.92	0.92	0.92	0.83

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the TIMSS 2019 *Teachers' Job Satisfaction* Scale – Grade 8

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item				
			BTBG08A	BTBG08B	BTBG08C	BTBG08D	BTBG08E
Benchmarking Participants							
Ontario, Canada	0.92	75	0.83	0.86	0.90	0.88	0.87
Quebec, Canada	0.89	70	0.82	0.79	0.86	0.86	0.85
Moscow City, Russian Fed.	0.89	70	0.83	0.79	0.88	0.88	0.81
Gauteng, RSA (9)	0.90	72	0.79	0.87	0.87	0.87	0.83
Western Cape, RSA (9)	0.88	69	0.83	0.88	0.86	0.83	0.74
Abu Dhabi, UAE	0.91	74	0.80	0.90	0.90	0.87	0.84
Dubai, UAE	0.91	74	0.79	0.85	0.91	0.90	0.84

Relationship Between the TIMSS 2019 *Teachers' Job Satisfaction* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Australia	0.12	0.01	0.02	0.07	0.00	0.00
Bahrain	0.03	0.00	0.00	0.10	0.01	0.01
Chile	0.01	0.00	0.01	0.04	0.00	0.01
Chinese Taipei	0.04	0.00	0.00	0.01	0.00	0.00
Cyprus	0.04	0.00	0.01	0.04	0.00	0.00
Egypt	0.17	0.03	0.03	0.05	0.00	0.01
England	0.19	0.04	0.05	0.11	0.01	0.01
Finland	0.03	0.00	0.00	0.03	0.00	0.00
France	0.01	0.00	0.00	0.06	0.00	0.00
Georgia	-0.06	0.00	0.00	0.03	0.00	0.00
Hong Kong SAR	0.08	0.01	0.00	0.05	0.00	0.01
Hungary	0.14	0.02	0.02	0.07	0.01	0.01
Iran, Islamic Rep. of	0.04	0.00	0.00	0.02	0.00	0.01
Ireland	0.13	0.02	0.01	0.10	0.01	0.01
Israel	0.04	0.00	0.01	0.05	0.00	0.00
Italy	0.01	0.00	0.00	-0.02	0.00	0.00
Japan	0.08	0.01	0.00	0.02	0.00	0.00
Jordan	0.04	0.00	0.00	0.22	0.05	0.04
Kazakhstan	-0.12	0.01	0.02	-0.07	0.00	0.00
Korea, Rep. of	-0.02	0.00	0.00	0.06	0.00	0.00
Kuwait	-0.04	0.00	0.00	0.08	0.01	0.01
Lebanon	0.11	0.01	0.01	0.09	0.01	0.01
Lithuania	0.06	0.00	0.00	0.07	0.00	0.01
Malaysia	-0.03	0.00	0.00	0.10	0.01	0.02
Morocco	0.07	0.00	0.01	-0.01	0.00	0.00
New Zealand	0.00	0.00	0.00	0.03	0.00	0.00
Norway (9)	0.04	0.00	0.00	0.00	0.00	0.00
Oman	0.08	0.01	0.00	0.11	0.01	0.01
Portugal	0.08	0.01	0.00	0.04	0.00	0.00
Qatar	0.08	0.01	0.01	-0.04	0.00	0.00
Romania	0.06	0.00	0.01	0.10	0.01	0.01
Russian Federation	0.05	0.00	0.00	0.06	0.00	0.00
Saudi Arabia	-0.03	0.00	0.00	0.10	0.01	0.02
Singapore	0.12	0.01	0.01	0.06	0.00	0.00
South Africa (9)	0.00	0.00	0.00	-0.02	0.00	0.00
Sweden	0.03	0.00	0.00	0.04	0.00	0.00
Turkey	0.03	0.00	0.01	0.11	0.01	0.01
United Arab Emirates	0.18	0.03	0.03	0.13	0.02	0.01
United States	0.05	0.00	0.00	0.11	0.01	0.01
International Median	**0.04**	**0.00**	**0.00**	**0.06**	**0.00**	**0.00**

Relationship Between the TIMSS 2019 *Teachers' Job Satisfaction* Scale and TIMSS 2019 Achievement – Grade 8

Country	Pearson's Correlation with Mathematics Achievement		Variance in Mathematics Achievement Accounted for by Difference Between Regions of the Scale (η^2)	Pearson's Correlation with Science Achievement		Variance in Science Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	r	r^2		r	r^2	
Benchmarking Participants						
Ontario, Canada	0.01	0.00	0.00	0.04	0.00	0.00
Quebec, Canada	0.08	0.01	0.01	-0.05	0.00	0.02
Moscow City, Russian Fed.	0.14	0.02	0.02	0.02	0.00	0.00
Gauteng, RSA (9)	0.02	0.00	0.00	-0.08	0.01	0.01
Western Cape, RSA (9)	0.01	0.00	0.00	0.21	0.04	0.03
Abu Dhabi, UAE	0.25	0.06	0.04	0.21	0.05	0.03
Dubai, UAE	0.08	0.01	0.01	0.05	0.00	0.01

BOSTON COLLEGE

timss.bc.edu

ISBN: 978-1-889938-53-0
LOC: 2020901029

IEA

TIMSS & PIRLS
International Study Center
Lynch School of Education
BOSTON COLLEGE