TIMSS

Technical Report
Volume III

TIMSS

Third International Mathematics and Science Study

Technical Report

Volume III: Implementation and Analysis

Final Year of Secondary School (Population 3)

Edited by

Michael O. Martin
Dana L. Kelly

with contributors

Raymond J. Adams
Michael Bruneforth
Jean Dumais
Pierre Foy
Eugenio Gonzalez
Dirk Hastedt
Greg Macaskill
Ina V.S. Mullis
Knut Schwippert
Heiko Sibberns
Margaret L. Wu

Boston College • Chestnut Hill, Massachusetts, USA

Third International Mathematics and Science Study Technical Report, Volume III: Implementation and Analysis – Final Year of Secondary School /edited by Michael O. Martin, Dana L. Kelly
Publisher: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College

Library of Congress Catalog Card Number: 99-60717

ISBN: 1-889938-12-2

For more information about TIMSS contact:

TIMSS International Study Center
Center for the Study of Testing, Evaluation, and Educational Policy
Campion Hall
School of Education
Boston College
Chestnut Hill, MA 02467
United States

This report is also available on the World Wide Web:
http://www.csteep.bc.edu/timss

Funding for the international coordination of TIMSS is provided by the U.S. National Center for Education Statistics, the U.S. National Science Foundation, the IEA, and the Canadian government. Each participating country provides funding for the national implementation of TIMSS.

Printed and bound in the United States.

Contents

Foreword

IEA's Third International Mathematics and Science Study (TIMSS) brought together educators, policymakers, and researchers from 41 countries to study student achievement in mathematics and science, and the factors influencing that achievement. TIMSS was an ambitious and demanding collaborative effort that required considerable resources and expertise, and the dedication of all involved. The TIMSS International Study Center at Boston College has been responsible for directing the course of the study and for orchestrating the contributions of the many participants. To date the results of the study have been summarized in six international reports published by the International Study Center.

A study like TIMSS faces many technical challenges, and is heavily dependent on the technology of educational measurement for its success. TIMSS has placed great emphasis on documenting the technical aspects of the project, and has produced a wide range of technical documentation. In addition to a three-volume series of technical reports (of which this volume is the third), TIMSS has produced two large user databases with accompanying user guides and supplementary documentation, so that secondary analysts can have complete access to the TIMSS data, and a technical volume detailing all of the quality control measures taken to assure the quality of the TIMSS data.

The first volume in this series, the *TIMSS Technical Report, Volume I: Design and Development*, describes the design and development of TIMSS, including the development of the achievement tests and questionnaires, the sample design and field operations procedures, and the plans for quality assurance activities. The second volume, *TIMSS Technical Report, Volume II*, documents the implementation and analysis of TIMSS for students in the primary and middle school years (Populations 1 and 2 in the terminology of TIMSS). The implementation of the sample design, the calculation of sampling weights, procedures for the estimation of sampling variability, steps involved in the international data verification, the TIMSS scaling model, and the analysis of the achievement and background data, are all presented in that volume for those two populations.

I am pleased to introduce the third and final volume in the series, *TIMSS Technical Report, Volume III*, which, since it documents the implementation of TIMSS at the final year of secondary school (Population 3 in TIMSS terms), is a parallel volume to Volume II. Together with the international reports presenting the study results, the international databases, and the earlier technical volumes that have already been published, this volume completes the first round of reports from the TIMSS International Study Center. The technical volumes should prove indispensable to those educators, analysts, and policymakers who seek deeper understanding of the techniques and methodology underpinning the TIMSS results.

Albert E. Beaton
TIMSS International
Study Director
Boston College

TIMSS was truly a collaborative effort among hundreds of individuals around the world. Staff from the national research centers of the participating countries, the international management, advisors, and funding agencies worked closely to design and implement the most ambitious study of international comparative achievement in mathematics and science ever undertaken. The design was implemented in each country by the TIMSS national research center staff, with the cooperation and assistance of schools, and the participation of the students and teachers. This volume documents the efforts of those involved in the implementation of the very ambitious TIMSS design, and the steps undertaken to analyze and report the international results for students in the final year of secondary school.

It is impossible to acknowledge individually everyone who contributed to the implementation and analysis of TIMSS. Chapter authors have recognized significant contributors where appropriate, and the Acknowledgments section at the end of the volume further acknowledges the National Research Coordinators and special advisors. The financial support provided by the National Center for Education Statistics of the U.S. Department of Education, the U.S. National Science Foundation, and the participating countries was essential in allowing us to complete the technical documentation of the study. We gratefully acknowledge their continuing support of our efforts.

This report would not have been possible without the efforts of many people. We are very grateful to the authors for their timely contributions, and for their cooperation throughout the editing process. We are especially grateful to Albert Beaton, the TIMSS International Study Director, for his constant help and support. His insistence on the central importance of technical documentation in a study like TIMSS was a continuous source of inspiration.

Several individuals at the TIMSS International Study Center at Boston College deserve special recognition for the production of this report. José R. Nieto coordinated the production of the report, including designing the layout and cover, scheduling production tasks, and assembling the text and tables. Rachel Saks was instrumental in seeing this report through to completion and Sarah Andrews diligently implemented many text changes throughout the revision process. Special thanks go to Maria Sachs for editing the text.

Michael O. Martin
Dana L. Kelly
Boston College

1 Introduction

Michael O. Martin
Dana L. Kelly
Boston College

TIMSS represents the continuation of a series of studies conducted by the International Association for the Evaluation of Educational Achievement (IEA). Since its inception in 1959, the IEA has conducted more than 15 studies of cross-national achievement in curricular areas such as mathematics, science, language, civics, and reading. IEA conducted its First International Mathematics Study (FIMS) in 1964, and the Second International Mathematics Study (SIMS) in 1980-82. The First and Second International Science Studies (FISS and SISS) were conducted in 1970-71 and 1983-84, respectively. Since the subjects of mathematics and science are related in many respects, the third studies were conducted together as an integrated effort.[1] The number of participating countries, the number of grades tested, and testing in both mathematics and science resulted in TIMSS becoming the largest, most complex IEA study to date and the largest international study of educational achievement ever undertaken.

Traditionally, IEA studies have systematically worked toward gaining a deeper insight into how various factors contribute to the overall outcomes of schooling. Particular emphasis has been placed on refining our understanding of students' opportunity to learn as that opportunity becomes defined and implemented by curricular and instructional practices. In an effort to extend what had been learned from previous studies and provide contextual and explanatory information, TIMSS was expanded beyond the already substantial task of measuring achievement in two subject areas to include a thorough investigation of curriculum and how it is delivered in classrooms around the world.

Continuing the approach of previous IEA studies, TIMSS defined three conceptual levels of curriculum. The intended curriculum is composed of the mathematics and science instructional and learning goals as defined at the system level. The implemented curriculum is the mathematics and science curriculum as interpreted by teachers and made available to students. The attained curriculum is the mathematics and science content that students have learned and their attitudes towards these subjects. To aid in interpretation and comparison of results, TIMSS also collected extensive information about the social and cultural contexts for learning, many of which are related to variations among education systems.

[1] In the time elapsed since SIMS and SISS, curriculum and testing methods have evolved considerably. The resulting changes in items and methods as well as differences in the populations tested make comparisons of TIMSS results with those of previous studies very difficult.

To gather information about the intended curriculum, mathematics and science specialists in each participating country worked section by section through curriculum guides, textbooks, and other curricular material to categorize them in accordance with detailed specifications drawn from the TIMSS mathematics and science curriculum frameworks (Robitaille et al., 1993). Initial results from this component of TIMSS can be found in two companion volumes: *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Mathematics* (Schmidt, McKnight, Valverde, Houang, and Wiley, 1997) and *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Science* (Schmidt, Raizen, Britton, Bianchi, and Wolfe, 1997).

To measure student achievement, TIMSS tested more than half a million students in mathematics and science at five grade levels involving the following three populations:

- Population 1. Students enrolled in the two adjacent grades that contained the largest proportion of 9-year-old students at the time of testing (third- and fourth-grade students in most countries).

- Population 2. Students enrolled in the two adjacent grades that contained the largest proportion of 13-year-old students at the time of testing (seventh- and eighth-grade students in most countries).

- Population 3. Students in their final year of secondary education. As an additional option, countries could test two subgroups of these students: students having taken advanced mathematics, and students having taken physics.

All countries that participated in TIMSS were to test students in Population 2. Many TIMSS countries also tested the mathematics and science achievement of students in Population 1 and of students in Population 3. Subsets of students in the fourth and eighth grades also had the opportunity to participate in a "hands-on" performance assessment. Together with the achievement tests, TIMSS administered a broad array of background questionnaires. The data collected from students, teachers, and school principals, as well as the system-level information collected from the participating countries, provide an abundance of information for further study and research. TIMSS data make it possible to examine differences in current levels of performance in relation to a wide range of variables associated with the classroom, school, and national contexts within which education takes place. The results of the assessments of Population 1 and Population 2 students have been published in:

> *Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study* (Mullis, Martin, Beaton, Gonzalez, Kelly, and Smith, 1997)

Science Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study (Martin, Mullis, Beaton, Gonzalez, Smith, and Kelly, 1997)

Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (Beaton, Mullis, Martin, Gonzalez, Kelly and Smith, 1996)

Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (Beaton, Martin, Mullis, Gonzalez, Smith, and Kelly, 1996)

Performance Assessment in IEA's Third International Mathematics and Science Study (Harmon, Smith, Martin, Kelly, Beaton, Mullis, Gonzalez, and Orpwood, 1997)

These reports have been widely disseminated and are available on the Internet (http://www.csteep.bc.edu/timss). The entire TIMSS international database containing the achievement and background data underlying these reports also has been released and is available at the TIMSS website.

The most recent TIMSS report, *Mathematics and Science Achievement in the Final Year of Secondary School: IEA's Third International Mathematics and Science Study* (Mullis, Martin, Beaton, Gonzalez, Kelly, and Smith, 1998), focuses on the mathematics and science literacy of all students in their final year of upper secondary school, and on the advanced mathematics and physics achievement of final-year students having taken courses in those subjects. This population, Population 3, was the most challenging to assess, largely because of the diversity of upper secondary systems and the complex sample design and test design required.

This technical report, the third in a series of technical reports documenting the TIMSS procedures and analyses, describes the implementation and analysis of the assessment of students in their final year of secondary school in 24 countries (see Figure 1.1). Previous volumes in the series documented the design and development of the study (Martin and Kelly, 1996) and the implementation and analysis of the assessment of students in Populations 1 and 2 (Martin and Kelly, 1997).

1.1 PARTICIPATING COUNTRIES AND STUDENTS

Figure 1.1 shows the countries that participated in the assessment of students in their final year of secondary school in mathematics and science literacy, advanced mathematics, and physics. Each participating country designated a national center to conduct the activities of the study and a National Research Coordinator (NRC) to assume responsibility for the successful completion of these tasks.[2] For the sake of comparability, all testing was conducted at the end of the school year. Most countries tested the mathematics and science achievement of their students at the end of the 1994-95 school year, most often in May and June of 1995. The three countries on a Southern Hemi-

[2] The Acknowledgments section lists the National Research Coordinators.

sphere school schedule (Australia, New Zealand, and South Africa) tested from August to December 1995, which was late in the school year in the Southern Hemisphere. Students in Australia were tested in September to October; students in New Zealand were tested in August; and students in South Africa were tested in August to December 1995. Three countries tested their final-year students (or a subset of them) at the end of the 1995-96 school year. Iceland tested its final-year students in 1996; Germany tested its gymnasium students in 1996; and Lithuania tested the students in vocational schools in 1996. In Germany and Lithuania, all other students included in the TIMSS assessment were tested in 1995.

Table 1.1 Countries Participating in Testing of Students in Their Final Year of Secondary School

Mathematics and Science Literacy	Advanced Mathematics	Physics
• Australia	• Australia	• Australia
• Austria	• Austria	• Austria
• Canada	• Canada	• Canada
• Cyprus	• Cyprus	• Cyprus
• Czech Republic	• Czech Republic	• Czech Republic
• Denmark	• Denmark	• Denmark
• France	• France	• France
• Germany	• Germany	• Germany
• Hungary	• Greece	• Greece
• Iceland	• Israel	• Israel
• Israel	• Italy	• Italy
• Italy	• Lithuania	• Latvia
• Lithuania	• Russian Federation	• Norway
• Netherlands	• Slovenia	• Russian Federation
• New Zealand	• Sweden	• Slovenia
• Norway	• Switzerland	• Sweden
• Russian Federation	• United States	• Switzerland
• Slovenia		• United States
• South Africa		
• Sweden		
• Switzerland		
• United States		

As can be imagined, testing students in their final year of secondary school was a special challenge for TIMSS. The 24 countries participating in this component of the testing vary greatly with respect to the nature of their upper secondary education systems. Some countries provide comprehensive education to students in their final years of school, while in other countries students might attend more specialized academic, vocational, or technical schools. Some countries fall between these extremes, their students being enrolled in academic, vocational, technical, or general programs of study within the same schools. Across countries the definitions of academic, vocational, and technical programs also vary, as do the kinds of education and training students in these programs receive.

The differences across countries in how education systems are organized, how students proceed through the upper secondary system, and when students leave school posed a challenge in defining the target populations to be tested in each country and interpreting the results. In order to make valid comparisons of students' performance across countries, it is critical that there be an understanding of which students were tested in each country, that is, how the target population was defined. It also is important to know how each upper secondary education system is structured and how the tested students fit into the system as a whole. In order to provide a context for interpreting the achievement results presented in this report, TIMSS summarized the structure of the upper secondary system for each country, specified the grades and tracks (programs of study) in which students were tested for TIMSS, and provided this information in the international report (Mullis et al., 1998).

Understandably, it was difficult for some countries to test all of the final-year students, particularly those in on-site occupational training. This, combined with the fact that by the final year of secondary school not all students are attending school, meant that countries differ with respect to the age-eligible cohort that was tested. To give some indication of the proportion of the entire school-leaving age cohort that was covered by the testing in each country, TIMSS developed its own index – the TIMSS Coverage Index or TCI.

1.2 THE TESTS FOR FINAL-YEAR STUDENTS

Three tests were developed for the TIMSS assessment of students in the final year of secondary school: the mathematics and science literacy test; the advanced mathematics test; and the physics test. The tests were developed through an international consensus involving input from experts in mathematics, science, and measurement. The TIMSS Subject Matter Advisory Committee, including distinguished scholars from 10 countries, ensured that the mathematics and science literacy tests represented current conceptions of literacy in those areas, and that the advanced mathematics and physics tests reflected current thinking and priorities in the fields of mathematics and physics education. The items underwent an iterative development and review process, with multiple pilot tests. Every effort was made to ensure that the items exhibited no bias towards or against particular countries. Item specifications were checked against data from the curriculum analysis. Items were rated for suitability by subject matter specialists in the participating countries, and a thorough statistical item analysis of data collected in the pilot testing was conducted. The final forms of the test were endorsed by the NRCs of the participating countries.[3]

[3] For a full discussion of the TIMSS test development effort, see Garden and Orpwood (1996), Robitaille and Garden (1996), and Orpwood and Garden (1998).

The mathematics and science literacy test was designed to test students' general knowledge and understanding of mathematical and scientific principles. The mathematics items cover number sense, including fractions, percentages, and proportionality. Algebraic sense, measurement, and estimation are also covered, as are data representation and analysis. Reasoning and social utility are emphasized in several items. A general criterion in selecting the items was that they should involve the types of mathematics questions that could arise in real-life situations and that they be contextualized accordingly. Similarly, the science items selected for use in the TIMSS literacy test were organized according to three areas of science – earth science, life science, and physical science – and included a reasoning and social utility component. The emphasis was on measuring how well students can use their knowledge in addressing real-world problems having a science component. The test was designed to enable reporting for mathematics literacy and science literacy separately as well as overall.

In order to examine how well students understand advanced mathematics concepts and can apply knowledge to solve problems, the advanced mathematics test was developed for students in their final year of secondary school having taken advanced mathematics. This test enabled reporting of achievement overall and in three content areas: numbers and equations; calculus; and geometry. In addition to items representing these three areas, the test also included items related to probability and statistics and to validation and structure, but because there were few such items, achievement in these areas was not reported separately.

The physics test was developed for students in their final year of secondary school who had taken physics, in order to examine how well they understand and can apply physics principles and concepts. It enabled reporting of physics achievement overall and in five content areas: mechanics; electricity and magnetism; heat; wave phenomena; and modern physics – particle physics, quantum physics and astrophysics, and relativity.

1.3 MANAGEMENT AND OPERATIONS

Like all previous IEA studies, TIMSS was essentially a cooperative venture among independent research centers around the world. While country representatives came together to plan the study and to agree on instruments and procedures, participants were each responsible for conducting TIMSS in their own country in accordance with the international standards. Each national center provided its own funding and contributed to the support of the international coordination of the study. A study of the scope and magnitude of TIMSS offers a tremendous operational and logistical challenge. In order to yield comparable data, the achievement survey must be replicated in each participating country in a timely and consistent manner. This was the responsibility of the NRC in each country. Among the major tasks of the NRCs in this regard were the following:

- Meeting with other NRCs and international project staff to plan the study and develop instruments and procedures

- Defining the school populations from which the TIMSS samples were to be drawn, selecting the sample of schools using an approved random sampling procedure, contacting the school principals and securing their agreement to participate in the study, and selecting the classes to be tested, again using an approved random sampling procedure

- Translating and adapting all of the tests, questionnaires, and administration manuals into the language of instruction of the country (and sometimes more than one language) prior to data collection

- Assembling, printing, and packaging the test booklets and questionnaires, and shipping the survey materials to the participating schools

- Ensuring that the tests and questionnaires were administered in participating schools, either by teachers in the school or by an external team of test administrators, and that the completed test protocols were returned to the TIMSS national center

- Conducting a quality assurance exercise in conjunction with the test administration, whereby some testing sessions were attended by an independent observer to confirm that all specified procedures were followed

- Recruiting and training individuals to score the free-response questions in the achievement tests, and implementing the plan for scoring the student responses, including the plan for assessing the reliability of the scoring procedure

- Recruiting and training data entry personnel for keying the responses of students, teachers, and principals into computerized data files, and conducting the data entry operation using the software provided

- Checking the accuracy and integrity of the data files prior to shipping them to the IEA Data Processing Center in Hamburg

In addition to their role in implementing the TIMSS data collection procedures, NRCs were responsible for conducting analyses of their national data and for reporting on the results of TIMSS in their own countries.`

The TIMSS International Study Director was responsible for the overall direction and coordination of the project. The TIMSS International Study Center, located at Boston College in the United States, was responsible for supervising all aspects of the design and implementation of the study at the international level. This included the following:

- Planning, conducting, and coordinating all international TIMSS activities, including meetings of the International Steering Committee, NRCs, and advisory committees

- Developing and field testing the data collection instruments

- Developing sampling procedures for efficiently selecting representative samples of students in each country, and monitoring sampling operations to ensure that they conformed to TIMSS requirements

- Designing and documenting operational procedures to ensure efficient collection of all TIMSS data

- Designing and implementing a quality assurance program encompassing all aspects of the TIMSS data collection, including monitoring of test administration sessions in participating countries

- Supervising the checking and cleaning of the data from the participating countries, the construction of the TIMSS international database, the computation of sampling weights, and the scaling of the achievement data

- Analyzing the international data and writing and disseminating the international reports

The International Study Center was supported in its work by the following advisory committees:[4]

- The International Steering Committee, which advised on policy issues and on the general direction of the study

- The Subject Matter Advisory Committee, which advised on all matters relating to mathematics and science subject matter, particularly the content of the achievement tests

- The Technical Advisory Committee, which advised on all technical issues related to the study, including study design, sampling design, achievement test construction and scaling, questionnaire design, database construction, data analysis, and reporting

- The Performance Assessment Committee, which developed the TIMSS performance assessment and advised on the analysis and reporting of the performance assessment data

- The Free-Response Item Coding Committee, which developed the coding rubrics for the free-response items

- The Quality Assurance Committee, which helped to develop the TIMSS quality assurance program

- The Advisory Committee on Curriculum Analysis, which advised the International Study Director on matters related to the curriculum analysis

4 See the Acknowledgments section for membership of TIMSS committees.

Several important TIMSS functions, including test and questionnaire development, translation checking, sampling consultations, data processing, and data analysis, were conducted by centers around the world under the direction of the TIMSS International Study Center. In particular, the following centers have played important roles in the TIMSS project.

- The IEA Data Processing Center (DPC), located in Hamburg, Germany, was responsible for checking and processing all TIMSS data and for constructing the international database. The DPC played a major role in developing and documenting the TIMSS field operations procedures

- Statistics Canada, located in Ottawa, Canada, was responsible for advising NRCs on their sampling plans, for monitoring progress in all aspects of sampling, and for the computation of sampling weights

- The Australian Council for Educational Research (ACER), located in Melbourne, Australia, participated in the development of the achievement tests, conducted psychometric analyses of field trial data, and was responsible for the development of scaling software and for scaling the achievement test data

- The International Coordinating Center (ICC) in Vancouver, Canada, was responsible for the international project coordination prior to the establishment of the International Study Center in August 1993. Since then, the ICC has provided support to the International Study Center, particularly in managing translation verification in the achievement test development process, and has published several monographs in the TIMSS monograph series

- As Sampling Referee, Keith Rust of Westat, United States, worked with Statistics Canada and the NRCs to ensure that sampling plans met the TIMSS standards, and advised the International Study Director on all matters relating to sampling

1.4 SUMMARY OF THIS REPORT

The variation across countries regarding the nature of upper secondary education systems, including what constitutes the in-school population, what programs of study students follow, and when students finish secondary school, posed many challenges in sampling schools and students. In Chapter 2 of this report, Jean Dumais describes the implementation of the TIMSS sample design for Population 3: how students were stratified according to their academic preparation, how schools and students were sampled, how TIMSS quantified the coverage of the school-leaving age cohort with the TIMSS Coverage Index (TCI), the response rates for each country, and how TIMSS documented the extent to which the sampling guidelines were followed in each country.

To ensure the availability of comparable, high-quality data for analysis, TIMSS took a set of rigorous quality control steps to create the international database. TIMSS prepared manuals and software for countries to use in entering their data so that the information would be in a standardized international format before it was forwarded to the IEA Data Processing Center (DPC) in Hamburg for creation of the international database. Upon arrival at the Center, the data from each country underwent an exhaustive cleaning process. That process involved several iterative steps and procedures designed to identify, document, and correct deviations from the international instruments, file structures, and coding schemes. The process also emphasized consistency of information within national data sets and appropriate linking among the many student, teacher, and school data files. Following the data cleaning and file restructuring by the DPC, Statistics Canada computed the sampling weights and the Australian Council for Educational Research computed the item statistics and scale scores. These additional data were merged into the database by the DPC. Throughout, the International Study Center reviewed the data and managed the data flow. In Chapter 3, Heiko Sibberns, Dirk Hastedt, Michael Bruneforth, Knut Schwippert, and Eugenio Gonzalez describe the TIMSS data management, including procedures for cleaning and verifying the data and the links across files, restructuring of the national data files to the standard international format, the various data reports produced throughout the cleaning process, and the computer systems used to undertake the data cleaning and construction of the database.

Within countries, TIMSS used a two-stage sample design for Populations 3. The first stage involved selecting 120 public and private schools within each country. Within each school, the basic approach required countries to use random procedures to select 40 students. The actual number of schools and students selected depended in part on the structure of the education system – tracked or untracked – and on where the student subpopulations were in the system. The complex sampling approach required the use of sampling weights to account for the differential probabilities of selection and to adjust for non-response in order to ensure the computation of proper survey estimates. Statistics Canada was responsible for computing the sampling weights for the TIMSS countries. In Chapter 4, Jean Dumais and Pierre Foy describe the derivation of school and student weights.

Because the statistics presented in the TIMSS reports are estimates of national performance based on samples of students, rather than the values that could be calculated if every student in every country had answered every question, it is important to have measures of the degree of uncertainty of the estimates. The complex sampling approach that TIMSS used had implications for estimating sampling variability. Because of the effects of cluster selection and the effects of certain adjustments to the sampling weights, standard procedures for estimating the variability of sample statistics generally underestimate the true variability of the statistics. To avoid this problem, TIMSS used the jackknife procedure to estimate the standard errors associated with each statistic presented in the international reports. In Chapter 5, Eugenio Gonzalez and Pierre Foy describe the jackknife technique and its application to the TIMSS data in estimating the variability of the sample statistics.

Prior to scaling, the TIMSS cognitive data were thoroughly checked by the IEA Data Processing Center, the International Study Center, and the national centers. The national centers were contacted regularly and given multiple opportunities to review the data for their countries. In conjunction with the Australian Council for Educational Research, the International Study Center conducted a review of item statistics for each of the mathematics and science literacy, advanced mathematics, and physics items in each of the countries to identify poorly performing items. In Chapter 6, Ina Mullis and Michael Martin describe the procedures used to ensure that the cognitive data included in the scaling and the international database are comparable across countries.

The complexity of the TIMSS test design and the desire to compare countries' performance on a common scale led TIMSS to use item response theory to summarize the achievement results. TIMSS reported scale scores for mathematics literacy; science literacy; advanced mathematics; three advanced mathematics content areas; physics; and five physics content areas. These scales were based on a variant of the Rasch item response model. The model, developed by Adams, Wilson, and Wang (1997), includes refinements that enable reliable scores to be produced even though individual students responded to relatively small subsets of the total item pools. This approach was preferred for developing comparable estimates of performance for all students, since students answered different test items depending on which of the test booklets they received. In Chapter 7, Greg Macaskill, Ray Adams, and Margaret Wu describe the scaling methodology and procedures used to produce the TIMSS achievement scores, including the estimation of international item parameters and the derivation and use of plausible values to provide estimates of performance.

TIMSS reported achievement from a number of perspectives. Mean achievement and percentiles of distribution were reported by country for mathematics and science literacy, advanced mathematics, and physics, and significant differences between countries (adjusted for multiple comparisons) were also reported. To show whether or not countries may have achieved higher performance because they tested fewer students and, in particular, a more elite group of students, TIMSS showed the relationship between the TIMSS Coverage Index and achievement for mathematics and science literacy, advanced mathematics, and physics. TIMSS also reported achievement for the school-leaving age cohort, regardless of the coverage of this cohort by the sample; achievement was reported for the top 25 percent of students in mathematics and science literacy, and the top 10 percent and 5 percent of students in both advanced mathematics and physics. TIMSS also compared countries' achievement on the final-year mathematics and science literacy test with achievement on the Population 2 mathematics and science tests, in relationship to the international averages. In Chapter 8, Eugenio Gonzalez describes the analyses undertaken to report the achievement scale scores in these various ways in the international reports.

REFERENCES

Adams, R.J., Wilson, M.R., and Wang, W.C. (1997). The multidimensional random coefficients multinomial logit. *Applied Psychological Measurement, 21,* 1-24.

Beaton, A.E., Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., Smith, T.A., and Kelly, D.L. (1996). *Science achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Boston College.

Beaton, A.E., Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Kelly, D.L., and Smith, T.A. (1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Boston College.

Garden, R.A. and Orpwood, G. (1996). Development of the TIMSS achievement tests. *TIMSS technical report, volume I: Design and development*. Chestnut Hill, MA: Boston College.

Harmon, M., Smith, T.A., Martin, M.O., Kelly, D.L., Beaton, A.E., Mullis, I.V.S., Gonzalez, E.J., and Orpwood, G. (1997). *Performance assessment in IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Boston College.

Martin, M.O., and Kelly, D.L., Eds. (1996). *TIMSS technical report, volume I: Design and development*. Chestnut Hill, MA: Boston College.

Martin, M.O., and Kelly, D.L., Eds. (1997). *TIMSS technical report, volume II: Implementation and analysis, primary and middle school years*. Chestnut Hill, MA: Boston College.

Martin, M.O., Mullis, I.V.S., Beaton, A.E., Gonzalez, E.J., Smith, T.A., and Kelly, D.L. (1997). *Science achievement in the primary school years: IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., and Smith, T.A. (1998). *Mathematics and science achievement in the final year of secondary school: IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., and Smith, T.A. (1997). *Mathematics achievement in the primary school years: IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Boston College.

Orpwood, G. and Garden, R.A. (1998). *TIMSS monograph no. 4: Assessing mathematics and science literacy*. Vancouver, B.C.: Pacific Educational Press.

Robitaille, D.F. and Garden (1996). *TIMSS monograph no. 2: Research questions and study design*. Vancouver, Canada: Pacific Educational Press.

Robitaille, D.F., McKnight, C., Schmidt, W., Britton, E., Raizen, S., and Nicol, C. (1993). *TIMSS monograph no. 1: Curriculum frameworks for mathematics and science.* Vancouver, B.C.: Pacific Educational Press.

Schmidt, W.H., McKnight, C.C., Valverde, G.A., Houang, R.T., and Wiley, D.E. (1997). *Many visions, many aims: A cross-national investigation of curricular intentions in school mathematics.* Dordrecht, the Netherlands: Kluwer Academic Publishers.

Schmidt, W.H., Raizen, S.A., Britton, E.D., Bianchi, L.J., and Wolfe, R.G. (1997). *Many visions, many aims: A cross-national investigation of curricular intentions in school science.* Dordrecht, the Netherlands: Kluwer Academic Publishers.

2 Implementation of the TIMSS Sampling Design

Jean Dumais
Statistics Canada

2.1 THE TARGET POPULATION

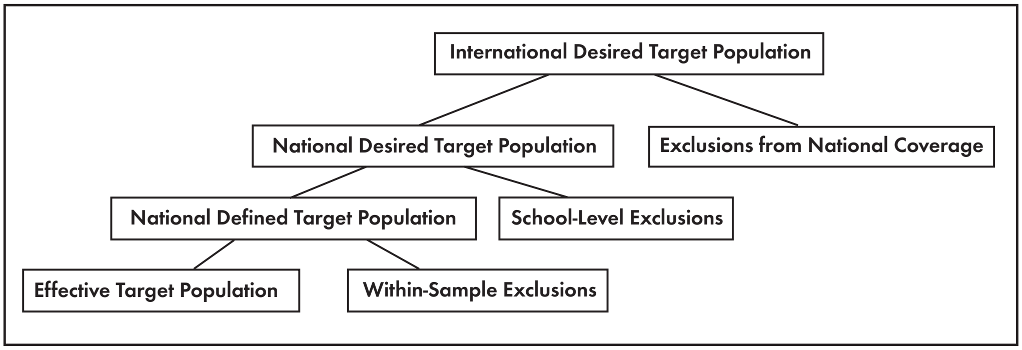
The selection of valid and efficient samples is crucial to the quality and success of an international comparative study such as TIMSS. The accuracy of the survey results depends on the quality of the sampling information available when planning the sample, and on the care with which the sampling activities themselves are conducted. For TIMSS, National Research Coordinators (NRCs) worked on all phases of sampling with staff from Statistics Canada. NRCs were trained in how to select the school and student samples and how to use of the sampling software. In consultation with the TIMSS sampling referee (Keith Rust, Westat), staff from Statistics Canada reviewed the national sampling plans, sampling data, sampling frames, and sample selection. This documentation was used by the International Study Center jointly with Statistics Canada, the sampling referee, and the Technical Advisory Committee to evaluate the quality of the samples.

The assessment of final-year students was intended to measure what might be considered the "yield" of the elementary and secondary education systems of a country with regard to mathematics and science. This was done by assessing the mathematics and science literacy of all students in the final year of secondary school, the advanced mathematics knowledge of students having taken advanced mathematics courses, and the physics knowledge of students having taken physics. The International Desired Population, then, was all students in the final year of secondary school, with those having taken advanced mathematics courses and those having taken physics courses as two overlapping sub-populations. Students repeating the final year were not part of the desired population. For each secondary education track in a country, the final grade of the track was identified as being part of the target population, allowing substantial coverage of students in their final year of schooling. For example, grade 10 could be the final year of a vocational program, and grade 12 the final year of an academic program. Both of these grade/track combinations are considered part of the population (but grade 10 in the academic track is not). Appendix A of Mullis et al. (1998) describes the structure of the upper secondary education systems and the students tested in each country. Appendix B of this volume gives more details of the population definition and sample design for each country.

2.2 COVERAGE OF THE TIMSS TARGET POPULATION

The stated objective in TIMSS was that the effective population, the population actually sampled by TIMSS, be as close as possible to the International Desired Population. Figure 2.1 illustrates the relationship between the desired populations and the excluded populations at the country, school, and student levels.

Figure 2.1 Relationship Between the Desired Populations and Exclusions



Using the International Desired Population as a basis, participating countries had to operationally define their population for sampling purposes. Occasionally, NRCs had to restrict coverage at the country level, for example by excluding remote regions or a segment of the education system. In these few situations, countries were permitted to define a National Desired Population that did not include part of the International Desired Population. Exclusions could be based on geographic areas or language groups. Table 2.1 shows differences in coverage between the International and National Desired Populations. Most participants achieved 100 percent coverage (20 out of 24). The countries with less than 100 percent coverage are footnoted in tables in the international report. Israel and Lithuania, as a matter of practicality, needed to define their tested populations according to the structure of their school systems. Latvia, which participated only in the physics assessment, limited its testing to Latvian-speaking schools. Because coverage fell below 65 percent, the Latvian results have been labeled Latvia (LSS), for Latvian Speaking Schools, in the tables presenting results for the physics assessment. Italy was unable to include 4 of its 20 regions.

Within the National Desired Population, countries could exclude a small percentage – less than 10 percent – of certain kinds of schools or students that would be very difficult or resource-intensive to test, such as schools for students with special needs, or schools that were very small or located in extremely remote areas. Some countries also excluded students in particular tracks or school types. These exclusions are also shown in Table 2.1. The countries with particularly high exclusions are so footnoted in the achievement tables in the report.

Table 2.1 Coverage of TIMSS Target Population
The International Desired Population is defined as follows:
Population 3 – All students in final year of secondary school

Country	International Desired Population		National Desired Population	
	Country Coverage	Notes on Coverage	Sample Exclusions	Notes on Exclusions
Australia	100%		5.5%	
Austria	100%		18.2%	Colleges and courses lasting less than 3 years excluded
Canada	100%		8.9%	
Cyprus	100%		22.0%	Private and vocational schools excluded
Czech Republic	100%		6.0%	
Denmark	100%		2.3%	
France	100%		1.0%	
Germany	100%		11.3%	
Greece	100%		85.0%	Only students having taken advanced mathematics and physics included
Hungary	100%		0.2%	
Iceland	100%		0.1%	
Israel	74%	Hebrew public education system	0.0%	
Italy	70%	Four regions did not participate	0.9%	
Latvia (LSS)	50%	Latvian speaking students	85.0%	Only students having taken physics included
Lithuania	84%	Lithuanian speaking students	0.0%	
Netherlands	100%		21.6%	Apprenticeship programs excluded
New Zealand	100%		0.0%	
Norway	100%		3.8%	
Russian Federation	100%		43.0%	Vocational schools and non-Russian-speaking students excluded
Slovenia	100%		6.0%	
South Africa	100%		0.0%	
Sweden	100%		0.2%	
Switzerland	100%		2.5%	
United States	100%		3.7%	

Because population coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.

Participants could exclude schools from the sampling frame if they were geographically remote, were extremely small, had a curriculum or structure different from the mainstream, or provided instruction only to students in the "within-school" exclusion categories. The general TIMSS rules for defining within-school exclusions follow.

- **Educable mentally disabled students.** These are students who are considered, in the professional opinion of the school principal or other qualified staff members, to be educable mentally disabled students, or who have been so diagnosed in psychological tests. This includes students who are emotionally or mentally unable to follow even the general instructions of the TIMSS test. It does not include students who merely exhibit poor academic performance or discipline problems.

- **Functionally disabled students.** These are students who are permanently physically disabled in such a way that they could not perform in the TIMSS tests. Functionally disabled students who could perform in the TIMSS test were included in the testing.

- **Non-native-language speakers.** These are students who cannot read or speak the language of the test and so could not overcome the language barrier of testing. Typically, students who had received less than one year of instruction in the language of the test were excluded, but this definition was adapted in different countries. Some countries opted to test students in more than one language.

2.3 TIMSS COVERAGE INDEX

Historically, an important difference between education systems was the proportion of an age cohort that successfully completed upper secondary education. In order to avoid unwittingly comparing the elite students in one country with the more general population in another, therefore, it is important to be aware of the extent to which the upper secondary system in each country includes the total student population.

So as to learn how much of the school-leaving age cohort was still in school and represented by the TIMSS sample, a TIMSS Coverage Index (TCI) was computed for each country. The TCI is an estimate of the percentage of the school-leaving age cohort covered by the TIMSS final-year student sample. It reflects any omissions from the sample, such as students who were excluded because of handicap or who had dropped out of school, and, in some countries, tracks or educational programs that were not covered by the TIMSS sample. The TCI was computed by forming a ratio of the size of the student population covered by the TIMSS sample, as estimated from the sample itself, to the size of the school-leaving age cohort, which was derived from official population census figures supplied by each country. The TCI was defined as follows:

$$TCI = \frac{Total\ enrollment\ in\ TIMSS\ Grades\ \mathrm{1995}}{(Total\ national\ population\ aged\ \mathrm{15\text{-}19}\ in\ \mathrm{1995})/5}$$

The *numerator* in this expression is the total enrollment in the grades tested by TIMSS, estimated from the weighted sample data. This estimate corresponds to the size of the population to which the TIMSS results generalize, and makes appropriate provision for student non-response. It does not include students who are no longer attending school, or students who were excluded from the sample on grounds of physical or other disability. It also does not include students who were repeating the final grade.

The *denominator* in the expression is an estimate of the school-leaving age cohort size. Since the age at which upper secondary students may leave school varies, TIMSS estimated the size of the school-leaving age cohort by taking the average of the size of the 1995 age cohorts for 15-, 16-, 17-, 18-, and 19-year-olds in each country. (Although the general procedure was to base the estimate on the 15-19 age group, there were exceptions. For example, in Germany, the estimate was based on the 17-19 age group.) This information was provided by National Research Coordinators from official population census figures in their countries. This approach reflects the fact that students in the final year of secondary school are likely to be almost entirely a subset of the population of 15- to 19-year-olds in most countries.

Table 2.2 Computation of TCI: Estimated Percentage of School-Leaving Age Cohort Covered by TIMSS Sample Final Year of Secondary School

Country	Estimated School-Leaving Age Cohort Size (A)	Estimated Number of Students Represented by Sample (B)	Estimated Number of Students Excluded from Sample (C)	Estimated Number of Other Students Not Represented by Sample (D)	TIMSS Coverage Index (TCI)† (B/A)
Australia	250,852	170,849	9,944	70,059	68%
Austria	93,168	70,721	15,682	6,765	76%
Canada	374,499	263,241	25,559	85,699	70%
Cyprus	9,464	4,535	1,279	3,650	48%
Czech Republic	177,180	137,467	8,821	30,892	78%
Denmark	65,683	37,872	872	26,939	58%
France	760,452	637,935	6,509	116,008	84%
Germany	870,857	655,916	83,514	131,427	75%
[1] Greece	146,400	14,668	83,119	48,613	10%
Hungary	170,524	111,281	201	59,042	65%
Iceland	4,231	2,308	2	1,921	55%
Israel	-	-	-	-	-
Italy	739,268	380,834	3,459	354,975	52%
[2] Latvia (LSS)	33,096	979	5,548	26,569	3%
Lithuania	52,140	22,160	0	29,980	43%
Netherlands	187,087	145,916	40,293	878	78%
New Zealand	53,284	37,549	4	15,731	70%
Norway	52,180	43,806	1,747	6,627	84%
Russian Federation	2,145,918	1,031,187	777,913	336,818	48%
Slovenia	30,354	26,636	1,706	2,012	88%
South Africa	766,334	374,618	0	391,716	49%
Sweden	101,058	71,333	168	29,557	71%
Switzerland	79,547	65,174	1,671	12,702	82%
United States	3,612,800	2,278,564	88,642	1,245,594	63%

† TIMSS Coverage Index (TCI): Estimated percentage of school-leaving age cohort covered by TIMSS sample.

[1] Greece sampled only students having taken advanced mathematics and physics.

[2] Latvia (LSS) sampled only students having taken physics.

Because population coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.

A dash (-) indicates data are not available.

In countries with high TCIs most of the students are still in school and are covered in the TIMSS sample. Countries with low TCIs have fewer students still in school, or have excluded some components of their system from their sample (or both). Table 2.2 presents the TCI for each country, and also shows the two parts of the portion of the school-leaving age cohort not covered by the TIMSS sample: system components and students excluded by the country, and others – primarily young people who chose not to complete upper secondary education. The percentage of the age cohort covered by the TIMSS sample (the TCI), the percentage excluded from the sample, and the percentage of others not covered combine to form 100 percent of the school-leaving age cohort. For example, Australia has a TCI of 68.1 percent, which indicates that the TIMSS sample of final-year students covers just over two-thirds of the school-leaving age cohort. Of the rest, 4 percent have been excluded from the sample, and the remaining 27.9 percent are presumably no longer attending school. The TCI for Cyprus is lower (47.9 percent), partly because Cyprus excluded students in private schools and in vocational programs (13.5 percent), and partly because a greater percentage of the age cohort is no longer attending school (38.6 percent).

In order to quantify the coverage of the advanced mathematics and physics samples and help interpret the achievement results for these students, TIMSS computed a Mathematics TIMSS Coverage Index (MTCI) and a Physics TIMSS Coverage Index (PTCI), as shown in Table 2.3. The MTCI is the overall TCI multiplied by the percentage of the final-year sample having taken advanced mathematics. For example, in Australia 23.1 percent of the final-year sample had taken advanced mathematics. Multiplying this by the TCI (68.1 percent, from Table 2.2) gives a MTCI of 15.7 percent, as shown in the second column of Table 2.3. This implies that about 16 percent of the school-leaving age cohort in Australia had taken advanced mathematics in upper secondary school. Similarly, the PTCI for Australia is 12.6 percent, as shown in the fourth column of Table 2.3.

Table 2.3 TIMSS Coverage Indices (TCIs) for Advanced Mathematics and Physics Final Year of Secondary School

Country	Percentage of Students in Sample Having Taken Advanced Mathematics	Mathematics TIMSS Coverage Index (MTCI)*	Percentage of Students in Sample Having Taken Physics	Physics TIMSS Coverage Index (PTCI)†
Australia	23.1%	15.7%	18.5%	12.6%
Austria	43.9%	33.3%	43.5%	33.1%
Canada	22.3%	15.6%	19.4%	13.7%
Cyprus	18.5%	8.8%	18.5%	8.8%
Czech Republic	14.1%	11.0%	14.1%	11.0%
Denmark	35.7%	20.6%	5.5%	3.2%
France	23.8%	19.9%	23.8%	19.9%
Germany	34.9%	26.3%	11.2%	8.4%
[1] Greece	-	10.0%	-	10.0%
[2] Israel	-	-	-	-
Italy	27.4%	14.1%	16.7%	8.6%
[3] Latvia	-	-	-	3.0%
Lithuania	6.1%	2.6%	-	-
Norway	-	-	10.0%	8.4%
Russian Federation	4.2%	2.0%	3.2%	1.5%
Slovenia	85.9%	75.4%	43.9%	38.6%
Sweden	23.0%	16.2%	23.1%	16.3%
Switzerland	17.4%	14.3%	17.3%	14.2%
United States	21.8%	13.7%	22.9%	14.5%

* MTCI: Estimated percentage of school-leaving age cohort covered by TIMSS sample of advanced mathematics students.
† PTCI: Estimated percentage of school-leaving age cohort covered by TIMSS sample of physics students.
[1] Greece sampled only students having taken advanced mathematics and physics.
[2] The MTCI and the PTCI could not be computed for Israel.
[3] Latvia sampled only students having taken physics.
A dash (-) indicates data are not available.
Note: Hungary, Iceland, the Netherlands, New Zealand, and South Africa did not participate in the advanced mathematics and physics testing. Norway did not participate in the advanced mathematics testing and Lithuania did not participate in the physics testing.

2.4 SAMPLE DESIGN

One of the goals of TIMSS was to assess the mathematics and science literacy of all students while also assessing the advanced mathematics and physics knowledge of students with preparation in these subjects. To that end, a sampling design had to be developed that ensured that students were stratified according to their level of preparation in mathematics and physics, so that appropriate test booklets could be assigned to them. According to the TIMSS design each student is characterized as having taken advanced mathematics (M) or not (O), and as having taken physics (P) or not (O). Combining these two-way classifications yields four mutually exclusive and exhaustive categories of students:

OO Students having studied neither advanced mathematics nor physics

OP Students having studied physics but not advanced mathematics

MO Students having studied advanced mathematics but not physics

MP Students having studied both advanced mathematics and physics

Four kinds of student test booklets were assigned to students on the basis of this classification scheme (OO, OP, MO, MP), so that each student completed one 90-minute booklet. Students classified as OO received either booklet 1A or 1B, the two booklets containing items related to mathematics and science literacy. Students classified as OP received either booklet 1A or 1B, or one of the three booklets containing physics material (2A, 2B, or 2C). Students classified as MO received either booklet 1A or 1B, or one of the three booklets containing advanced mathematics material (3A, 3B, or 3C). Students classified as MP also received one booklet, which could have been any one of the booklets (1A, 1B, 2A, 2B, 2C, 3A, 3B, 3C, or 4). Booklet 4 contained mathematics and science literacy, advanced mathematics, and physics items. The TIMSS test design is described in detail in Adams and Gonzalez (1996).

Table 2.4 Assignment of Test Booklets According to Student Classification

Student Type	Booklet Assigned
OO	1A or 1B
OP	1A, 1B, 2A, 2B or 2C
MO	1A, 1B, 3A, 3B or 3C
MP	1A, 1B, 2A, 2C, 3A, 3B, 3C or 4

The student samples used for estimating the parameters of the three populations of interest (all students in their final year of secondary school, final-year students having taken advanced mathematics, and final-year students having taken physics) were constructed by combining the students who had been assigned the appropriate booklets. Thus, the student sample for estimating proficiency in mathematics and science literacy was made up of all the students from each of the four student groups (OO, OP, MO, and MP) who were assigned one of the literacy booklets (booklets 1A or 1B) or the combined literacy, mathematics, and physics booklet (booklet 4). This ensured that each type of student was properly represented in the final-year sample. The sample for estimating proficiency in advanced mathematics consisted of students from the MO and MP groups who were assigned one of the mathematics booklets (booklets 3A, 3B, or 3C) or the combined literacy, advanced mathematics, and physics booklet (booklet 4). The sample for estimating proficiency in physics consisted of students from the OP and MP groups who were assigned one of the physics booklets (booklets 2A, 2B, or 2C) or the combined literacy, advanced mathematics, and physics booklet (booklet 4).

2.5 REQUIREMENTS FOR SAMPLING PRECISION

The general standards for sampling precision established for TIMSS are discussed in Foy, Rust, and Schleicher (1996). The sampling precision requirement for mathematics and science literacy was a confidence interval of no more than ±0.1 standard deviation units at the 95 percent confidence level. Although efforts were made to ensure the same precision for the advanced mathematics and physics scales, it was recognized that circumstances in participating countries would make this sometimes difficult to achieve. Because of this, TIMSS participants, in consultation with the sampling coordinators, were permitted to design samples for these scales that would achieve confidence intervals of ±0.15 standard deviation units at the 95 percent confidence level.

The sampling design was a two-stage process, with schools sampled with probability-proportional-to-size (PPS) in the first stage, and a fixed number of students sampled in the second stage. To meet the TIMSS standard for sampling precision, approximately 120 schools were required in each country. Within each sampled school, students were classified OO, OP, MO, MP, and 40 students sampled at random, 10 from each category. Because the organization of school systems at upper secondary level varies considerably across countries, each country had to work with the sampling consultants to adapt the basic design to the local situation.

The basic design was well suited to comprehensive systems, where schools cater to all kinds of students, and students must be classified individually or on the basis of the courses they have taken. However, many of the TIMSS countries operate tracked systems, where students are assigned to particular types of schools on the basis of their academic interests and abilities. In such countries it was often possible to stratify whole schools in terms of whether or not they contained advanced mathematics or physics classes. In such systems it was sometimes possible to refine the basic design so as to achieve the required sampling precision with a smaller sample of schools. Although there was no analytic requirement to sample students in whole classes rather than individually, some countries found it more convenient to do so, even though the increased clustering effect sometimes necessitated larger sample sizes.

2.6 SCHOOL SAMPLING

The sample-selection method used for first-stage sampling was based on a systematic probability-proportional-to-size technique. Countries were encouraged to stratify schools by important demographic variables (e.g., geographical region, public/private) as well as by school type. Small schools were handled either by assigning them to separate strata or by combining them with larger schools to form pseudo-schools for sampling purposes. Some very large countries introduced a preliminary sample stage before schools were sampled, in which the country was divided into primary sampling units. Within each stratum, schools were listed in order of any implicit stratification variables, and then further sorted according to their measure of size. Schools were then sampled using a random-start fixed-interval procedure that ensured selection with probability-proportional-to-size.

Sometimes a sampled school was unable to participate in the assessment. In such cases, it was replaced by a *replacement school*. The mechanism for selecting replacement schools, established a priori, identified the next school on the ordered school-sampling list as the replacement for each sampled school, and the one after that as a second replacement, should it be necessary. Since schools were grouped by stratification variables and by size on the sampling frame, a replacement school should have characteristics similar to the originally selected school.

2.7 STUDENT SAMPLING

Whereas schools were sampled with probability- proportional-to-size, the basic design called for a fixed number of students to be sampled within each school in the second stage of sampling. This gives selection probabilities for students that are inversely proportional to school size. The combined school and student selection probabilities result in an overall selection probability that is equal for all students in each explicit stratum. In untracked schools, students were classified into one of the four groups (OO, OP, MO, MP), and a sample of 10 students was drawn from each group. If just three student types were present (for example if there were no OP students, as sometimes happened) three samples of 13 students were drawn. In schools with no advanced mathematics or physics students, all 40 students were sampled from the OO group. In some tracked systems, schools frequently consisted either of only OO students or of only MP students. In these situations all 40 students were sampled from the appropriate group. Detailed procedures for sampling students were specified within schools for a variety of school organizations. These procedures are presented in Schleicher and Siniscalco (1996).

2.8 PARTICIPATION RATES

Weighted and unweighted participation rates were computed for each participating country, at the school level and at the student level for each assessment (mathematics and science literacy, advanced mathematics, and physics). Overall response rates (combined school and student response rates) also were computed for each assessment.

2.8.1 School-Level Participation Rates

The general formula for computing weighted school-level participation rates is shown in the following equation:

$$R_{wgt}(sch) = \frac{\sum_{part} MOS_i / \pi_i}{\sum_{elig} MOS_i / \pi_i}$$

For each sampled school, the ratio of its measure of size (MOS) to its selection probability (π_i) was computed. The weighted school-level response rate is the sum of the ratios for all participating schools divided by the sum of the ratios for all eligible schools. The unweighted school-level response rates were computed in a similar way, where all school ratios were set to one. This becomes simply the number of participating schools in the sample divided by the number of eligible schools in the sample. Since in most cases, in selecting the sample, the value of π_i was set proportional to MOS_i within each explicit stratum, weighted and unweighted rates were generally similar.

2.8.2 Student-Level Participation Rates

The general formula for computing student-level participation rates is shown in the following equation:

$$R_{wgt}(stu) = \frac{\sum_{part} 1/p_j}{\sum_{elig} 1/p_j}$$

where p_j denotes the probability of selection of the student, incorporating all stages of selection. Thus the weighted student-level participation rate is the sum of the inverse of the selection probabilities for all participating students divided by the sum of the inverse of the selection probabilities for all eligible students. The unweighted student participation rates were computed in a similar way, but with each student contributing equal weight.

2.8.3 Overall Participation Rates

The overall participation rate was calculated as the product of the weighted school-level participation rate without replacement schools and the weighted student-level participation rate. School and student sample sizes and participation rates are presented in Tables 2.5 to 2.11.

Table 2.5 School Sample Sizes – Mathematics and Science Literacy Final Year of Secondary School

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	132	132	71	16	87
Austria	182	182	74	95	169
Canada	389	389	333	4	337
Cyprus	29	28	28	0	28
Czech Republic	150	150	150	0	150
Denmark	130	130	122	0	122
France	71	71	56	0	56
Germany	174	174	121	31	152
Hungary	204	204	204	0	204
Iceland	30	30	30	0	30
Israel	125	125	52	0	52
Italy	150	150	93	8	101
Lithuania	168	142	142	0	142
Netherlands	141	141	52	27	79
New Zealand	79	79	68	11	79
Norway	171	171	122	9	131
Russian Federation	175	165	159	4	163
Slovenia	172	172	79	0	79
South Africa	185	140	90	0	90
Sweden	157	157	145	0	145
Switzerland	401	401	378	5	383
United States	250	250	190	21	211

Table 2.6 School Sample Sizes – Advanced Mathematics Final Year of Secondary School

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	132	132	68	15	83
Austria	182	119	48	66	114
Canada	389	389	306	3	309
Cyprus	29	21	21	0	21
Czech Republic	90	90	90	0	90
Denmark	130	130	115	0	115
France	69	69	61	0	61
Germany	76	76	53	23	76
Greece	60	60	45	15	60
Israel	125	125	44	0	44
Italy	59	59	41	1	42
Lithuania	29	29	29	0	29
Russian Federation	132	117	112	1	113
Slovenia	172	159	73	0	73
Sweden	157	157	101	0	101
Switzerland	198	198	195	2	197
United States	250	250	180	19	199

Because population coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.

Table 2.7 School Sample Sizes – Physics Final Year of Secondary School

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	132	132	69	16	85
Austria	182	119	48	66	114
Canada	389	389	304	3	307
Cyprus	29	21	21	0	21
Czech Republic	90	90	90	0	90
Denmark	130	130	77	0	77
France	69	69	61	0	61
Germany	74	74	52	22	74
Greece	60	60	45	15	60
Israel	125	125	46	0	46
Italy	29	29	20	0	20
Latvia (LSS)	45	45	38	0	38
Norway	70	70	63	3	66
Russian Federation	132	98	83	1	84
Slovenia	172	172	52	0	52
Sweden	157	157	101	0	101
Switzerland	198	198	195	2	197
United States	250	250	184	19	203

Because population coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.

Table 2.8 Student Sample Sizes Final Year of Secondary School

Country	Number of Students Sampled in Participating Schools	Number of Students Withdrawn†	Number of Students Excluded	Number of Students Eligible	Number of Students Absent
Australia	4130	37	0	4093	1040
Austria	3693	140	0	3553	398
Canada	11782	732	0	11050	1470
Cyprus	1224	15	0	1209	38
Czech Republic	4188	43	0	4145	326
Denmark	5208	0	0	5208	672
France	4096	275	0	3821	600
Germany	6971	94	117	6760	1666
Greece	1246	261	0	985	180
Hungary	5493	265	0	5228	137
Iceland	2500	132	2	2366	663
Israel	2568	0	0	2568	29
Italy	2426	148	3	2275	192
Latvia (LSS)	780	6	0	774	66
Lithuania	4196	1	0	4195	574
Netherlands	1882	181	20	1681	211
New Zealand	2687	580	1	2106	343
Norway	4056	76	65	3915	349
Russian Federation	5356	536	0	4820	182
Slovenia	3755	37	1	3717	282
South Africa	3695	906	0	2789	32
Sweden	5362	184	12	5166	589
Switzerland	5939	258	0	5681	262
United States	14812	603	293	13916	3082

† Sampled students who reported that they were repeating the final year, were incorrectly classified, or were otherwise ineligible.
Because population coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.

Table 2.9 Participation Rates – Mathematics and Science Literacy Final Year of Secondary School

Country	School Participation		Student Participation (Weighted Percentage)	Overall Participation	
	School Participation Before Replacement (Weighted Percentage)	School Participation After Replacement (Weighted Percentage)		Overall Participation Before Replacement (Weighted Percentage)	Overall Participation After Replacement (Weighted Percentage)
Australia	48.8	66.2	78.1	38.1	51.8
Austria	35.9	90.9	79.7	28.6	72.5
Canada	82.2	82.6	82.7	68.0	68.3
Cyprus	100.0	100.0	98.2	98.2	98.2
Czech Republic	100.0	100.0	92.2	92.2	92.2
Denmark	54.9	54.9	88.9	48.8	48.8
France	80.3	80.3	85.6	68.7	68.7
Germany	88.7	100.0	80.1	71.0	80.1
Hungary	100.0	100.0	97.7	97.7	97.7
Iceland	100.0	100.0	73.6	73.6	73.6
Israel	48.8 **	48.8 **	98.3 **	48.0 **	48.0 **
Italy	59.9	65.0	94.8	56.8	61.6
Lithuania	97.1	97.1	87.9	85.4	85.4
Netherlands	35.8	56.3	87.6	31.3	49.3
New Zealand	87.0	100.0	80.6	70.1	80.6
Norway	74.1	80.0	88.9	65.9	71.1
Russian Federation	93.0	99.3	90.9	84.6	90.3
Slovenia	45.6	45.6	92.8	42.3	42.3
South Africa	65.0	65.0	99.4	64.6	64.6
Sweden	95.3	95.3	86.5	82.4	82.4
Switzerland	87.0	89.1	95.0	82.6	84.6
United States	77.1	85.1	74.6	57.6	63.5

** Unweighted participation rates.

Table 2.10 Participation Rates – Advanced Mathematics Final Year of Secondary School

Country	School Participation		Student Participation (Weighted Percentage)	Overall Participation	
	School Participation Before Replacement (Weighted Percentage)	School Participation After Replacement (Weighted Percentage)		Overall Participation Before Replacement (Weighted Percentage)	Overall Participation After Replacement (Weighted Percentage)
Australia	47.3	63.6	86.7	40.9	55.2
Austria	36.7	95.5	84.6	31.0	80.8
Canada	84.6	85.2	90.4	76.4	76.9
Cyprus	100.0	100.0	96.0	96.0	96.0
Czech Republic	100.0	100.0	92.1	92.1	92.1
Denmark	54.9	54.9	89.2	49.0	49.0
France	89.9	89.9	86.1	77.4	77.4
Germany	78.6	100.0	77.6	61.0	77.6
Greece	76.2	100.0	86.5	65.9	86.5
Israel	48.8 **	48.8 **	99.6 **	48.6 **	48.6 **
Italy	70.3	70.9	95.1	66.9	67.5
Lithuania	100.0	100.0	92.1	92.1	92.1
Russian Federation	97.6	99.4	96.5	94.2	95.9
Slovenia	45.6	45.6	93.0	42.4	42.4
Sweden	95.3	95.3	92.9	88.6	88.6
Switzerland	99.0	99.0	88.2	87.4	87.4
United States	75.7	84.7	79.6	60.2	67.4

** Unweighted participation rates.

Table 2.11 Participation Rates – Physics Final Year of Secondary School

Country	School Participation		Student Participation (Weighted Percentage)	Overall Participation	
	School Participation Before Replacement (Weighted Percentage)	School Participation After Replacement (Weighted Percentage)		Overall Participation Before Replacement (Weighted Percentage)	Overall Participation After Replacement (Weighted Percentage)
Australia	63.2	63.9	84.9	53.7	54.2
Austria	36.7	95.5	84.6	31.0	80.8
Canada	79.7	80.2	91.0	72.6	73.0
Cyprus	100.0	100.0	96.0	96.0	96.0
Czech Republic	100.0	100.0	92.1	92.1	92.1
Denmark	54.9	54.9	86.1	47.3	47.3
France	89.9	89.9	86.1	77.4	77.4
Germany	76.8	100.0	81.7	62.7	81.7
Greece	76.2	100.0	86.5	65.9	86.5
Israel	48.8 **	48.8 **	99.6 **	48.6 **	48.6 **
Italy	69.3	69.3	96.6	67.0	67.0
Latvia (LSS)	84.4	84.4	90.8	76.6	76.6
Norway	77.7	94.3	88.0	68.4	83.0
Russian Federation	97.6	98.8	96.2	93.9	95.1
Slovenia	45.6	45.6	94.2	43.0	43.0
Sweden	95.3	95.3	92.9	88.6	88.6
Switzerland	99.0	99.0	88.2	87.4	87.4
United States	77.0	84.3	80.3	61.8	67.7

** Unweighted participation rates.

Because population coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.

2.9 COMPLIANCE WITH SAMPLING GUIDELINES

Figures 2.2, 2.3, and 2.4 show how countries have been grouped in tables reporting achievement results for mathematics and science literacy, advanced mathematics, and physics, respectively. Countries that complied with the TIMSS guidelines for school and student sampling, and that achieved acceptable participation rates – 85 percent of both the schools and students, or a combined rate (the product of school and student participation) of 75 percent – with or without replacement schools, are shown in the first panel. Countries that met the guidelines only after including replacement schools are so noted.

Countries that did not reach at least 50 percent school participation without the use of replacements schools, or that failed to reach the sampling participation standard even with their use, are shown in the second panel of Figures 2.2 - 2.4. Countries that did not meet the guidelines for student sampling are shown in the third panel, and countries that met neither these requirements nor participation rate requirements are shown in the bottom panel. Unweighted results only are included for Israel[1] because Israel had difficulties meeting several sampling guidelines.

[1] This is effectively implemented by assigning a weight of 1 to all students in the sample for Israel.

Figure 2.2 Countries Grouped for Reporting Achievement According to Their Compliance with Guidelines for Sample Implementation and Participation Rates Mathematics and Science Literacy - Final Year of Secondary School

Countries satisfying guidelines for sample participation rates and sampling procedures	
[2] Cyprus	† New Zealand
Czech Republic	[2] Russian Federation
Hungary	Sweden
[1] Lithuania	Switzerland
Countries not satisfying guidelines for sample participation rates	
Australia	Iceland
[2] Austria	[1] Italy
Canada	Norway
France	United States
Countries with unapproved student sampling	
† Germany	
Countries with unapproved sampling procedures and low participation rates	
Denmark	Slovenia
[2] Netherlands	South Africa

† Met guidelines for sample participation rates only after replacement schools were included.
[1] National Desired Population does not cover all of International Desired Population.
[2] National Defined Population covers less than 90 percent of National Desired Population.

Figure 2.3 Countries Grouped for Reporting Achievement According to Their Compliance with Guidelines for Sample Implementation and Participation Rates Advanced Mathematics – Final Year of Secondary School

Countries satisfying guidelines for sample participation rates and sampling procedures	
Canada	† Greece
2 Cyprus	1 Lithuania
Czech Republic	2 Russian Federation
France	Sweden
† Germany	Switzerland
Countries not satisfying guidelines for sample participation rates	
Australia	1 Italy
2 Austria	United States
Countries with unapproved sampling procedures and low participation rates	
Denmark	Slovenia

† Met guidelines for sample participation rates only after replacement schools were included.
1 National Desired Population does not cover all of International Desired Population.
2 National Defined Population covers less than 90 percent of National Desired Population.

Figure 2.4 Countries Grouped for Reporting Achievement According to Their Compliance with Guidelines for Sample Implementation and Participation Rates Physics – Final Year of Secondary School

Countries satisfying guidelines for sample participation rates and sampling procedures	
Canada	[1] Latvia (LSS)
[2] Cyprus	† Norway
Czech Republic	[2] Russian Federation
France	Sweden
† Germany	Switzerland
† Greece	

Countries not satisfying guidelines for sample participation rates	
Australia	United States
[2] Austria	

Countries with unapproved sampling procedures and low participation rates	
Denmark	Slovenia

† Met guidelines for sample participation rates only after replacement schools were included.
[1] National Desired Population does not cover all of International Desired Population.
[2] National Defined Population covers less than 90 percent of National Desired Population.

2.10 SAMPLING WEIGHTS

Appropriate estimation of population characteristics based on the TIMSS samples requires that the TIMSS sample design be taken into account in all analyses. This is accomplished in part by assigning a weight[2] to each respondent, where the sampling weight properly accounts for the sample design, takes into account any stratification or disproportional sampling of subgroups, and includes adjustments for non-response.[3]

The students within each country were selected using probability sampling. A consequence of this is that each student had a known probability of selection. The inverse of this selection probability is the sampling weight. In a properly selected and weighted sample, the sum of the weights for the sample approximates the size of the population. In TIMSS, the sum of the sampling weights for a country sample is an estimate of the size of the population of students within the country in the sampled grade(s). The sampling weights must be used whenever population estimates are required. The use of

[2] The computation of sampling weights is described in Chapter 4.

[3] Sampling weights can be computed only when the probability of selection is known for all students.

the appropriate sampling weights ensures that the subgroups that constitute the sample are properly and proportionally represented in the computation of population estimates.

Tables 2.12 presents the sample sizes and the estimate of the population size (sum of the weights) for the entire final-year sample, for advanced mathematics students, and for physics students, respectively, for each participating country.

Table 2.12 Sample Information

Country	Students in their Final Year of Secondary School		Students Having Taken Advanced Mathematics		Students Having Taken Physics	
	Sample Size	Estimated Population Size	Sample Size	Estimated Population Size	Sample Size	Estimated Population Size
Australia	1941	170847	645	39498	661	31619
Austria	1962	70602	782	31063	777	30795
Canada	5232	263241	2781	58606	2367	51179
Cyprus	534	4556	391	837	368	837
Czech Republic	2167	137459	1101	19446	1087	19428
Denmark	2714	37872	1388	13527	654	2073
France	1590	637935	1071	151531	1110	151531
Germany	2289	967705	2296	262789	723	87888
Greece	-	-	456	14620	459	14668
Hungary	5091	111281	-	-	-	-
Iceland	1703	2308	-	-	-	-
Italy	1616	380834	398	104477		
Latvia (LSS)	-	-	-	-	708	979
Lithuania	2887	22161	734	1360	-	-
Netherlands	1470	145916	-	-	-	-
New Zealand	1763	37549	-	-	-	-
Norway	2518	43806	-	-	1048	4369
Russian Federation	2289	1031187	1638	42858	1233	32975
Slovenia	1622	26644	1536	22881	747	11706
South Africa	2757	374618	-	-	-	-
Sweden	3068	71243	1001	16408	1012	16459
Switzerland	3308	65140	1404	11343	1371	11276
United States	5807	2278258	2785	496852	3114	522784

REFERENCES

Adams R.J. and Gonzalez, E.J. (1996). The TIMSS test design. In Martin M.O. and Kelly, D.L. (Eds.), *Third International Mathematics and Science Study technical report, volume I: Design and development*. Chestnut Hill, MA: Boston College.

Foy, P., Rust, K., and Schleicher, A. (1996). Sample design. In M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study technical report, volume I: Design and development*. Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., and Smith, T.A. (1998). *Mathematics and science achievement in the final year of secondary school: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College.

Schleicher, A. and Siniscalco, M.T. (1996). Field Operations. In M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study technical report, volume I: Design and development*. Chestnut Hill, MA: Boston College.

3 Data Management and Construction of the TIMSS Database

Heiko Sibberns
Dirk Hastedt
Michael Bruneforth
Knut Schwippert
IEA Data Processing Center

Eugenio J. Gonzalez
Boston College

The procedures for processing the TIMSS data for Population 3 (final year of secondary school) were similar to those developed for Populations 1 and 2 (third and fourth grades, and seventh and eighth grades in most countries, respectively) and described in the *TIMSS Technical Report, Volume II* (Martin and Kelly, 1997). In accordance with the TIMSS reporting schedule, the Population 3 data were processed after the data for TIMSS Populations 1 and 2 were completed. The main procedural differences resulted from the absence of teacher questionnaires for the final-year population and that there were three achievement tests: (1) mathematics and science literacy; (2) advanced mathematics; and (3) physics. Otherwise, the data structure for Population 3 was identical to that of the younger populations.

The TIMSS data were processed through a closely cooperative procedure involving the TIMSS International Study Center at Boston College, the IEA Data Processing Center, the Australian Council for Educational Research, Statistics Canada, and the national research centers of the participating countries. Under the general direction of the International Study Center, each institution was responsible for specific aspects of the data processing.

The data processing consisted of six general tasks: data entry, creation of the international database, calculation of sampling weights, scaling of achievement data, analysis of the background data, and creation of the reporting tables. While each task is crucial to ensuring the quality and accuracy of the results, data entry and the creation of the international database take center stage, since those tasks feed into the remaining four. The scaling of the TIMSS Population 3 data is discussed in Chapter 7, the weighting procedures in Chapter 4, and the analysis and reporting in Chapter 8. This chapter describes the process followed in data entry, the creation of the international database, and the steps taken to ensure the quality and accuracy of the international database. It also describes the responsibilities of each participant in creating the international database. In particular, this chapter outlines the flow of the data files among the centers involved in the data processing; the structure of the data files submitted by each country for processing, and the resulting files that are part of the international database; the rules, methods, and procedures used for data verification and manipulation; the data products created during data cleaning and provided to the national centers; and the computer software used in that process.

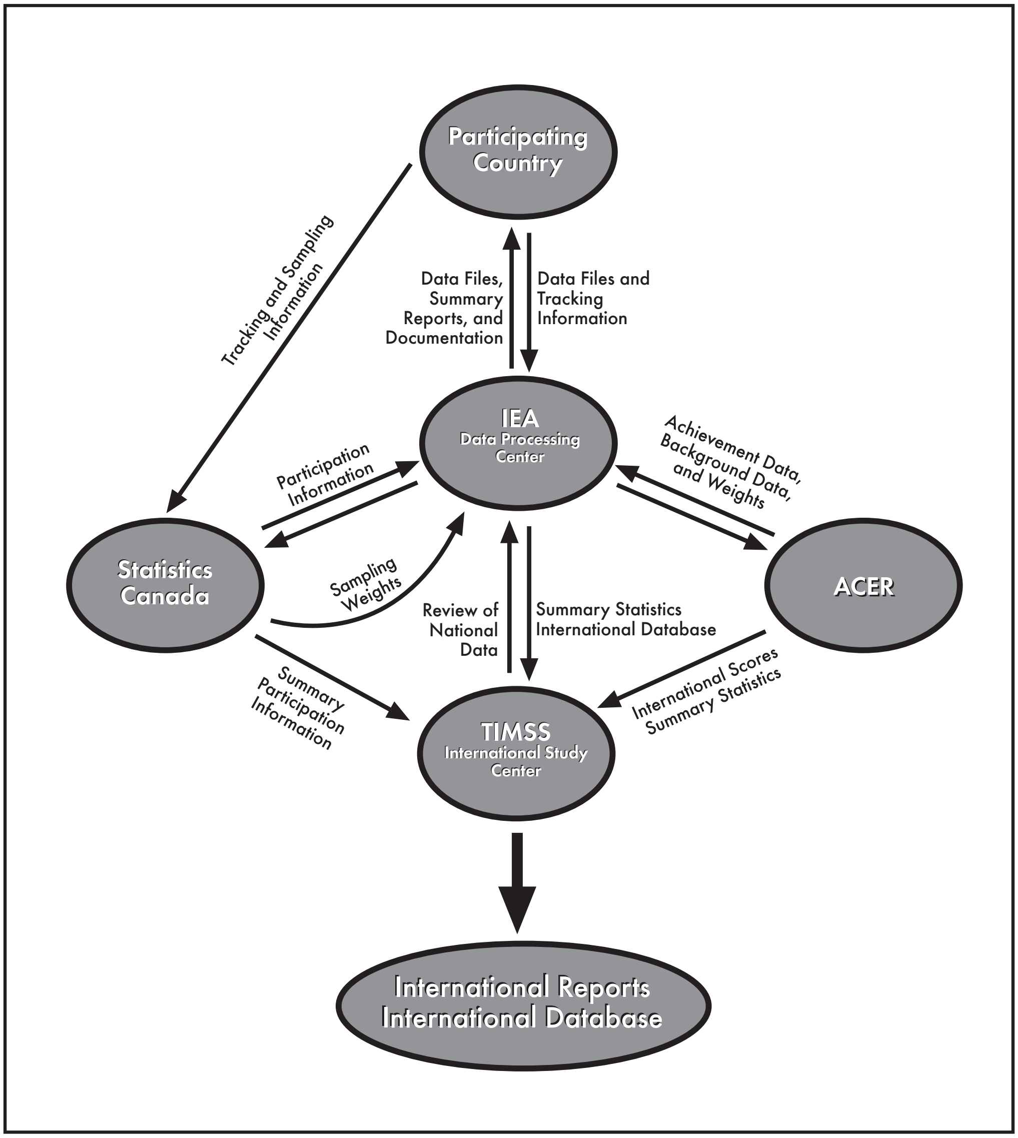
The TIMSS international database for Populations 1 and 2 was released for public use in September 1997, and that for Population 3 in June 1998. Both are available at the TIMSS website (http://www.csteep.bc.edu./timss) and through IEA Headquarters. The databases are each accompanied by a User's Guide and full documentation (see Gonzalez and Smith, 1997 and Gonzalez, Smith, and Sibberns, 1998).

3.1 DATA FLOW

The data collected with the TIMSS survey instruments were entered into data files of a common international format at the national research centers of the participating countries. These data files were then submitted to the IEA Data Processing Center for cleaning and verification. The main responsibilities of the IEA Data Processing Center at this point were to check that the data files submitted matched the international standard and to make modifications where necessary, apply standard cleaning rules to the data to verify their consistency and accuracy, interact with the National Research Coordinators (NRCs) to ensure the accuracy of the data contained in the files, produce summary statistics of the background and achievement data for review by the TIMSS International Study Center; and finally, upon feedback from the individual countries and the TIMSS International Study Center, to construct the international database. The IEA Data Processing Center was also responsible for distributing the national data files to each of the participating countries.

Once verified and in the international file format, the achievement data were sent to the Australian Council for Educational Research (ACER), where basic item statistics were produced for item review and an initial country-level scaling was conducted. At this time, the staff at the TIMSS International Study Center undertook a thorough review of the item statistics (see Chapter 6). At the same time Statistics Canada received from the IEA Data Processing Center data files containing participation information for students in the sample. This information, together with information provided by the NRC, was used by Statistics Canada to calculate sampling weights, population coverage, and participation rates at the school and student level. The sampling weights were then sent to the TIMSS International Study Center for verification and forwarded to ACER to be used in the scaling. When the review of the item statistics was completed and the IEA Data Processing Center had updated the database accordingly, the revised data files were sent to ACER. ACER was then responsible for computing the international item difficulties and for scoring individual students on the international scales. Once the sampling weights and international scale scores were verified at the TIMSS International Study Center, they were sent to the IEA Data Processing Center for inclusion in the international database and distributed to the national research centers. The International Study Center prepared the international report tables and published the reports of the study results. Figure 3.1 is a pictorial representation of the flow of the data files.

Figure 3.1 Flow of TIMSS Data Files



A very important part of the data processing was the interaction of the staff at the TIMSS International Study Center, the staff at the IEA Data Processing Center, and the National Research Coordinators. At specific stages of data verification, the IEA Data Processing Center returned countries' data files for checking. These data files were accompanied by computer printouts with summary statistics to be reviewed by the NRC, together with specific questions pertaining to the data.

3.2 DATA ENTRY AT THE NATIONAL RESEARCH CENTERS

Each TIMSS national research center was responsible for transforming the information from the achievement booklets and questionnaires into computer data files. Participating countries were provided with data-entry software adapted specifically for the purpose of TIMSS, DATAENTRYMANAGER (DEM), together with codebooks for data entry. The codebooks contained information about the variable names used for each variable in the survey instruments, and about field length, field location, labels, valid ranges, default values, and missing codes. The codebooks were designed to be used with DEM in the data entry process. Although DEM was the recommended software, some of the participating countries elected to use a different data entry system. Data files were accepted from the countries provided they conformed to the parameters set in the international codebooks. In order to facilitate data entry, the codebooks and data files were structured to match the test instruments and questionnaires. This meant that there was a data file for each survey instrument. Each country was responsible for submitting four data files if participating in the testing of students in the final year of secondary school: Student Background, Achievement, Coding Reliability, and School Background. Each file had its own codebook.

Although generally collected during the same session, the student background data were entered separately from the student achievement data because the tests and questionnaires were administered as separate instruments. This was done to prevent students from looking back or ahead at their work in the achievement booklet and, most important, because the open-ended achievement items had to be scored following administration. Setting the system to enter the student background data in a file separate from the achievement data allowed the data manager of each country to start entering student background data without having to wait for scoring to finish.

The Student Background data file contains one record for each student in the sample. Entries were made in this file even if the student was excluded from the testing session. This file was used to record the information given by the students in the student questionnaire and other information on identification, participation, and sampling.

The Achievement data file contains one record for each student who was administered a test booklet. A record was also created for any student whose booklet was lost, but not for students who did not respond to the written assessment. The necessary information for these students was contained in the Student Background data file.

In order to check the reliability of the free-response item coding, the free-response items in a random sample of 10 percent of booklets were coded independently by a second coder. The Coding Reliability file contains one record for each student whose responses to the free-response items were coded by a second coder.

The School Background data file contains one record for each originally sampled school, whether the school participated in the survey or not. They also contain records for schools that participated as replacement schools. This file was used to register the information from the school questionnaire and on the participation status of schools.

Table 3.1 presents the total number of files and records of each type received from the participating countries.

Table 3.1 Population 3 Data Files Received by the IEA Data Processing Center

File	Files	Observations
Achievement	23	99107
Student Background	23	99107
Coding Reliability	14	9337
School Background	23	1282

In addition to the data files, countries were also required to submit supporting documentation of their field procedures and copies of their national instruments (translated tests and questionnaires). The documentation included a report of their survey activities, a series of data management forms with clear indications of any changes made in the survey instruments or the structure of the database, and copies of all sampling and tracking forms. These materials were archived at the IEA Data Processing Center and kept for reference purposes during data processing.

Each country was provided with a program called LINKCHK to carry out checks on the data files before submitting them to the IEA Data Processing Center. The program was designed to help NRCs perform an initial check of the system of student, teacher, and school identification numbers after data entry, both within and between files. The reports produced by the LINKCHK program allowed countries to correct problems in the identification system before transferring the data to the IEA Data Processing Center.

3.3 DATA CLEANING AT THE IEA DATA PROCESSING CENTER

Once the data were entered into data files at the national research center, the data files were submitted to the IEA Data Processing Center for checking and input into the international database. This process is generally referred to as data cleaning. The goals of the TIMSS data cleaning were to identify, document, and, where necessary and possible, correct deviations from the international file structure, and to correct key punch errors, systematic deviations from the international data formats, problems in linking observations between files, inconsistent tracking information between and within files, and inconsistencies within and across observations. The main objective of the process was to ensure that the data adhered to international formats and reflected accurately and consistently the information collected within each country.

Data cleaning involved several steps. Some of these were repeated in an iterative fashion until satisfactory results were achieved. During the first step of data cleaning, all incoming data files were checked and reformatted if necessary so that their file structure conformed to the international format. As a second step, all problems with identification variables, linkage across files, codes used for different groups of variables, and participation status were detected and corrected. The distribution for each variable was examined, with particular attention to variables that presented implausible or inconsistent distributions based on the information from the country involved.

During this stage, a series of data summary reports was generated for each country. The reports contained listings of codes used for each variable and pointed to outliers and changes in the structure of the data file. They also contained univariate statistics.

The reports were sent to each participating country, and the NRC was asked to review the data and advise how to best resolve inconsistencies. In many cases the NRC was obliged to go back to the original booklets from which the data had been entered initially.

During the data cleaning process two main procedures were used to make necessary changes in the data. Inconsistencies that could unambiguously be solved were corrected automatically by a data cleaning program. Errors that could not be solved using standard cleaning routines had to be solved case by case by the DPC staff. In either case, all changes made in the data were documented. A database was created in which each change was recorded, and it was possible to reconstruct the original database received from a country.

In the following section each of the steps mentioned above is described in more detail.

3.3.1 Standardization of National File Structure

The first step in the data processing at the international level was to verify the compatibility of the national datasets with the international file structure as defined in the TIMSS international codebook. This was necessary before the standard cleaning with the Data Processing Center cleaning software could be performed.

Although the TIMSS international codebooks distributed with the data entry software gave clear and detailed instructions about the structure and format of the files each country was to submit to the IEA Data Processing Center, some countries opted to enter and submit their data files in other formats, using structures different from the international standard. For the most part, these differences were due to specific national circumstances.

The *TIMSS Guide to Checking, Coding, and Entering TIMSS Data* (TIMSS, 1995) asked countries to prepare and send their data files using the DEM software, which produces an extended dBase format. Some data files, however, were received in ASCII fixed format (raw data), SPSS format, and SAS format.

After the national files were converted into the extended dBase format, the structure of the files was inspected and deviations from the international file structure were identified. A standard software tool automatically scanned the file structure of the country files and reported the following deviations:

- International variables dropped
- National variables added
- Different variable length or number of decimal positions
- Different coding schemes or out of range values
- Specific national variables
- Gang-punched variables

Together with the inspection of the national data files, the data management and tracking forms submitted by each NRC were reviewed. As a result of this initial review, the Data Processing Center outlined and implemented necessary changes in the national data to make the files compatible with the international format. In most cases programs had to be prepared to fit the file structures and specificities of each country.

As part of the file standardization process files were merged where applicable (for example, the Student Background and the Achievement data files) and the file structure was changed to facilitate data analysis, since direct correspondence to the instruments was no longer necessary. The changes made in the files during the cleaning process are summarized below. In general, variables created during data entry for verification were purposely dropped from all files, and new variables were added (e.g., reporting variables, derived variables, sampling weights, and achievement scores). What follows is a brief description of the changes performed in the files received from the countries.

3.3.1.1 Student Background File

Several new variables were added to the beginning of each record to represent students' participation status in the two testing sessions and in completing the student background questionnaire. The students' ages computed from the date of testing and the date of birth were also added to the files, as were sampling weights and several achievement scores for advanced mathematics, physics, and mathematics and science literacy.

3.3.1.2 Achievement File

The structure of the Achievement files produced by each country reflected the structure of the nine test booklets. During data entry, once the version of the booklet was indicated, the data software displayed only the variables representing the items in that particular booklet. A variable was created for each item in a booklet, and the order of these variables reflected the order of the items within a booklet. This kept data entry and programming of the data entry software to a simple and rectangular structure. However, it also meant that a lot of redundant variables were created during data entry, since an item administered in more than one booklet was coded as a different variable for each booklet in which it occurred. A useful feature of the redundancy is that it allowed the student's booklet to be identified easily even if there was a keypunch error when the identification of the booklet was entered.

After final cleaning, the Achievement files were restructured so that each item appeared in just one location in the student records, regardless of the test booklet it came from. This new structure reflects the item clusters used to assemble the booklets (Adams and Gonzalez, 1996) rather than not the booklet layout. The variables for the items that were not administered to the student were coded as "not administered." The structure of the Achievement file is presented schematically in Figure 3.2.

Figure 3.2 Revised Structure of the Student Achievement File

IDs	Tracking	Achievement Item Cluster					
		A	B	C	...	K	L

3.3.1.3 Coding Reliability File

The structure of the Coding Reliability file prepared in each country also mirrored the structure of the nine test booklets. Again, a variable was created for each free-response item in a booklet, and the order reflected the order of appearance of the items within the booklets. In the final international data file the variables were rearranged so that each item was represented by only one variable regardless of the booklet in which it appears. All other variables representing items not included in the booklet administered to the student were coded as "not administered."

The final international version of the Coding Reliability file includes both the data from the 10 percent sample of students selected for reliability coding and the original data for these students. This enables the user of the file to compare the codes without having to merge any files. A third set of variables was included in the final international version of the file to reflect the agreement between the two codes assigned to the answers to the free-response items.

3.3.1.4 School Background File

The file structure of the cleaned school data sets in the international database is identical to the structure used for data entry. The file includes a School Identification number (ID) block and the variables in order of their appearance in the school questionnaire. The major change was that school weighting variables and student achievement scores aggregated on the school level were added to the school file.

3.3.2 Standard Cleaning

After the data received from the countries were transformed into the international format, a set of standard cleaning rules was applied to each data file. These rules were applied using software the IEA Data Processing Center developed to identify and in many cases correct inconsistencies in the data. Some inconsistencies could not be solved automatically but had to be reviewed carefully and appropriate corrections devised.

In particular, the following problems were sought and corrected whenever possible (for further details, please refer to Jungclaus and Bruneforth (1996)):

- Problems with identification, tracking, and other indicator variables
- Problems with split variables, i.e. variables where respondents were allowed to check more than one option

- Problems with the variable indicating the achievement booklet administered to the student
- Problems with filter and dependent questions

After as many problems as possible were solved at the IEA Data Processing Center (by reviewing the instruments and national documentation or by applying the cleaning rules), the Data Processing Center cleaning software was used a second time to create a report of remaining data problems. These reports were summarized and sent to the NRCs with specific questions and, in some cases, suggestions for resolution.

3.3.3 Item Cleaning

After applying the cleaning rules described above, the achievement data underwent a careful and detailed review.

For this purpose, an item analysis was performed using the item analysis software QUEST developed by ACER (Adams and Khoo, 1993). National scores in mathematics, physics, and literacy based on the Rasch model were calculated and several reports were generated with these data. Some data problems, such as items with inadvertent changes in the coding scheme or switched response options, were detected and corrected at this point. Reports with summary item statistics were sent to the NRCs for their review.

The coding reliability data were compared with the achievement data. For this purpose, the percentage of agreement between the codes assigned by the two coders was calculated on two levels: agreement between the number of score points assigned to an item and agreement on the two-digit diagnostic code.

After this initial review by the IEA Data Processing Center, reports were generated with item statistics. The TIMSS International Study Center used these reports to conduct a thorough review of the achievement item data. Details of this process are presented in Chapter 6 of this report.

3.3.4 Country-Specific Cleaning

Some of the anomalies detected by the checking procedure had to be solved case by case. During this process, it was important to find individual solutions that followed general guidelines, so that the solutions could be uniformly applied to similar problems in other countries.

The corrections made in this cleaning step were based on the NRCs' review of the preliminary statistics from the IEA Data Processing Center, the NRC field operations reports and instruments sent with the data, and the NRCs' comments on the data almanacs produced by the TIMSS International Study Center. In particular, the following steps were performed on a country-by-country basis to correct the data:

- Correcting switched options/categories in categorical background variables
- Deleting data entered for questions that were not included in the international versions of the questionnaires
- Deleting data entered in error
- Collapsing categories to match the international coding scheme
- Deleting data not internationally comparable due to translation problems
- Copying data from one observation to another if the information requested was identical for both observations
- Adding dummy records to the files to ensure correct linkage across files

None of these steps were performed without the cooperation of the NRCs, who had to confirm or reject the suggested data changes. More important, in many cases they had to give detailed advice about the changes to be made in the coding scheme.

3.3.5 Other General Cleaning

After transforming the data files into the international format, performing the standard cleaning on them, and reviewing the achievement data, two other checks were made: statistical checks and consistency checks.

3.3.5.1 Statistical Checks

Statistical checks were designed to find outliers for continuous variables, variables with very high percentages of missing values, and categorical variables with different numbers of options than the international version of the instruments. Statistical checks were performed separately for each country. As preparatory steps, descriptive statistics were computed for each variable within each country and these statistics were stored in a database. The information compiled in this way was used as outlined below.

Outlier Detection

In order to check variables for extreme values, an outlier was defined as a value in a variable that is over 5 standard deviations above the mean for that variable, or with a value twice as large as the 90th percentile for the variable. Any such variables detected were carefully examined.

For some of the variables found by this procedure (e.g., number of students in a school), additional information was used to judge the plausibility of the detected outlying values. If the file contained obvious miskeys, the variable was coded to “Invalid.” Cases that could not be resolved at the Data Processing Center were reported to NRCs and treated according to their suggestions.

High Percentages of Missing Observations

Variables were flagged for investigation if more than 99 percent of the cases had missing values. If such a variable was detected, the corresponding question in the questionnaire was examined. Often in such cases the question was not completed by the respondents because it was not applicable. In some cases this was due to a data entry error.

Additional Response Options for Categorical Variables

The observed values for categorical variables were compared with the valid codes specified by the international codebook. If additional codes were found, the corresponding question in the questionnaire was examined. It was possible that the additional code was due to key-punch error during data entry. Where it was determined that this was the case, the corresponding categories were recoded to "Invalid." If, on the other hand, the question that was asked allowed additional categories, the NRCs were asked to help find a way to make the new code internationally comparable. If recoding was possible, the original value for the variable was kept in a separate country-specific variable. If it was not possible to recode to meet the international coding scheme, the original data were kept in a separate variable and the international variable was coded to an explicit missing code.

Response Options with a Frequency of Zero in Categorical Variables

If a frequency of zero was detected for an option of a categorical variable, the corresponding question in the questionnaire was checked as a precaution. If a category in the original version of the question was missing, the NRC was contacted to verify that the correct categories were retained. However, if the category was not missing in the questionnaire but was not checked by any respondent, the data were not changed. Quite often, variables belonging to groups of questions had zero frequencies for one or more of the categories. For example, the school questionnaire asked for the frequency of different types of student behavior in schools. Some forms of behavior did not happen often; thus the corresponding categories had a frequency of zero.

3.3.5.2 Consistency Checks

Consistency checks dealt with problems that were discovered in the first phase of the cleaning process, but not corrected at that time because information about the problems across countries was needed to decide on the rules to be applied. The following sections describe the checks applied to all countries and the inconsistencies that were corrected.

Student's Gender, Date of Birth, Age, and Date of Testing

If a student's sex as reported in the background questionnaire differed from that in the tracking information, the tracking version was replaced by the background questionnaire version. The same substitution procedure was followed with regard to students' dates of birth. If the date of testing was missing, it was replaced by the modal value of the student's class when available.

School File

The questions concerning the same course of instruction were checked for consistent answers. If all students followed the same course of instruction (filter = Yes) and the majority of answers was consistent with the filter, all answers in the "No" list were recoded to "Not applicable." If, on the other hand, valid answers could be found in the "No" list and only missing values could be found in the "Yes" list, the filter was changed to "No." Uncertain cases were reported and recoded directly if possible. Sometimes the appropriate response could be deduced from the answering pattern found in the data.

3.4 DATA PRODUCTS

3.4.1 Data Almanacs

Together with their data files, each country received data almanacs produced by the TIMSS International Study Center that contained weighted summary statistics, for each participating country, on each variable included in the survey instruments. There were two types of display. The display for categorical variables included an estimate of the size of the student population, the sample size, the weighted percentage of students who were not administered the question, the percentage of students choosing each of the options on the question, and the percentage of students who did not choose any of the valid options. The percentage of students to whom the question did not apply was also presented in the almanac. For continuous variables the display included an estimate of the size of the student population, the sample size, the weighted percentage of students who were not administered the question, the percentage who did not respond, the percentage to whom the question did not apply, the mean, mode, minimum, maximum, and the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles. An example of such data displays is presented in Figures 3.3 and 3.4. These data almanacs were sent to each of the participating countries for review. When necessary, they were accompanied by specific questions about the data presented in them. These almanacs were also used by the TIMSS International Study Center during the data review and in the production of the reporting tables.

Figure 3.3 Example Data Almanac Display for a Categorical Student Background Variable

1Third International Mathematics and Science Study - 1995 Assessment May 29, 1998 61
Student Background Variables - Students in the Final Year of Secondary School (INTMSL4=1)
Question: Do you intend to attend a four-year college or university? (CSBGFED4)
Location: SQ3-14A4

3.POPULATION 3
GEN\CONT EDUC\ATTEND <4 YR COLLEGE U>

Country	Population	Sample	%NA	1.YES %	2.NO %	NOT APP. %	Other %
Australia	170847	1941	2.0	65.7	25.5	8.8	2.9
Austria	70602	1962	3.7	32.1	41.1	26.7	4.4
Canada	263241	5232	0.7	62.5	33.7	3.7	5.3
Cyprus	4556	534	0.4	61.4	21.3	17.3	7.0
Czech Republic	137459	2167	0.4	24.6	35.8	39.6	0.8
Denmark	37872	2714	9.5	38.5	35.8	25.7	12.8
France	637935	1590	0.5	26.8	62.7	10.5	2.5
Germany	967705	2289	73.2	62.1	34.5	3.4	74.2
Hungary	111281	5091	3.2	9.3	74.0	16.6	12.0
Iceland	2308	1703	1.9	65.4	28.1	6.5	5.9
Israel	1357	1357	10.1	64.4	28.4	7.2	19.4
Italy	380834	1616	0.8	44.2	23.2	32.6	7.0
Lithuania	22161	2887	2.1	50.5	36.2	13.3	8.3
Netherlands	145916	1470	2.2	17.1	60.6	22.3	3.6
New Zealand	37549	1763	0.7	53.0	37.7	9.3	2.8
Norway	43806	2518	1.5	43.0	46.1	10.9	6.1
Russian Federation	1031187	2289	0.3	60.6	38.0	1.4	6.0
Slovenia	26644	1622	2.7	62.5	25.9	11.6	5.4
South Africa	374618	2757	2.1	75.2	19.1	5.7	8.8
Sweden	71243	3068	2.0	43.8	40.7	15.5	5.9
Switzerland	65140	3308	0.6	19.4	50.4	30.1	1.9
United States	2278258	5807	2.8	69.3	26.8	3.9	7.0

Figure 3.4 Example Data Almanac Display for a Numerical School Background Variable

```
Third International Mathematics and Science Study - 1995 Assessment                May 29, 1998  20
School Background Variables - Students in the Final Year of Secondary School (INTMSL4=1)

Question: How many classroom teachers are on the staff of your school?  (FTE) (CCBGFTE4)
Location: SCQ3-3D

*POPULATION ID*=3.POPULATION 3
```

Country	Population	Cases	%Not Ad.	%Omit	%Not Ap.	Mean	Mode	Min	P5	P10	Q1	Median	Q3	P90	P95	Max
Australia	170847	86	4.9	2.3	0.0	53.6	43.0	19.0	27.4	27.4	36.0	50.0	60.0	80.0	85.3	180.0
Austria	70602	169	8.5	3.0	0.0	45.0	45.0	4.0	13.3	23.0	35.0	59.4	86.0	110.0	.	385.0
Canada	263228	336	17.9	4.7	0.0	59.6	52.0	2.0	16.0	23.7	38.5	56.2	76.9	99.3	120.0	153.0
Cyprus	4556	28	12.9	0.0	0.0	52.6	29.0	20.0	20.0	29.0	41.7	51.4	61.0	75.0	76.0	113.0
Czech Republic	137459	150	0.2	1.5	0.0	29.3	50.0	1.0	9.0	10.5	19.0	28.1	37.3	49.0	53.0	91.0
Denmark	29095	91	0.8	1.8	0.0	60.4	46.0	6.5	28.0	36.0	52.0	62.0	73.0	80.0	85.0	110.0
France	637935	56	4.5	8.5	0.0	116.5	35.0	10.0	18.0	38.0	79.0	113.0	134.0	190.0	260.0	260.0
Germany	967705	152	35.3	3.7	0.0	56.5	48.0	2.0	15.0	25.0	34.0	53.0	75.0	103.0	106.0	112.0
Hungary	111281	204	15.0	5.8	0.5	47.4	35.0	1.0	15.5	21.0	34.3	46.0	60.5	76.5	82.5	94.0
Iceland	2308	30	5.7	10.2	0.0	43.1	8.5	8.0	8.5	10.0	33.2	44.6	55.0	59.0	77.5	77.5
Israel	1357	52	8.8	15.9	0.0	103.2	70.0	8.0	34.0	59.0	90.0	120.0	180.0	300.0	.	300.0
Italy	380834	101	1.4	8.9	0.0	78.3	69.0	9.0	15.0	30.0	48.0	73.0	107.0	130.0	154.0	190.0
Lithuania	22161	142	24.4	4.6	0.0	62.7	30.0	15.0	30.0	30.0	47.0	63.0	75.0	91.0	108.0	145.0
Netherlands	145916	79	100.0	0.0	0.0	.	.	.	.	.	.	.	.	.	.	.
New Zealand	37549	79	4.5	0.0	0.0	117.3	999.0	9.6	24.0	27.8	36.0	48.0	64.0	89.7	999.0	999.0
Norway	43806	131	9.1	3.1	0.0	58.2	40.0	16.0	23.0	28.0	40.0	51.8	70.0	97.0	120.0	160.0
Russian Federation	1031187	163	0.0	8.4	0.0	54.1	65.0	7.0	19.0	22.0	35.0	52.0	71.0	87.0	96.0	132.0
Slovenia	26644	79	66.9	2.6	0.0	58.2	33.0	18.0	31.0	32.0	41.0	55.0	74.0	89.0	91.0	152.0
South Africa	374618	90	69.9	0.7	0.0	23.3	16.0	1.0	1.0	12.0	16.0	23.0	28.0	40.0	51.0	51.0
Sweden	62244	126	4.5	13.4	0.0	84.6	70.0	3.0	13.0	37.0	61.0	80.0	110.0	134.0	148.0	190.0
Switzerland	65140	383	17.5	9.6	0.0	53.7	40.0	0.3	7.0	8.0	13.0	39.0	70.0	102.0	200.0	320.0
United States	2278258	211	17.1	5.7	0.0	71.2	42.0	1.0	13.0	19.0	35.5	63.0	93.0	126.0	156.0	263.0

3.4.2 Versions of the National Data Files

Building the international database was an iterative process. The IEA Data Processing Center provided NRCs with a new version of their countries' data files whenever a major step in data processing was completed. This also guaranteed that the NRCs had a chance to review their data and run their own checks to validate the data files.

Three versions of the data files were sent out to the countries before the TIMSS international database was made available. Each country received its own data only. The first version of the data files was sent to the NRC as soon as that country's data had been cleaned. These files contained national Rasch scores calculated by the Data Processing Center. Documentation, with a list of the cleaning checks and all corrections applied to the data, was included to enable the NRC to review the cleaning process. Univariate statistics for the background data and item statistics for the achievement data were also provided for statistical review. A second version of the data files was sent to the NRCs when the weights and the international achievement scores were available and had been merged with the files. A third version of the data was sent together with the data almanacs after final updates had been made, to enable the NRCs to validate the results presented in the first international reports.

3.4.3 Reports

Several reports were produced during data processing at the IEA Data Processing Center to inform and assist the NRCs, the TIMSS International Study Center, and other institutions involved in TIMSS. The NRCs were provided with diagnostic reports and univariate statistics to help them in checking their data. The TIMSS International Study Center and ACER were provided with international item statistics. The International Study Center also received international coding reliability statistics and international univariate statistics. A report was made to the TIMSS International Study Center and the TIMSS Technical Advisory Committee about each country's deviations and cleaning status as well as the major problems encountered during its data cleaning. The report also included general statistics about the number of observations per file and subpopulation and student response rates.

3.5 COMPUTER SOFTWARE

dBase was used as the standard database program for handling the incoming data. Tools for pre-cleaning and programs such as LINKCHCK (described earlier), and MANCORR and CLEAN (described below) were developed using CLIPPER for manipulating data and some data processing. Statistical analyses (e.g., univariate statistics) for data cleaning and review were carried out with SAS. The final data sets were also created using SAS. For item statistics, the Data Processing Center used the QUEST software (Adams and Khoo, 1993).

The main programs that were developed by the Data Processing Center for TIMSS are described below. Most of the programs that were written for country-specific cleaning needs are not listed. The programming resources in the main cleaning process were spent largely in developing this set of programs.

3.5.1 MANCORR

The most time-consuming and error-prone part of data cleaning is the direct or "manual" editing of errors uncovered by the review process. Based on the Data Processing Center's experience in the IEA Reading Literacy Study and the pilot phases of TIMSS, the data editing program MANCORR was developed. It is easy to use and generates automatic reports of all data manipulation. Its main advantage compared with other editors is that all changes in the data are documented in a log database, from which reports can be generated. As updated data were received from countries, the time-intensive manual changes could be automatically repeated. An "Undo" function allowed the restoration of original values that had been modified with the MANCORR program. The report on which changes were made in the data, by whom, and when was important for internal quality control and review. The MANCORR program was designed using CLIPPER in order to manipulate DATAENTRYMANAGER files.

3.5.2 CLEAN

The central program for data cleaning in TIMSS was the diagnostic program CLEAN, developed with CLIPPER. This program was based on the programs used in the IEA Reading Literacy Study and the TIMSS field tests. It checked all the TIMSS files separately, but also checked the linkages across files and made between-file comparisons. Then corrections were made according to the rules described above. An important feature of the program is that it can be used on a data set as often as necessary. It could first be used to make automatic corrections, and subsequently for creating a report only, without making corrections. Thus it was possible to run a check on the files at all stages of work until the file format was changed to the SAS format. This meant that the program was used not only for initial checks but also to check the work done at the Data Processing Center.

A feature of the TIMSS data cleaning tools is that all deviations are reported to a database, so that reports can be generated by type of problem or by record. Reports previously generated by the program could be compared automatically with newer reports to see which problems had been solved, and even more important, whether additional deviations were introduced during manual correction. Last, the databases (which included all reported deviations) were used to generate the final reports to be sent to the countries. These reports showed which deviations were initially in the data, which were solved automatically, which were solved manually, and which remained unchanged.

3.5.3 Programs Creating Meta Databases

Using SAS, several programs were developed by the Data Processing Center for reviewing and analyzing both the background data and the test items. For the background data, a meta database containing information provided by the initial analysis and by the international codebook was created. Another meta database containing the relevant item parameters was created for the achievement test items. Later, all statistical checks and reports used these databases instead of running the statistics over all data sets again and again. If the data for one country were changed, then statistics had to be recalculated only for this country; the tabulation program, which accessed only the meta database, could then be applied, since the other countries' values remained unchanged. This reduced the computing time for certain procedures from hours to a few minutes. Both databases are the base sources of several reports produced at both the national and international levels (e.g., for the univariate and item analysis reports).

The univariates and item statistics were prepared on a variable-by-country or country-by-variable basis to allow review at the national level and international comparison of individual variables.

3.5.4 Export programs

As mentioned above, SAS was the main program for analyzing the data. Using SAS, export programs were developed and tested to create output data sets for data distribution that are readable by either SAS or SPSS.

3.6 CONCLUSION

The structures and processes designed for the data processing of TIMSS, the largest international empirical educational study ever conducted, successfully met the tremendous challenge. In planning for TIMSS data processing, the major problems were anticipated and provision for dealing with them was incorporated into the data processing system. Even the most complicated school systems were handled adequately by the admittedly complex record identification system. This system had been criticized during the planning phase as too complicated, but it proved to be just barely general enough to identify observations unambiguously and allow the linkage of files in every education system.

The Data Processing Center was closely involved in the planning phase of the study. The study thus benefitted from the Center's knowledge and experience in data processing. For example, it was anticipated that national adaptations and country-specific options would create problems not only during data processing but also in later analysis. Accordingly, international definitions were established that minimized such problems. Most of the problems encountered during data processing arose because countries modified the internationally agreed procedures without notifying the Data Processing Center. The adaptation of record identification systems by some countries (because they felt the international system was too complex) created a lot of unexpected work.

Minor modifications, such as adding new categories to questions, switching the order of options, leaving out international response categories, or changing open-ended questions to multiple-choice questions, were easy to recode to match the international definitions unless countries completely restructured the questionnaires, resulting in the need for additional resources and energy to check and reorganize the data. This shows how important it is in any international study to verify translations of the national questionnaires and to ensure internationally comparable data.

Some problems arose due to communications difficulties. Early and continuous involvement of the data processing staff helped minimize the amount of time and work required, by the countries, the International Study Center, and the Data Processing Center, to produce clean data. It was very important that the data processing staff was easily accessible to the participating countries so that they could get help whenever they had problems. Modern technology, such as the capability to send facsimiles, as well as the Internet, makes the will to communicate, and not the distance between the participants, the most important factor in a successful study. TIMSS demonstrated this with the successful communication between the Data Processing Center in Hamburg, the TIMSS International Study Center at Boston College, Statistics Canada in Ottawa, and the Australian Council of Educational Research in Melbourne. The idea of a decentralized study proved feasible and workable. The time difference between the institutions involved occasionally even helped speed up the work: TIMSS was worked on around the clock.

REFERENCES

Adams, R.J. and Gonzalez, E.J. (1996). The TIMSS test design. In M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study technical report, Volume I: Design and development.* Chestnut Hill, MA: Boston College.

Adams, R.J. and Khoo, S. (1993). *Quest: The interactive test analysis system.* Melbourne: Australian Council for Educational Research.

Gonzalez, E.J. and Smith, T.A., Eds. (1997). *User Guide for the TIMSS international database: Primary and middle school years – 1995 assessment.* Chestnut Hill, MA: Boston College.

Gonzalez, E.J., Smith, T.A., and Sibberns, H (Eds.). (1998). *User guide for the TIMSS international database: Final year of secondary school- 1995 assessment*. Chestnut Hill, MA: Boston College.

Jungclaus, H. and Bruneforth, M. (1996). Data consistency checking across countries. In M.O. Martin and I.V.S. Mullis (Eds.), *Third International Mathematics and Science Study: Quality assurance in data collection.* Chestnut Hill, MA: Boston College.

Martin, M.O. and Kelly, D.L. (Eds.) (1998). *TIMSS technical report, volume II*: *Implementation and analysis - primary and middle school years.* Chestnut Hill, MA: Boston College.

Third International Mathematics and Science Study (TIMSS). (1995). *TIMSS guide to checking, coding, and entering the TIMSS data (Doc. Ref.: ICC918/NRC449).* Chestnut Hill, MA: Boston College.

4 Calculation of Sampling Weights

Jean Dumais
Pierre Foy
Statistics Canada

4.1 INTRODUCTION

This chapter describes the procedures for computing sampling weights for the TIMSS assessment of students in the final year of secondary school (Population 3), in which 24 countries took part. TIMSS Population 3 is defined as all students in their final year of secondary education, that is, all students who upon successful completion of that final school year would either enter the labor market or tertiary education. This definition is meant to be as inclusive as possible.

The TIMSS sampling design was intended to provide estimates of the mathematics and science literacy of all students in the final year of secondary school, while also assessing the advanced mathematics and physics knowledge of students with preparation in these subjects. In addition to characterizing the entire population of final-year students, therefore, the design had to produce accurate estimates of two overlapping subpopulations: students with preparation in advanced mathematics, and students with preparation in physics. In several countries where the overlap was complete (all students that study advanced mathematics also study physics) there were just two groups, those that studied advanced mathematics and physics and those that did not. In countries with clearly defined tracks for upper secondary students, these two groups were often in different schools, which further simplified the sampling procedure. However, in general the situation was more complicated, and a more complex design was required. This design is summarized below, and is described in more detail in Chapter 2.

An essential aspect of the sampling design was that students were stratified according to their level of preparation in mathematics and physics, so that appropriate test booklets could be assigned to them. As described in Chapter 2, each student was characterized as having taken advanced mathematics (M) or not (O), and as having taken physics (P) or not (O). Combining these two-way classifications yields four mutually exclusive and exhaustive categories of students:

OO Students having studied neither advanced mathematics nor physics

OP Students having studied physics but not advanced mathematics

MO Students having studied advanced mathematics but not physics

MP Students having studied both advanced mathematics and physics

In some school systems, students in each group were readily identifiable by virtue of their track assignment or school type, whereas in others it was necessary to categorize individual students in terms of their course-taking history.

Four kinds of student test booklets were assigned to students on the basis of this classification, so that each student completed one 90-minute test booklet. Students classified as OO received either booklet 1A or 1B, the two booklets containing items related to mathematics and science literacy. Students classified as OP received either booklet 1A or 1B, or one of the three booklets containing physics material (2A, 2B, or 2C). Students classified as MO received either booklet 1A or 1B, or one of the three booklets containing advanced mathematics material (3A, 3B, or 3C). Students classified as MP also received one booklet, which could be 1A, 1B, 2A, 2B, 2C, 3A, 3B, 3C or 4. Booklet 4 contained mathematics and science literacy, advanced mathematics, and physics items.

In reporting the achievement of advanced mathematics students, the sample consisted of all MO and MP students that were assigned one of the mathematics booklets (booklets 3A, 3B, or 3C) or the combined literacy, advanced mathematics, and physics booklet (booklet 4). Similarly, the sample of physics students consisted of all OP and MP students that were assigned one of the physics booklets (booklets 2A, 2B, or 2C) or the combined booklet (booklet 4). The sample for reporting on mathematics and science literacy comprised all students in each of the OO, OP, MO, and MP strata that were assigned one of the literacy booklets (booklets 1A or 1B) or the combined booklet (booklet 4).

The basic sample design (intended for use in comprehensive systems where all four kinds of students could be found in all schools) was straightforward. It consisted of a two-stage procedure where schools were sampled with probability proportional to size in the first stage, and an equal number of OO, OP, MO, and MP students was sampled in the second stage. However, implementation varied from country to country, depending on the structure of the education system, and was often quite complex. Some chose to stratify the schools explicitly, others did not; some sampled individual students while others preferred to sample entire classrooms. Details on sampling plans for individual countries are provided in Appendix B of this report. Given the number of variations on the basic design, and the frequent necessity of using different sampling fractions for each student type, the derivation of appropriate sampling weights was a very important step in ensuring the computation of proper survey estimates.

TIMSS made use of item response theory (IRT) methods to derive scales for mathematics and science literacy, advanced mathematics, and physics. The IRT methodology provides an estimate of the proficiency on the scale for each student, even though each student completed only one booklet and hence responded to only part of the assessment item pool. For example, the literacy scale is based on the contents of booklets 1A and 1B, but students in the literacy sample completed just one of these booklets. Most of the international reporting is by scale, so it was necessary to have sampling weights appropriate to this level, i.e., for students that took either booklet 1A or 1B. However, TIMSS also reports student performance on selected individual items, and these are specific to particular booklets. Consequently, it was necessary to compute sampling weights that could be used at the booklet level also, i.e., for the students that completed booklet 1A, or for the students that completed booklet 1B.

4.2 GENERAL WEIGHTING PROCEDURE

Although the basic sampling design specified just a two-stage procedure, since participants could sample either intact classes or individual students it was convenient for computational reasons to think of classes and students as separate sampling stages; and the distribution of booklets at random within schools effectively adds another sampling stage. Computationally, therefore, the sampling weights were assembled in four steps which reflected the multi-stage nature of the sample design. The first step produced a school-level weighting factor. The second step produced a classroom-level weighting factor. The third step produced a student-level weighting factor. The last step produced booklet-level and a scale-level weights. Non-response adjustments were also made to the weighting factors. The overall estimation weight attached to each student record was the product of the four intermediate weights: the school weight, the classroom weight, the student weight, and the scale or booklet weight.

4.2.1 The School Weight

The school weight represents the inverse of the first-stage selection probability of a sampled school. The TIMSS design requires that school selection probabilities be proportional to the school size (PPS), defined as enrollment in the target population. Participants were encouraged to stratify schools explicitly by factors that would improve the precision of the sample or guarantee coverage of special populations. This was in addition to the requirement to stratify students by academic preparation so as to identify the reporting populations. The basic school weight for the *i*th sampled school in a given explicit stratum is thus defined as

$$BW_i^{sc} = \frac{M}{n \times m_i}$$

where n is the number of sampled schools in the stratum, m_i is the measure of size for the *i*th school, and

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

where N is the total number of schools in the stratum.

A few countries opted for simple random sampling of schools rather than PPS; this means that every school has the same unit size ($m_i \equiv 1$) and that M = N.

In two large participating countries (the United States and the Russian Federation) it was necessary to introduce an extra stage of sampling whereby geographical regions were sampled prior to sampling schools. For those countries the basic school weight incorporated a weighting factor to reflect this additional front-end sampling stage.

This weighting factor was calculated in the same way as the school weight since the geographical regions were also sampled with probability proportional to size. The resulting school weight was simply the product of the "region" weight and the school weight as described earlier.

The basic school weight was adjusted to reflect non-response among sampled schools. From the originally selected sample of n schools, occasionally schools were unable or unwilling to take part in the assessment. Whenever possible, these schools were replaced with replacement schools selected at the same time as the originals. In the end, the number of participating schools, n_p say, was sometimes smaller than the planned school sample size. Therefore the basic weight was adjusted to account for the reduction in sample size.

The school-level adjustment for non-response was calculated as follows within each explicit stratum:

$$A^{sc} = \frac{n}{n_p}$$

and the final school weight for the ith school thus becomes

$$FW_i^{sc} = BW_i^{sc} \times A^{sc}.$$

4.2.2 The Classroom Weight

The classroom weight is the inverse of the probability of selection of a sampled classroom within a sampled school. For many of the participants, the classroom weight was irrelevant since students were sampled directly within the school, in accordance with the basic sampling design, rather than via a sampled classroom. In such cases, the classroom weight was simply set at one (1.0). Classroom sampling was used only when all the students in the class belonged to the same sub-population, and consequently classroom weights were calculated independently for each sub-population.

For sub-population g within the ith school, let C_{gi} be the total number of classrooms. In most cases, one classroom only was selected with equal probability, and so the probability of selection was one divided by C_{gi}, and the reciprocal of this probability is the classroom weight. In those schools, the classroom weight assigned to the classroom from sub-population g in the ith school was

$$FW_{gi}^{cl} = C_{gi}$$

In a few instances, countries chose more than one classroom to better represent certain sub-populations. If c_{gi} is the number of classrooms selected at random, then

$$FW_{gi}^{cl} = \frac{C_{gi}}{c_{gi}}.$$

4.2.3 The Student Weight

The student weight is the inverse of the probability that a student within a sampled school or classroom will be sampled for the TIMSS testing. Let the number of enrolled students (after removing students that were out-of-scope or excluded) in school i and sub-population g (perhaps in a classroom) be N_{gi}. If the sample size is n_{gi}, then the basic student weight[1] is

$$BW_{gi}^{st} = \frac{N_{gi}}{n_{gi}}.$$

Occasionally a sampled student did not take part in the assessment, because of absence through illness or for some other reason, and so it was necessary to have a correction for student non-response. If there were r_{gi} students that responded, then the student non-response adjustment is

$$A_{gi}^{st} = \frac{n_{gi}}{r_{gi}}$$

and the final student-level estimation weight is:

$$FW_{gi}^{st} = BW_{gi}^{st} \times A_{gi}^{st}.$$

4.2.4 The Booklet Weights

Each sampled student was randomly assigned one of the nine test booklets. The possibilities for booklet assignment varied across sub-populations: OO students could receive one of the booklet 1 series only (the mathematics and science literacy booklets 1A or 1B); OP students could receive one of the booklet 2 series (the physics booklets 2A, 2B, or 2C) or one of the booklet 1 series (since all students are eligible to receive a literacy booklet); MO students could receive one of the booklet 3 series (the advanced mathematics booklets 3A, 3B, or 3C) or one of the booklet 1 series; and MP students could receive booklet 4 (which combines literacy, advanced mathematics, and physics questions) or any of the booklets in series 1, 2, or 3. The random assignment (or rotation, since booklets were actually distributed systematically within schools or classes) of booklets to students constituted another stage of sampling, and consequently had to be included in the calculation of weights. The booklet weights represent the booklet assignments as they were implemented in the student sample. There is one weight for each booklet series distributed within each sub-population. The set of students that were assigned the same booklet can be thought of as a sub-sample of the total sample. The booklet weights may be used when the focus of the analysis is on individual items rather than on summary scales.

[1] When classroom sampling was used, all students in the class were included and in that case $BW_{gi}^{st} = 1.0$.

To compute the booklet weights, we need to know for each sub-population g (g = OO, OP, MO, MP) how many booklets of each kind were distributed among the r_{gi} participants. Let r_{gi}^{b} be the number of participants in school i and sub-population g who received booklet b (b = 1, 2, 3, 4). Then we have:

$$W_{gi}^{1} = \frac{r_{gi}}{r_{gi}^{1}}, \text{ for all } g$$

for the mathematics and science literacy booklets, 1A and 1B,

$$W_{gi}^{2} = \begin{cases} 0, \text{ if } g \in \{OO, MO\} \\ \frac{r_{gi}}{r_{gi}^{2}}, \text{ if } g \in \{OP, MP\} \end{cases}$$

for the physics booklets, 2A, 2B and 2C,

$$W_{gi}^{3} = \begin{cases} 0, \text{ if } g \in \{OO, MO\} \\ \frac{r_{gi}}{r_{gi}^{3}}, \text{ if } g \in \{MO, MP\} \end{cases}$$

for the advanced mathematics booklets, 3A, 3B, 3C, and finally

$$W_{gi}^{4} = \begin{cases} 0, \text{ if } g \in \{OO, OP, MO\} \\ \frac{r_{gi}}{r_{gi}^{4}}, \text{ if } g \in \{MP\} \end{cases}$$

for the combined booklet, booklet 4.

4.2.5 The Scale Weights

The booklet weights permit properly weighted analyses of student responses to individual items and were necessary since such analyses are an important aspect of the international reports. However, most of the TIMSS reporting made use of IRT scales, which summarize student performance across all of the items in a subject area. Because TIMSS requires more than one booklet to cover a subject area, but each student responded to only one booklet, the IRT scales had to combine item responses from different booklets, and hence from different students, and the scale weights had to reflect this. The scale weights are rooted in the booklet sub-samples and in the sub-populations.

The mathematics and science literacy estimation weight was based on all students that were assigned booklet series 1 or 4, and was constructed as follows:

$$W_{gi}^{MSL} = \frac{r_{gi}}{r_{gi}^{1,4}}, \text{ for all } g$$

The weight for the physics scale involved students in the OP and MP sub-populations that were assigned booklet series 2 or 4, as follows:

$$W_{gi}^{P} = \begin{cases} 0, \text{ if } g \in \{\text{OO, MO}\} \\ \dfrac{r_{gi}}{r_{gi}^{2,4}}, \text{ if } g \in \{\text{OP, MP}\} \end{cases}$$

Finally, the weight for the advanced mathematics scale involved students in the MO and MP sub-populations that were assigned booklet series 3 or 4, as follows:

$$W_{gi}^{AM} = \begin{cases} 0, \text{ if } g \in \{\text{OO, OP}\} \\ \dfrac{r_{gi}}{r_{gi}^{3,4}}, \text{ if } g \in \{\text{MO, MP}\} \end{cases}$$

4.2.6 The Adjustments for Unbalanced Booklet Rotation

In many instances, there were fewer students from a sub-population in a school or class than the number of different booklets to be rotated. Uncorrected, this situation could make estimates of the population size vary with the choice of weight series. Since the estimated number of physics students (say) should be the same regardless of whether it was estimated using the "booklet weight" or the "scale weight," this is not a desirable situation. Adjustment factors for booklet and scale weights were devised to correct for the potential relative rarity of certain booklets in the sample.

First, an estimate of the size of each sub-population g, g=OO, OP, MO, MP, was computed:

$$SI\hat{Z}E_g = \sum_{i}^{n}\sum_{j}^{r_{gi}} FW_i^{sc} \times FW_{gi}^{cl} \times FW_{gi}^{st}.$$

Then, an estimate was constructed of the size of the sub-population g using in turn each weight series b (b = 1, 2, 3, 4, MSL, P, AM) as defined in the preceding sections:

$$SI\hat{Z}E_g^{b} = \sum_{i}^{n}\sum_{j}^{r_{gi}} FW_i^{sc} \times FW_{gi}^{cl} \times FW_{gi}^{st} \times W_{gi}^{b}.$$

The correction factor is therefore:

$$K_g^b = \frac{SI\hat{Z}E_g}{SI\hat{Z}E_g^{b}}$$

for each booklet series b (b = 1, 2, 3, 4, MSL, P, AM) and each sub-population g, (g=OO, OP, MO, MP).

Hence, the final booklet or scale weight becomes:

$$FW_{gi}^{\mathrm{b}} = W_{gi}^{\mathrm{b}} \times K_g^{\mathrm{b}}.$$

4.2.7 The Complete Weight

At the end of the process, the estimation weight assigned to a student *j* depends on the school the student attends, the classroom, if classroom sampling has been used, the sub-population the student belongs to, and the booklet the student was assigned. Both booklet-based and scale-based weights were computed.

$$CW_{gij}^{\mathrm{b}} = FW_i^{sc} \times FW_{gi}^{cl} \times FW_{gi}^{st} \times FW_{gi}^{\mathrm{b}}.$$

Further details of the weights that were computed and are available in the TIMSS user database may be found in Gonzalez, Smith, and Sibberns (1998).

REFERENCES

Gonzalez, E. J., Smith, T. A., and Sibberns, H. (Eds.). (1998). *User guide for the TIMSS international database: Final year of secondary school – 1995 assessment*. Chestnut Hill, MA: Boston College.

5 Estimation of Sampling and Imputation Variability

Eugenio J. Gonzalez
Boston College

Pierre Foy
Statistics Canada

5.1 OVERVIEW

One aim of TIMSS was to obtain accurate and cost-effective estimates of student performance in the populations under study – mathematics and science literacy among students in their final year of secondary school, and advanced mathematics and physics among final-year students with preparation in these subjects. To that end, TIMSS made extensive use of probability sampling techniques to sample students from national student populations.[1] Statistics computed from these national probability samples were used as estimates of population parameters. Because some uncertainty is involved in generalizing from samples of people to populations, the important statistics in the TIMSS international report (Mullis, Martin, Beaton, Gonzalez, Kelly, and Smith, 1998) are presented together with their standard errors, which are a measure of this uncertainty.

The TIMSS item pool was far too extensive to be administered in its entirety to any one student, and so a complex test design was developed whereby each student was given a single test booklet containing only a part of the entire assessment.[2] The results for all of the booklets were then aggregated using item response theory to provide results for the entire assessment. A consequence of this approach was that each student responded to just a few items from each content area in the assessment, and therefore multiple imputation or "plausible values" (see Chapter 7 of this volume) had to be used to derive reliable indicators of student proficiency. Since each proficiency estimate incorporates some uncertainty it is customary to generate a number of estimates (usually five) for each student, and to use the variability among the five estimates as a measure of this imputation uncertainty, or error. In the TIMSS international report the imputation error for each variable has been combined with the sampling error for that variable to provide a standard error incorporating both.

5.2 ESTIMATING SAMPLING VARIANCE

The TIMSS sampling design applies stratified multistage cluster-sampling techniques to the problem of selecting efficient and accurate samples of students while working with schools and classes. Such complex designs capitalize on the structure of the student population (i.e., students grouped in classes within schools) to derive student

1 See Foy, Rust, and Schleicher (1996) for details of the TIMSS sampling design.

2 See Adams and Gonzalez (1996) for details of the TIMSS test design.

samples that permit efficient and economical data collection. However, complex sampling designs make the task of computing standard errors to quantify sampling variability more difficult.

When, as in TIMSS, the sampling design involves multistage cluster sampling, there are several options for estimating sampling errors that avoid the assumption of simple random sampling (see Wolter, 1985). The jackknife repeated replication technique (JRR) was chosen in TIMSS because it is computationally straightforward and provides approximately unbiased estimates of the sampling errors of means, totals, and percentages in complex sample designs.

The particular variation on the JRR technique used in TIMSS is described in Johnson and Rust (1992). This method assumes that the primary sampling units (PSUs) can be paired in a manner consistent with the sample design, and each pair regarded as members of a pseudo-stratum for variance estimation purposes. Note that when using the JRR technique for the estimation of sampling variability, the approach will appropriately reflect the combined effect of the between- and within-PSU contributions to the sampling variance. The general use of the JRR entails systematically assigning pairs of schools to sampling zones, and randomly selecting one of these schools to have its contribution doubled, and the other to have it zeroed, so as to construct a number of "pseudo-replicates" of the original sample. The statistic of interest is computed once for all of the original sample, and once again for each pseudo-replicate sample. The variation between the estimates from each of the replicate samples and the original sample estimate is the jackknife estimate of the sampling error of the statistic.

5.3 CONSTRUCTION OF SAMPLING ZONES FOR SAMPLING VARIANCE ESTIMATION

To apply the JRR technique used in TIMSS it is necessary to pair the sampled schools and assign them to a series of groups known as sampling zones. This is done by working through the list of sampled schools in the order in which they were selected and assigning the first and second schools to the first sampling zone, the third and fourth schools to the second zone, and so on. A maximum of 75 zones was used, allowing for a total of 150 schools per country. In countries where more than 150 schools were sampled, it was sometimes necessary to combine two schools before assigning them to a sampling zone.

Sampling zones were constructed within design domains, or explicit strata. Where there was an odd number of schools in an explicit stratum, either by design or because of school nonresponse, the students in the remaining school were randomly divided to make up two "quasi" schools for the purposes of calculating the jackknife standard error. Each zone then consisted of a pair of schools or "quasi" schools. Table 5.1 shows the range of sampling zones used in each country.

Table 5.1 Sampling Zones - Population 3

Country	Mathematics and Science Literacy Sample		Advanced Mathematics Sample		Physics Sample	
	First Zone	Last Zone	First Zone	Last Zone	First Zone	Last Zone
Australia	1	49	1	49	1	49
Austria	1	75	1	58	1	58
Canada	1	75	1	75	1	75
Cyprus	1	28	1	21	1	21
Czech Republic	1	75	1	45	1	45
Denmark	1	65	1	65	1	65
France	32	70	1	31	1	31
Germany	1	75	1	75	1	75
Greece	-	-	1	30	1	30
Hungary	1	75	-	-	-	-
Iceland	1	30	-	-	-	-
Italy	1	52	7	51	-	-
Latvia (LSS)	-	-	-	-	1	19
Lithuania	1	73	1	29	-	-
Netherlands	1	40	-	-	-	-
New Zealand	1	39	-	-	-	-
Norway	1	66	-	-	1	33
Russian Federation	1	41	1	41	1	41
Slovenia	1	39	1	39	1	39
South Africa	1	52	-	-	-	-
Sweden	1	75	1	75	1	75
Switzerland	1	75	20	75	20	75
United States	1	33	1	33	1	33

A dash (-) indicates the country did not participate in the assessment of this subject area.

5.4 COMPUTING SAMPLING VARIANCE USING THE JRR METHOD

The JRR algorithm used in TIMSS assumes that there are H sampling zones within each country, each one containing two sampled schools selected independently. When computing a statistic t from the sample for a country, the formula for the JRR variance estimate of the statistic t is then given by the following equation:

$$Var_{jrr}(t) = \sum_{h=1}^{H} [t(J_h) - t(S)]^2$$

where H is the number of pairs in the sample for the country. The term $t(S)$ corresponds to the statistic for the whole sample (computed with any specific weights that may have been used to compensate for the unequal probability of selection of the different elements in the sample or any other post-stratification weight). The element $t(J_h)$ denotes the same statistic using the hth jackknife replicate. This is computed using all cases except those in the hth zone of the sample; for those in the hth zone, all cases associated with one of the randomly selected units of the pair are removed, the elements associated with the other unit in the zone are included twice. In practice, this is effectively accomplished by recoding to zero the weights for the cases of the element of the pair to be excluded from the replication, and multiplying by two the weights of the remaining element within the hth pair.

The computation of the JRR variance estimate for any statistic in TIMSS requires the computation of the statistic up to 76 times for any given country: once to obtain the statistic for the full sample, and up to 75 times to obtain the statistics for each of the jackknife replicates (J_h). The number of times a statistic needs to be computed for a given country depends on the number of implicit strata or sampling zones defined for that country.

Doubling and zeroing the weights of the selected units within the sampling zones is accomplished effectively by creating replicate weights that are then used in the calculations. Gonzalez, Smith, and Sibberns (1998) provide examples of how this approach allows standard statistical software such as SAS or SPSS to be used to compute JRR estimates of sampling variability in TIMSS. The replicate weight approach requires the user to temporarily create a new set of weights for each pseudo-replicate sample. Each replicate weight is equal to k times the overall sampling weight, where k can take values of 0, 1 or 2 depending on whether the case is to be removed from the computation, left as it is, or have its weight doubled. The value of k for an individual student record for a given replicate depends on the assignment of the record to the specific PSU and zone.

Within each zone the members of the pair of schools are assigned an indicator (u_i), coded randomly to 1 or 0 so that one of the members of each pair has a value of 1 on the variable u_i, and the remaining member a value of 0. This indicator determines whether the weights for the elements in the school in this zone are to be doubled or zeroed. The replicate weight ($W_h^{g,i,j}$) for the elements in a school assigned to zone h is computed as the product of k_h times their overall sampling weight, where k_h can take values of 0, 1, or 2 depending on whether the school is to be omitted, be included with its usual weight, or have its weight doubled for the computation of the statistic of interest. In TIMSS, the replicate weights are not permanent variables, but are created temporarily by the sampling variance estimation program as a useful computing device.

Replicate weights were created by the following procedure.

Each sampled student was assigned a vector of 75 weights, $W_h^{g,i,j}$, where h takes values from 1 to 75. The value of $W_0^{g,i,j}$ is the overall sampling weight, which is simply the product of the final school weight, the appropriate final classroom weight, and the appropriate final student weight, as described in Chapter 4.

The replicate weights for a single case were then computed as:

$$W_h^{g,i,j} = W_0^{g,i,j} * k_{hi}$$

where the variable k_h for an individual i takes the value $k_{hi} = 2*u_i$ if the record belongs to zone h, and $k_{hi} = 1$ otherwise.

In the TIMSS analysis, a total of 75 replicate weights were computed for each country regardless of the number of actual zones within the country. If a country had fewer than 75 zones, then the replicate weights W_h, where h was greater than the number of zones within the country, were each the same as the overall sampling weight. Although this involved some redundant computation, having 75 replicate weights for each country has no effect on the size of the error variance computed using the jackknife formula, but facilitated the computation of standard errors for a number of countries at one time.

Example SAS and SPSS programs used to compute standard errors in TIMSS are given in Gonzalez, Smith, and Sibberns (1998). Although standard errors presented in the international reports were computed using SAS programs developed at the International Study Center, they were also verified against results produced by the WesVarPC software (Westat, 1997). Results were compared with each other for accuracy.[3]

5.5 ESTIMATING IMPUTATION VARIANCE

The general procedure for estimating the imputation variance using plausible values is as follows: first compute the statistic (t), for each set of plausible values (M). The statistics t_m can be anything estimable from the data, such as a mean, the difference between means, or percentiles, and so forth. Each of these statistics will be called t_m, where m = 1, 2, …, 5.

Once the statistics are computed, the imputation variance is then computed as:

$$Var_{imp} = (1 + 1/M) \times Var(t_m)$$

where M is the number of plausible values used in the calculation, and Var(t_m) is the variance of the estimates computed using each plausible value.

5.6 COMBINING SAMPLING AND IMPUTATION VARIANCE

When reporting standard errors for proficiency estimates using plausible values, it is necessary to combine the sampling and imputation components of the error variance for the estimate. Under ideal circumstances and with unlimited computing resources, the user would compute the imputation variance for the plausible values and the JRR sampling variance for each of the plausible values. This would be equivalent to computing the same statistic up to 380 times (once overall for each of the five plausible values using the overall sampling weights, and then 75 times more for each plausible value using the complete set of replicate weights). However, an acceptable shortcut is to compute the JRR variance component using one plausible value, and then the imputation variance using the five plausible values. Using this approach, the same statistic needs to be computed only 80 times. Under this procedure the error variance component for a statistic is computed using the following formula:

$$Var(t_{pv}) = Var_{jrr}(t_1) + Var_{imp}$$

where $Var_{jrr}(t_1)$ is the sampling variance for the first plausible value. The User Guide for the TIMSS International Database (Gonzalez, Smith, and Sibberns, 1998) contains

3 Minor differences were occasionally found between the results obtained with WesVar and those obtained with software developed in-house. However, all these differences were due to the fact that the two programs did not always choose the same PSUs in forming jackknife replicates. When identical jackknife replicates were used for both programs, the results were identical.

programs in SAS and SPSS that compute each of these variance components for the TIMSS data. Tables 5.2 through 5.14 show, for each set of plausible values, the sample size, estimate of the population size, mean of the five plausible values, error due to sampling, error due to imputation, sampling and imputation errors combined, standard deviation of the plausible values, and its standard error. These statistics are presented for males and females separately in Appendix C.

Table 5.2 Sampling and Imputation Standard Errors - Mathematics and Science Literacy Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	1941	170847	525	9.5	0.5	9.5	94.9	4.8
Austria	1962	70602	519	5.3	0.8	5.4	80.4	3.1
Canada	5232	263241	526	2.5	0.7	2.6	83.2	1.6
Cyprus	534	4556	447	2.4	0.8	2.5	73.1	2.3
Czech Republic	2167	137459	476	10.5	0.5	10.5	91.8	3.3
Denmark	2714	37872	528	3.1	0.4	3.2	81.1	2.3
France	1590	637935	505	4.8	0.8	4.9	74.0	2.7
Germany	2289	967705	496	5.1	1.6	5.4	88.9	3.2
Hungary	5091	111281	477	3.0	0.5	3.0	84.3	2.4
Iceland	1703	2308	541	1.5	0.6	1.6	77.4	1.2
Italy	1616	380834	475	5.2	0.7	5.3	83.3	4.0
Lithuania	2887	22161	465	5.7	0.5	5.8	80.4	3.3
Netherlands	1470	145916	559	4.8	1.0	4.9	84.2	4.0
New Zealand	1763	37549	525	4.6	0.7	4.7	92.0	2.4
Norway	2518	43806	536	4.0	0.5	4.0	87.8	2.1
Russian Federation	2289	1031187	476	5.8	0.8	5.8	83.1	2.9
Slovenia	1622	26644	514	8.1	0.6	8.2	82.3	4.4
South Africa	2757	374618	352	9.3	0.5	9.3	87.6	8.7
Sweden	3068	71243	555	4.3	0.5	4.3	91.3	2.2
Switzerland	3308	65140	531	5.3	1.1	5.4	87.6	2.6
United States	5807	2278258	471	3.1	0.5	3.1	89.1	2.1

S.D. = standard deviation
S.E. = standard error

Table 5.3 Sampling and Imputation Standard Errors - Mathematics Literacy Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	1941	170847	522	9.3	0.4	9.3	97.2	4.9
Austria	1962	70602	518	5.3	0.7	5.3	79.8	2.8
Canada	5232	263241	519	2.8	0.6	2.8	90.1	1.7
Cyprus	534	4556	446	2.4	0.7	2.5	72.9	2.6
Czech Republic	2167	137459	466	12.3	0.8	12.3	99.4	3.5
Denmark	2714	37872	547	3.3	0.5	3.3	86.7	2.8
France	1590	637935	523	5.0	0.8	5.1	79.2	2.8
Germany	2289	967705	495	5.6	1.9	5.9	93.7	3.2
Hungary	5091	111281	483	3.1	0.6	3.2	92.3	2.2
Iceland	1703	2308	534	2.0	0.5	2.0	87.9	1.4
Italy	1616	380834	476	5.3	1.1	5.5	87.4	3.9
Lithuania	2887	22161	469	6.0	0.7	6.1	84.6	3.5
Netherlands	1470	145916	560	4.7	0.7	4.7	90.0	3.5
New Zealand	1763	37549	522	4.4	0.8	4.5	98.2	2.2
Norway	2518	43806	528	4.1	0.5	4.1	93.9	1.9
Russian Federation	2289	1031187	471	6.1	1.0	6.2	85.5	3.2
Slovenia	1622	26644	512	8.3	0.7	8.3	86.8	4.4
South Africa	2757	374618	356	8.3	0.6	8.3	81.3	8.5
Sweden	3068	71243	552	4.3	0.5	4.3	98.7	2.3
Switzerland	3308	65140	540	5.7	1.1	5.8	88.5	2.5
United States	5807	2278258	461	3.1	0.7	3.2	91.1	1.9

S.D. = standard deviation
S.E. = standard error

Table 5.4 Sampling and Imputation Standard Errors - Science Literacy Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	1941	170847	527	9.7	1.1	9.8	99.6	5.0
Austria	1962	70602	520	5.5	1.3	5.6	87.0	3.6
Canada	5232	263241	532	2.5	0.9	2.6	84.8	1.9
Cyprus	534	4556	448	2.8	1.0	3.0	82.9	2.7
Czech Republic	2167	137459	487	8.8	0.6	8.8	91.2	3.0
Denmark	2714	37872	509	3.6	0.4	3.6	86.9	2.4
France	1590	637935	487	5.0	1.2	5.1	78.8	2.4
Germany	2289	967705	497	4.9	1.3	5.1	90.7	3.5
Hungary	5091	111281	471	3.0	0.4	3.0	86.2	2.5
Iceland	1703	2308	549	1.4	0.6	1.5	75.4	1.4
Italy	1616	380834	475	5.3	0.6	5.3	86.7	3.9
Lithuania	2887	22161	461	5.7	0.5	5.7	84.0	3.2
Netherlands	1470	145916	558	5.1	1.4	5.3	85.5	4.5
New Zealand	1763	37549	529	5.1	1.0	5.2	94.4	3.2
Norway	2518	43806	544	4.1	0.6	4.1	91.2	2.5
Russian Federation	2289	1031187	481	5.6	0.8	5.7	91.0	2.8
Slovenia	1622	26644	517	8.1	0.6	8.2	83.8	4.7
South Africa	2757	374618	349	10.4	0.8	10.5	99.6	8.7
Sweden	3068	71243	559	4.4	0.4	4.4	91.0	2.2
Switzerland	3308	65140	523	5.2	1.2	5.3	93.6	2.7
United States	5807	2278258	480	3.2	0.4	3.3	93.9	2.5

S.D. = standard deviation
S.E. = standard error

Table 5.5 Sampling and Imputation Standard Errors - Advanced Mathematics Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	645	39498	525	11.6	1.2	11.6	109.2	7.9
Austria	782	31063	436	7.2	0.6	7.2	91.3	5.5
Canada	2781	58606	509	4.2	0.8	4.3	98.3	2.4
Cyprus	391	837	518	4.1	1.5	4.3	85.4	3.0
Czech Republic	1101	19446	469	11.1	1.3	11.2	106.3	9.3
Denmark	1388	13527	522	3.3	0.8	3.4	72.9	1.9
France	1071	151531	557	3.8	0.9	3.9	70.1	2.1
Germany	2296	262789	465	5.5	0.9	5.6	85.0	3.4
Greece	456	14620	513	5.9	1.2	6.0	104.9	6.0
Italy	398	104477	474	9.5	0.6	9.6	95.2	8.1
Lithuania	734	1360	516	2.4	0.9	2.6	85.1	3.2
Russian Federation	1638	42858	542	9.2	1.1	9.2	111.6	5.6
Slovenia	1536	22881	475	9.1	0.9	9.2	93.8	3.8
Sweden	1001	16408	512	4.3	0.8	4.4	85.8	2.9
Switzerland	1404	11343	533	4.8	1.4	5.0	90.5	2.7
United States	2785	496852	442	5.8	1.1	5.9	98.0	4.1

S.D. = standard deviation
S.E. = standard error

Table 5.6 Sampling and Imputation Standard Errors Numbers, Equations and Functions Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	645	39498	517	9.2	1.4	9.4	98.0	6.3
Austria	782	31063	412	7.4	0.9	7.4	91.4	5.8
Canada	2781	58606	512	3.8	1.0	3.9	84.3	2.8
Cyprus	391	837	510	5.4	1.6	5.7	92.8	2.9
Czech Republic	1101	19446	460	11.6	1.3	11.7	103.6	7.3
Denmark	1388	13527	504	2.6	0.7	2.7	61.8	1.7
France	1071	151531	548	4.1	0.4	4.1	56.2	2.9
Germany	2296	262789	457	5.0	0.8	5.0	80.1	3.6
Greece	456	14620	539	6.7	2.4	7.2	112.7	7.3
Italy	398	104477	460	9.2	0.3	9.2	103.2	8.7
Lithuania	734	1360	547	2.3	1.7	2.8	84.6	2.7
Russian Federation	1638	42858	555	8.7	1.4	8.8	106.7	5.6
Slovenia	1536	22881	491	9.8	0.7	9.9	105.8	5.0
Sweden	1001	16408	523	4.7	0.7	4.7	88.4	3.2
Switzerland	1404	11343	514	5.1	1.1	5.2	87.5	2.6
United States	2785	496852	459	5.2	1.1	5.3	86.6	3.8

S.D. = standard deviation
S.E. = standard error

Table 5.7 Sampling and Imputation Standard Errors - Calculus Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	645	39498	530	11.6	1.1	11.7	100.3	11.2
Austria	782	31063	439	6.4	1.3	6.5	88.9	5.1
Canada	2781	58606	503	3.4	1.2	3.6	92.8	3.1
Cyprus	391	837	561	5.0	1.4	5.2	99.8	3.8
Czech Republic	1101	19446	446	9.7	0.7	9.7	97.3	7.5
Denmark	1388	13527	508	3.3	0.3	3.3	85.6	2.2
France	1071	151531	560	2.9	0.7	3.0	64.0	2.3
Germany	2296	262789	454	4.3	0.7	4.4	85.0	3.2
Greece	456	14620	538	6.8	2.8	7.3	98.3	6.4
Italy	398	104477	520	10.4	0.4	10.4	107.3	6.2
Lithuania	734	1360	498	2.0	1.5	2.5	78.1	2.2
Russian Federation	1638	42858	537	9.0	1.3	9.1	106.4	7.4
Slovenia	1536	22881	471	6.6	0.8	6.6	71.2	2.5
Sweden	1001	16408	480	4.3	0.7	4.4	88.0	2.7
Switzerland	1404	11343	512	5.6	1.1	5.7	96.6	3.5
United States	2785	496852	450	4.0	1.0	4.1	97.3	4.1

S.D. = standard deviation
S.E. = standard error

Table 5.8 Sampling and Imputation Standard Errors - Geometry Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	645	39498	496	12.4	1.9	12.5	122.5	10.0
Austria	782	31063	462	7.8	0.5	7.9	97.7	6.5
Canada	2781	58606	499	3.7	0.9	3.8	95.0	1.9
Cyprus	391	837	517	4.7	1.6	4.9	99.1	4.4
Czech Republic	1101	19446	494	9.8	0.9	9.8	102.2	8.7
Denmark	1388	13527	527	2.9	1.2	3.1	70.5	2.2
France	1071	151531	544	3.7	0.9	3.8	76.8	2.5
Germany	2296	262789	487	5.4	0.5	5.5	75.1	4.0
Greece	456	14620	498	8.0	3.3	8.7	116.3	7.2
Italy	398	104477	480	9.5	0.5	9.5	103.7	9.0
Lithuania	734	1360	515	2.3	1.6	2.8	82.3	1.9
Russian Federation	1638	42858	548	9.0	1.6	9.2	103.6	5.5
Slovenia	1536	22881	476	7.5	0.9	7.6	83.4	3.3
Sweden	1001	16408	492	4.4	0.6	4.4	83.0	2.6
Switzerland	1404	11343	547	4.1	1.1	4.2	87.0	3.1
United States	2785	496852	424	5.1	0.7	5.1	96.5	4.4

S.D. = standard deviation
S.E. = standard error

Table 5.9 Sampling and Imputation Standard Errors - Physics Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	661	31619	518	5.8	2.1	6.2	82.2	3.6
Austria	777	30795	435	6.4	0.8	6.4	83.3	4.6
Canada	2367	51179	485	2.8	1.8	3.3	86.7	3.0
Cyprus	368	837	494	5.7	0.8	5.8	105.3	5.3
Czech Republic	1087	19428	451	6.2	0.7	6.2	82.2	5.9
Denmark	654	2073	534	3.9	1.3	4.2	84.6	3.9
France	1110	151531	466	3.5	1.5	3.8	65.9	3.1
Germany	723	87888	522	11.8	1.2	11.9	94.2	5.3
Greece	459	14668	486	5.4	1.2	5.6	87.3	3.7
Latvia (LSS)	708	979	488	21.5	0.6	21.5	100.3	10.6
Norway	1048	4369	581	6.1	2.3	6.5	90.5	2.5
Russian Federation	1233	32975	545	11.4	2.4	11.6	110.3	5.0
Slovenia	747	11706	523	15.3	2.5	15.5	108.9	8.7
Sweden	1012	16459	573	3.8	0.8	3.9	92.1	2.8
Switzerland	1371	11276	488	3.4	0.8	3.5	88.3	2.9
United States	3114	522784	423	3.2	0.7	3.3	59.9	3.2

S.D. = standard deviation
S.E. = standard error

Table 5.10 Sampling and Imputation Standard Errors - Mechanics Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	661	31619	507	5.9	1.2	6.1	87.6	3.9
Austria	777	30795	420	4.9	0.2	4.9	78.0	3.5
Canada	2367	51179	473	3.5	0.8	3.6	89.0	2.8
Cyprus	368	837	530	6.6	0.4	6.6	116.9	5.4
Czech Republic	1087	19428	469	5.9	0.8	6.0	80.5	6.3
Denmark	654	2073	529	4.7	1.2	4.9	87.3	3.9
France	1110	151531	457	3.9	1.8	4.3	74.8	3.8
Germany	723	87888	495	9.2	1.7	9.4	90.3	6.6
Greece	459	14668	514	6.3	1.8	6.5	90.7	4.4
Latvia (LSS)	708	979	489	18.1	0.6	18.1	91.6	9.1
Norway	1048	4369	572	5.9	2.3	6.4	89.5	4.0
Russian Federation	1233	32975	537	9.3	0.6	9.3	91.4	6.5
Slovenia	747	11706	552	17.2	1.9	17.3	120.2	11.3
Sweden	1012	16459	563	4.0	0.4	4.0	80.4	2.6
Switzerland	1371	11276	482	3.5	0.3	3.5	86.2	3.5
United States	3114	522784	420	2.7	0.5	2.8	59.0	2.7

S.D. = standard deviation
S.E. = standard error

Table 5.11 Sampling and Imputation Standard Errors - Electricity and Magnetism Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	661	31619	512	4.2	1.1	4.4	91.0	5.5
Austria	777	30795	432	6.2	1.2	6.3	93.6	4.8
Canada	2367	51179	485	3.4	1.4	3.7	83.2	4.2
Cyprus	368	837	502	6.2	1.0	6.3	114.5	7.0
Czech Republic	1087	19428	465	5.4	0.7	5.5	75.2	6.0
Denmark	654	2073	513	3.5	1.5	3.8	79.4	4.6
France	1110	151531	494	4.0	0.4	4.1	59.9	3.3
Germany	723	87888	512	9.6	2.3	9.9	92.0	5.4
Greece	459	14668	520	6.4	1.7	6.6	105.1	4.8
Latvia (LSS)	708	979	485	17.3	1.1	17.4	94.3	8.2
Norway	1048	4369	565	5.9	1.8	6.2	93.0	3.3
Russian Federation	1233	32975	549	9.0	1.7	9.2	107.2	4.9
Slovenia	747	11706	509	14.6	1.5	14.6	109.9	11.0
Sweden	1012	16459	570	3.2	0.7	3.3	88.1	3.3
Switzerland	1371	11276	480	4.5	0.8	4.5	93.7	3.3
United States	3114	522784	420	2.9	0.7	3.0	58.6	2.2

S.D. = standard deviation
S.E. = standard error

Table 5.12 Sampling and Imputation Standard Errors - Heat Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	661	31619	517	4.1	1.2	4.3	82.5	3.9
Austria	777	30795	445	5.5	0.9	5.6	93.0	4.9
Canada	2367	51179	508	3.7	2.1	4.2	85.3	5.0
Cyprus	368	837	476	6.7	0.4	6.7	145.6	5.8
Czech Republic	1087	19428	488	4.6	0.8	4.7	80.1	6.0
Denmark	654	2073	512	4.0	1.6	4.3	98.2	6.4
France	1110	151531	491	3.2	1.0	3.4	68.7	3.1
Germany	723	87888	496	5.8	2.6	6.4	97.7	5.0
Greece	459	14668	481	7.1	1.4	7.2	117.7	6.4
Latvia (LSS)	708	979	504	21.3	1.3	21.4	112.8	9.3
Norway	1048	4369	536	3.7	2.1	4.3	70.2	2.6
Russian Federation	1233	32975	530	10.3	1.6	10.4	105.4	7.3
Slovenia	747	11706	521	10.2	1.9	10.4	117.8	8.3
Sweden	1012	16459	522	4.2	0.9	4.3	81.9	2.4
Switzerland	1371	11276	509	3.4	0.9	3.6	89.5	2.7
United States	3114	522784	477	2.9	0.7	3.0	58.4	2.8

S.D. = standard deviation
S.E. = standard error

Table 5.13 Sampling and Imputation Standard Errors - Wave Phenomena Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	661	31619	519	6.6	1.8	6.9	98.0	6.0
Austria	777	30795	468	7.2	0.5	7.3	86.8	7.4
Canada	2367	51179	488	2.8	1.5	3.2	80.0	2.6
Cyprus	368	837	507	6.5	0.4	6.5	119.1	7.4
Czech Republic	1087	19428	447	5.4	0.4	5.4	75.9	3.4
Denmark	654	2073	537	5.2	1.7	5.5	97.1	5.6
France	1110	151531	463	3.6	0.2	3.6	73.2	2.7
Germany	723	87888	530	10.0	2.3	10.3	97.7	5.2
Greece	459	14668	453	5.2	0.9	5.3	93.3	4.9
Latvia (LSS)	708	979	498	17.5	1.0	17.6	91.1	11.9
Norway	1048	4369	560	5.1	1.7	5.4	90.4	2.9
Russian Federation	1233	32975	515	9.2	2.0	9.4	105.9	5.8
Slovenia	747	11706	514	11.4	1.4	11.5	115.1	7.0
Sweden	1012	16459	560	4.6	0.8	4.7	107.5	3.5
Switzerland	1371	11276	498	3.0	0.7	3.1	89.1	3.0
United States	3114	522784	451	2.1	0.6	2.2	53.3	1.6

S.D. = standard deviation
S.E. = standard error

Table 5.14 Sampling and Imputation Standard Errors Particle, Quantum, Astrophysics and Relativity Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	661	31619	521	5.6	1.3	5.8	88.0	4.0
Austria	777	30795	480	5.9	1.2	6.0	82.5	3.8
Canada	2367	51179	494	2.7	0.5	2.7	80.2	2.6
Cyprus	368	837	434	5.2	0.4	5.2	131.3	4.7
Czech Republic	1087	19428	453	4.9	0.9	4.9	87.3	4.1
Denmark	654	2073	544	4.7	1.2	4.9	81.4	4.7
France	1110	151531	474	3.3	0.7	3.4	60.9	4.0
Germany	723	87888	545	12.8	2.5	13.1	107.6	7.4
Greece	459	14668	447	4.9	0.6	4.9	92.8	5.3
Latvia (LSS)	708	979	488	19.0	1.3	19.0	95.3	9.8
Norway	1048	4369	576	5.1	1.4	5.3	84.0	4.2
Russian Federation	1233	32975	542	9.8	1.4	9.9	98.0	6.5
Slovenia	747	11706	511	15.0	1.9	15.1	112.8	10.1
Sweden	1012	16459	560	3.4	0.8	3.5	77.1	2.7
Switzerland	1371	11276	488	3.7	0.8	3.8	83.3	3.7
United States	3114	522784	456	2.4	0.6	2.5	49.3	2.9

S.D. = standard deviation
S.E. = standard error

REFERENCES

Adams, R. J., and Gonzalez, E. J. (1996). The TIMSS test design. In M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study technical report, volume I: Design and development*. Chestnut Hill, MA: Boston College.

Foy, P., Rust, K., and Schleicher, A. (1996). Sample design. In M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study technical report, volume I: Design and development*. Chestnut Hill, MA: Boston College.

Gonzalez, E. J., Smith, T. A., and Sibberns, H. (Eds.) (1998). *User guide for the TIMSS international database: Final year of secondary school – 1995 assessment.* Chestnut Hill, MA: Boston College.

Johnson, E. G., and Rust, K.F. (1992). Population references and variance estimation for NAEP data. *Journal of Educational Statistics*, *17*, 175-190.

Kish, L. (1965). *Survey sampling.* New York: John Wiley & Sons.

Mullis, I.V.S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., and Smith, T.A. (1998). *Mathematics and science achievement in the final year of secondary school: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College.

Westat, Inc. (1997). *A user's guide to WesVarPC.* Rockville, MD: Westat, Inc.

Wolter, K. M. (1985). *Introduction to variance estimation.* New York: Springer-Verlag.

6 Item Analysis and Review

Ina V.S. Mullis
Michael O. Martin
Boston College

6.1 CROSS-COUNTRY ITEM STATISTICS

In order to assess the statistical properties of the Population 3 (final year of secondary school) items before proceeding with item response theory (IRT) scaling (see Chapter 7), TIMSS computed a series of statistics for every item in every country. These basic item statistics (see Figure 6.1 for an example item) were produced by the IEA Data Processing Center. For each item, the display presents 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 a total score).[1] For multiple-choice items the display presents the percentage of students that chose each option, including the percentage that omitted or did not reach the item, and the point-biserial correlation between each option and the total score. For free-response items (which could have more than one score level), the display presents the difficulty and discrimination of each score level. As a prelude to the main IRT scaling, the display presents some statistics from a preliminary Rasch analysis, the Rasch item difficulty for each item, the standard error of this difficulty estimate, and an index of the goodness-of-fit of the item to the Rasch model (Wu, 1997).

The item-analysis display presents the difficulty level of each item separately for male and female students. As a guide to the overall statistical properties of the item, it also presents the international item difficulty (the mean of the item difficulties across countries) and the international item discrimination (the mean of the item discriminations).

As an aid to reviewers, the item-analysis display includes a series of "flags" signaling the presence of one or more conditions that might indicate a problem with an item. The following conditions are flagged:

- Item difficulty exceeds 95 percent in the sample as a whole
- Item difficulty is less than 25 percent for 4-option multiple-choice items in the sample as a whole (20 percent for 5-option items)

[1] For the purpose of computing the discrimination index, the total score was the percentage of items a student answered correctly in mathematics or science.

Figure 6.1 Example of Cross-Country Item Analysis

Population: '3' (C) Subject: 'General' Cluster:A Item: CSMGA01 Page 1

Country	Correct answer			Flags	Percentages for each alternative								Point biserials for each alternative								Rasch			Group difficulties				Internatinal Mean	
	N	DIFF	DISCR		A	B	C	D	E	W	OMIT	NR	A	B	C	D	E	W	OMIT	NR	RDIFF	SE	FIT	MAL	FEM	LOW	UPP	IDIFF	IDISCR
AUS	1958	42.2	0.30	.........sS.	5.6	42.2*	10.2	39.7			2.1	0.1	-0.12	.30*	-0.12	-0.14			-0.11	-0.02	1.11	0.05	1.09	49.7	37.1	.	.	39.5	0.26
AUT	2040	49.1	0.29	...q.....sS.	3.8	49.1*	5.8	38.2			2.1	0.1	-0.01	.29*	-0.10	-0.20			-0.12	0.00	0.53	0.05	1.07	60.0	39.6	.	.	39.5	0.26
CAN	5361	40.8	0.26	....Q....sS.	6.6	40.8*	10.2	40.9			1.5	0.1	-0.01	.26*	-0.11	-0.17			-0.05	-0.04	1.22	0.03	1.10	46.2	35.0	.	.	39.5	0.26
CHE	3458	46.9	0.35	.........sS.	5.1	46.9*	7.8	34.2			5.4	0.0	-0.01	.35*	-0.06	-0.26			-0.16	-0.05	1.16	0.04	1.01	56.7	36.7	.	.	39.5	0.26
CSK	2196	42.4	0.34	......B..sS.	8.5	42.4*	5.2	39.3			4.3	0.0	0.07	.34*	0.00	-0.36			-0.07	-0.06	0.90	0.05	1.10	50.5	33.9	.	.	39.5	0.26
CYP	538	31.6	0.25	...qQ.......	4.8	31.6*	9.9	49.6			3.9	0.0	-0.04	.25*	-0.12	-0.13			-0.05	0.00	0.75	0.10	1.08	33.6	29.5	.	.	39.5	0.26
DEU	2439	43.2	0.33	...q.....sS.	3.2	43.2*	4.3	37.7			8.2	0.4	-0.04	.33*	-0.11	-0.18			-0.18	-0.04	0.80	0.05	1.05	53.9	34.4	.	.	39.5	0.26
FRA	1865	30.4	0.26	....Q....s..	5.1	30.4*	14.5	44.6			4.9	0.2	-0.03	.26*	0.00	-0.19			-0.08	-0.05	1.38	0.05	1.06	32.9	27.6	.	.	39.5	0.26
GRC	353	54.4	0.20	......B.Fs..	7.1	54.4*	6.8	25.5			5.9	0.0	0.09	.20*	-0.06	-0.21			-0.05	0.00	0.27	0.12	1.14	57.1	48.1	.	.	39.5	0.26
HUN	5356	43.4	0.30	...q.....sS.	3.6	43.4*	4.0	42.7			6.4	0.0	-0.03	.30*	-0.06	-0.23			-0.09	0.00	0.58	0.03	1.04	46.9	39.8	.	.	39.5	0.26
ISL	1832	27.3	0.31	....Q.B..sS.	5.0	27.3*	6.4	59.1			1.9	0.1	0.00	.31*	0.04	-0.28			-0.07	-0.01	2.01	0.06	0.99	36.7	18.7	.	.	39.5	0.26
ISR	1357	20.9	0.20	p..qQ.B.FsS.	3.1	20.9*	8.5	58.1			8.0	1.4	-0.07	.20*	-0.09	0.07			-0.26	-0.06	1.91	0.07	1.16	26.6	14.9	.	.	39.5	0.26
LTU	2886	44.6	0.27	...q..B..s..	4.7	44.6*	4.4	38.6			7.2	0.4	0.04	.27*	-0.03	-0.19			-0.16	-0.04	0.38	0.04	1.10	47.1	43.4	.	.	39.5	0.26
MEX	3535	28.9	0.07	...qQbB.Fs..	4.0	28.9*	7.7	54.1			4.9	0.3	-0.12	.07*	-0.14	0.07			-0.03	0.00	0.37	0.04	1.14	31.0	26.4	.	.	39.5	0.26
NLD	1628	39.2	0.23	....Q.B.FsS.	8.8	39.2*	7.4	43.7			0.9	0.1	0.03	.23*	-0.14	-0.16			-0.06	0.03	1.47	0.05	1.14	46.2	31.6	.	.	39.5	0.26
NOR	2518	37.1	0.29	....Q....sS.	6.2	37.1*	8.4	43.3			5.0	0.0	-0.01	.29*	-0.07	-0.18			-0.14	-0.02	1.33	0.05	1.06	46.3	28.7	.	.	39.5	0.26
NZL	2308	37.5	0.27	....Q....sS.	5.5	37.5*	11.8	44.2			1.0	0.0	-0.10	.27*	-0.10	-0.14			-0.07	0.00	1.42	0.05	1.09	44.4	31.2	.	.	39.5	0.26
RUS	2599	49.6	0.36	...q.....sS.	3.2	49.6*	4.8	37.6			4.5	0.1	-0.01	.36*	-0.06	-0.32			-0.06	-0.03	0.35	0.04	1.01	60.2	42.8	.	.	39.5	0.26
SVN	1650	26.5	0.31	....Q....sS.	5.5	26.5*	6.9	50.3			10.2	0.0	-0.06	.31*	-0.03	-0.21			-0.04	0.00	1.52	0.06	1.06	32.9	19.4	.	.	38.9	0.26
SWE	3106	55.8	0.26	......B.FsS.	9.4	55.8*	7.5	25.7			1.3	0.0	0.05	.26*	-0.11	-0.24			-0.06	0.00	0.72	0.04	1.13	62.8	49.2	.	.	39.5	0.26
USA	5997	38.5	0.19	.....b..FsS.	8.0	38.5*	13.7	38.5			1.2	0.0	-0.10	.19*	-0.16	-0.01			-0.02	0.00	0.46	0.03	1.22	44.1	33.1	.	.	39.5	0.26
ZAF	3662	24.9	0.12	p...QbB.F.S.	12.8	24.9*	11.7	45.7			4.8	0.0	-0.10	.12*	-0.06	0.03			-0.06	0.00	-0.04	0.04	1.22	24.8	25.2	.	.	39.5	0.26

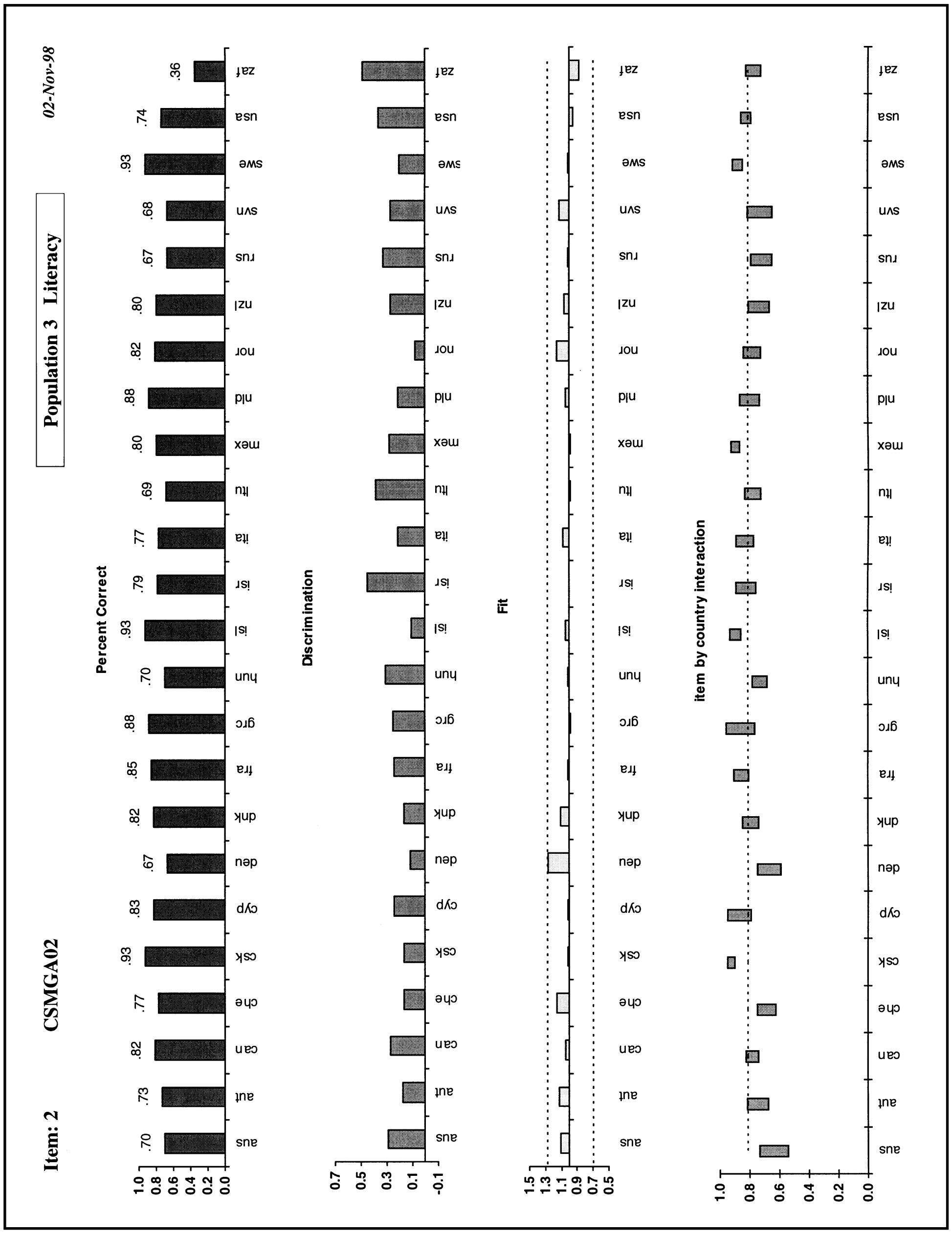
- Item difficulty exceeds 95 percent or is less than 25 percent (20 percent for 5-option items)
- Item difficulty exceeds 95 percent or is less than 25 percent (20 percent for 5-option items)
- One or more of the distracter percentages is less than 5 percent
- One or more of the distracter percentages is greater than the percentage for the correct answer
- Point-biserial correlation for one or more of the distracters exceeds zero
- Item discrimination (i.e., the point-biserial for the correct answer) is less than 0.2
- Item discrimination does not increase with each score level (for an item with more than one score level)
- Rasch goodness-of-fit index is less than 0.88 or greater than 1.12
- Difficulty levels on the item differ significantly for males and females
- Difference in item difficulty levels between males and females diverge significantly from the average difference between males and females across all the items making up the total score

Although not all of these conditions necessarily indicate a problem, the flags are a useful way to draw a reviewer's attention to potential sources of concern. The IEA Data Processing Center also produced information about the inter-rater agreement for the free-response items.

6.2 GRAPHICAL DISPLAYS

As a further aid to reviewing the psychometric characteristics of the items, the Australian Council for Educational Research (ACER) produced graphical representations of selected item statistics for each participating country (see Figure 6.2). This display presents, for each item, the difficulty level and discrimination for every country, together with the Rasch goodness-of-fit statistic and an indication of the item-by-country interaction. The item-by-country interaction chart plots a confidence interval for the probability of success on the item in each country against the average probability of success across all countries. The graphical representations allow comparisons of these statistics across countries at a glance.

Figure 6.2 Example of Graphical Displays of Cross-Country Item Statistics



6.3 SUMMARY INFORMATION FOR POTENTIALLY PROBLEMATIC ITEMS

Although the system of flagging potentially problematic conditions and the graphical summaries were both very helpful in identifying items with possible problems, the task of reviewing the characteristics of each item in each country was still considerable. To ensure that no serious item problem would go unnoticed, ACER also provided, for each item, a list of countries that exhibited one or more potentially serious characteristics (see Figure 6.3). Countries were listed in this display if the item had a significant item-by-country interaction (i.e., students in the country found the item easier or more difficult than items in general), or if they exhibited problematic discrimination (i.e., the point-biserial for a distracter was greater than .05, the point-biserial for the correct answer was negative, or, for items with more than one score point, the point-biserial did not increase with each score level). Countries were also listed if their data showed poor fit to the Rasch model for that item.

6.4 ITEM CHECKING PROCEDURES

Prior to the international scaling of the Population 3 achievement data by ACER, the International Study Center thoroughly reviewed the item statistics for all participating countries to ensure that items were performing comparably across countries. Although only a small number of items were found to be inappropriate for international comparisons, throughout the series of item-checking steps a number of reasons were discovered for differences in items across countries. Most of these were inadvertent changes in the items during printing, including omitting an item option or misprinting the graphics associated with an item. However, differences attributable to translation problems were found for an item or two in several countries.

In particular, items with the following problems were considered for possible deletion from the international database:

- Errors were detected during translation verification but were not corrected before test administration

- Data cleaning revealed more or fewer options than in the original version of the item

- The item-analysis information showed the item to have a negative biserial

- The item-by-country interaction results showed a very large negative interaction for a given country

- The item-fit statistic indicated the item did not fit the model

- For free-response items, the within-country scoring reliability data showed an agreement of less than 70 percent for the score level. Also, performance in items with more than one score level was not ordered by score, or correct levels were associated with negative point-biserials.

Figure 6.3 Example Summary Information for Items With Poor Statistics for Some Countries

	Item by Country Interactions		Discrimination			Fit
Country	Easier than Expected	Harder than Expected	Non-key PB is Positive	Key PB is Negative	Ability not Ordered	Fit Large
	Tolerance = #Name					
Item: 16	*CSEGA12*					
HUN	☐	☐	☐	☐	☐	☑
RUS	☐	☐	☐	☐	☐	☑
Item: 18	*CSMGB02*					
GRC	☐	☑	☑	☐	☐	☐
Item: 19	*CSMGB03*					
USA	☑	☐	☐	☐	☐	☐
Item: 20	*CSMGB04*					
FRA	☐	☑	☐	☐	☐	☐
ISR	☐	☐	☑	☐	☐	☐
SVN	☑	☐	☐	☐	☐	☐
Item: 21	*CSMGB05*					
CYP	☐	☐	☑	☐	☐	☐

The statistics and translation verification documentation were used as pointers towards checking actual booklets and contacting National Research Coordinators (NRCs). If a problem could be detected by the International Study Center (such as a negative point-biserial for a correct answer or too few options for the multiple-choice questions), the item was deleted from the international scaling. However, if there was a question about potential translation or cultural issues, then the NRC was queried, and the International Study Center abided by the decision made by the NRC. In several cases, NRCs consulted mathematics or science experts before making a decision.

Considering that the checking involved approximately 200 items for more than 20 countries, very few deviations from the international format were found. Tables 6.1 and 6.2 contain a list of the changes made in the international database for Population 3.

Table 6.1 Deleted Cognitive Items - Population 3

	Country	Item	Variable Name
Mathematics and Science Literacy	All	A09, Part A	CSEGA09A
		C10	CSMGC10
	Cyprus	C05	CSMGC05
		D12	CSMGD12
	Greece	C05	CSMGC05
		D12	CSMGD12
		A11, Part C	CSEGA11C
	France	B04	CSMGB04
		B06	CSMGB06
	Hungary	B08	CSMGB08
		B21	CSMGB21
		B26	CSSGB26
		C20	CSSGC20
		D15, Part B	CSSGD15B
		D16, Part A	CSSGD16A
		D16, Part B	CSSGD16B
	Switzerland	B06	CSMGB06
	Slovenia	A11, Part C	CSEGA11C
Advanced Mathematics	Cyprus	J02	CSMMJ02
	France	J18	CSEMJ18
	Greece	J02	CSMMJ02
	Israel	J14	CSMMJ14
		J16, Part B	CSSMJ16B
		L08	CSMML08
	Lithuania	K09	CSMMK09
	Switzerland	J02	CSMMJ02
		J17	CSSMJ17
	United States	J08	CSMMJ08
Physics	All	H11	CSMPH11
	Australia	H19, Part A	CSEPH19A
	Czech Republic	F06	CSMPF06
	Denmark	F07	CSMPF07
		H14	CSSPH14
	France	F15	CSEPF15
	Germany	G16	CSEPG16
		H14	CSSPH14

Table 6.2 Recodes Made to Population 3 Free-Response Item Codes

	Item	Variable	Recodes	Comment
Mathematics and Science Literacy	B25	CSSGB25	20 → 10 21 → 11 22 → 12 10 → 13 29 → 19	Category 10 was only 1 point category and generally had less than 1 percent of the students, which made distinction between 1 and 2 points unclear.
	B26	CSSGB26	10 → 23 19 → 29	Categories 10 and 19 contain correct answer.
	D02	CSSGD02	20 → 12 21 → 13	Discrimination between 20s and 10s not clear.
	D04	CSEGD04	20 → 10 21 → 11 22 → 12 29 → 19 10 → 73 11 → 74 19 → 75	Is a link item with Y01 at Population 2 and as with Population 2 only 20s had positive point-biserials in many countries.
	D17	CSSGD17	13 → 22	In some countries 10s had almost the same or even higher point-biserials than 20s.

REFERENCES

Wu, M.L. (1997). *The development and application of a fit test for use with marginal maximum likelihood estimation and generalised item response models*. Unpublished master's dissertation, University of Melbourne.

7 Scaling Methodology and Procedures for the Mathematics and Science Literacy, Advanced Mathematics, and Physics Scales

Greg Macaskill
Raymond J. Adams
Margaret L. Wu
Australian Council for Educational Research

Student achievement is reported in TIMSS mainly through scale scores derived using Item Response Theory (IRT) scaling. This approach allows the performance of a sample of students in a subject area to be summarized on a common scale or series of scales even when different students have been administered different items. The common scale makes it possible to report on relationships between students' characteristics (based on responses to the background questionnaires) and their performance in mathematics and science.

For Population 3, as for Populations 1 and 2, each student was administered only a subset of items within each content area in the three areas examine – advanced mathematics, physics and mathematics and science literacy. In this situation, to obtain reliable indices of student proficiency "plausible values" methodology was used. Some references to this work are given in Adams, Wu and Macaskill (1997).

This chapter gives details of the IRT model used in TIMSS to scale the Population 3 achievement data and includes a description of the model and the estimation process. For more details, see also the reference above and papers cited within this chapter.

7.1 THE TIMSS SCALING MODEL

The scaling model used in TIMSS was the multidimensional random coefficients logit model described by Adams, Wilson, and Wang (1997), with the addition of a multivariate linear model imposed on the population distribution. The scaling was done with the ConQuest software (Wu, Adams, and Wilson, 1997) that was developed in part to meet the needs of the TIMSS study.

7.1.1 The Multidimensional Random Coefficients Model

Assume that I items are indexed $i=1,\ldots,I$ with each item admitting $K_i + 1$ response alternatives $k=0,1,\ldots,K_i$. Use the vector valued random variable, $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iK_i})'$ where

$$X_{ij} = \begin{cases} 1 \text{ if response to item } i \text{ is in category } j \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

to indicate the $K_i + 1$ possible responses to item i.

A response in category zero is denoted by a vector of zeroes. This effectively makes the zero category a reference category and is necessary for model identification. The choice of this as the reference category is arbitrary and does not affect the generality of the model. We can also collect the $\mathbf{X}_i$ together into the single vector $\mathbf{X}' = (\mathbf{X}'_1, \mathbf{X}'_2, \ldots, \mathbf{X}'_I)$ which we call the response vector (or pattern). Particular instances of each of these random variables are indicated by their lower-case equivalents; $\mathbf{x}$, $\mathbf{x}_i$ and x_{ik}.

The items are described through a vector $\xi^T = (\xi_1, \xi_2, \ldots, \xi_p)$ of p parameters. Linear combinations of these are used in the response probability model to describe the empirical characteristics of the response categories of each item. These linear combinations are defined by design vectors $\mathbf{a}_{jk}$, ($j = 1,\ldots,I$; $k = 1,\ldots,K_i$) each of length p that can be collected to form a design matrix $\mathbf{A}' = (\mathbf{a}_{11}, \mathbf{a}_{12}, \ldots, \mathbf{a}_{1K_1}, a_{21}, \ldots, a_{2K_2}, \ldots, a_{1K_i})$.

The multidimensional form of the model assumes that a set of D traits underlie the individuals' responses. The D latent traits define a D-dimensional latent space and the individuals' positions in the D-dimensional latent space are represented by the vector $\theta = (\theta_1, \theta_2, \ldots, \theta_D)$.

An additional feature of the model is the introduction of a scoring function which allows the specification of the score or "performance level" that is assigned to each possible response to each item. To do this we introduce the notion of a response score b_{ijd} that gives the performance level of an observed response in category j of item I in dimension d. The scores across D dimensions can be collected first into a column vector $\mathbf{b}_{ik} = (b_{ik1}, b_{ik2}, \ldots, b_{ik1D})^T$, then into the scoring sub-matrix for item i, $\mathbf{B}_i = (\mathbf{b}_{i1}, \mathbf{b}_{i2}, \ldots, \mathbf{b}_{iD})^T$, and then into a scoring matrix $\mathbf{B} = (\mathbf{B}_1^T, \mathbf{B}_2^T, \ldots, \mathbf{B}_I^T)^T$ for the whole test. (By definition, the score for a response in the zero category is zero, but other responses may also be scored zero.)

The probability of a response in category k of item i is modeled as

$$\Pr(\mathbf{X}_{ij} = 1; \mathbf{A}, \mathbf{B}, \xi | \theta) = \frac{\exp(\mathbf{b}_{ij}\theta + \mathbf{a}'_{ij}\xi)}{\sum_{k=1}^{K_i} \exp(\mathbf{b}_{ik}\theta + \mathbf{a}'_{ik}\xi)}. \quad (2)$$

And for a response vector we have

$$f(\mathbf{x}; \xi | \theta) = \Psi(\theta, \xi) \exp[\mathbf{x}'(\mathbf{B}\theta + \mathbf{A}\xi)] \quad (3)$$

with

$$\Psi(\theta, \xi) = \left\{ \sum_{z \in \Omega} \exp[\mathbf{z}^T(\mathbf{B}\theta + \mathbf{A}\xi)] \right\}^{-1} \quad (4)$$

where Ω is the set of all possible response vectors.

7.2 THE POPULATION MODEL

The item response model is a conditional model, in the sense that it describes the process of generating item responses conditional on the latent variable, θ**.** The complete definition of the TIMSS model, therefore, requires the specification of a density, $f_\theta(\theta;\alpha)$ for the latent variable θ. We use **a** to symbolize a set of parameters that characterize the distribution of θ**.** The most common practice when specifying unidimensional marginal item response models is to assume that the students have been sampled from a normal population with mean m and variance s^2. That is:

$$f_\theta(\theta;\alpha) \equiv f_\theta(\theta;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(\theta-\mu)^2}{2\sigma^2}\right] \tag{5}$$

or equivalently

$$\theta = \mu + E \tag{6}$$

where $E \sim N(0, \sigma^2)$.

A natural extension of (5) is to replace the mean, m with the regression model $\mathbf{Y}_n^T\beta$, where $\mathbf{Y}_n$ is a vector of u, fixed and known values for student n, and β is the corresponding vector of regression coefficients. For example, $\mathbf{Y}_n$ could be constituted of student variables such as gender, socio-economic status, or major. Then the population model for student n becomes

$$\theta_n = \mathbf{Y}_n^T\beta + E_n \tag{7}$$

where we assume that the E_n are independently and identically normally distributed with mean zero and variance s^2 so that (7) is equivalent to

$$f_\theta(\theta_n;\mathbf{Y}_n, \mathrm{b}, \sigma^2) = (2\pi\sigma^2)^{-1/2}\exp\left[-\frac{1}{2\sigma^2}(\theta_n - \mathbf{Y}_n^T\beta)^T(\theta_n - \mathbf{Y}_n^T\beta)\right] \tag{8}$$

a normal distribution with mean $\mathbf{Y}_n^T\beta$ and variance s^2. If (8) is used as the population model then the parameters to be estimated are ***b***, s^2 and **x.**

The TIMSS scaling model takes the generalization one step further by applying it to the vector valued θ rather than the scalar valued θ, resulting in the multivariate population model

$$f_\theta(\theta_n;\mathbf{W}_n, \gamma, \Sigma) = (2\pi)^{-d/2}|\Sigma|^{-1/2}\exp\left[-\frac{1}{2}(\theta_n - \gamma\mathbf{W}_n)^T\Sigma^{-1}(\theta_n - \gamma\mathbf{W}_n)\right] \tag{9}$$

where γ is a $u \times d$ matrix of regression coefficients, Σ is a $d \times d$ variance-covariance matrix and $\mathbf{W}_n$ is a $u \times 1$ vector of fixed variables. If (9) is used as the population model then the parameters to be estimated are γ, Σ and **x.** In TIMSS we refer to the $\mathbf{W}_n$ variables as conditioning variables.

7.3 ESTIMATION

The ConQuest software uses maximum likelihood methods to provide estimates of γ, Σ and $\mathbf{x}$. Combining the conditional item response model (3) and the population model (9) we obtain the unconditional or marginal response model

$$f(\mathbf{x};\xi,\gamma,\Sigma) = \int_{\theta} f_{\mathrm{x}}(\mathbf{x};\xi|\theta) f_{\theta}(\theta;\gamma,\Sigma)d\theta \tag{10}$$

and it follows that the likelihood is

$$\Lambda = \prod_{n=1}^{N} f_{\mathrm{x}}(\mathbf{x}_n;\xi,\gamma,\Sigma) \tag{11}$$

where N is the total number of sampled students.

Differentiating with respect to each of the parameters and defining the marginal posterior as

$$h_{\theta}(\theta_n;\mathbf{W}_n,\xi,\gamma,\Sigma|\mathbf{x}_n) = \frac{f_{\mathbf{x}}(\mathbf{x}_n\xi|\theta_{\mathrm{n}}) f_{\theta}(\theta_n;\mathbf{W}_n\gamma,\Sigma)}{f_{\mathbf{x}}(\mathbf{x}_n;\mathbf{W}_n,\xi,\gamma,\Sigma)} \tag{12}$$

provides the following system of likelihood equations:

$$\mathbf{A}'\sum_{n=1}^{N}\left[\mathbf{x}_n - \int_{\theta_n} \mathrm{E}_z(z|\theta_n)h_{\theta}(\theta_n;\mathbf{Y}_n,\xi,\gamma,\Sigma|\mathbf{x}_n)d\theta_n\right] = 0 \tag{13}$$

$$\hat{\gamma} = \left(\sum_{n=1}^{N}\bar{\theta}_n\mathbf{W}_n^T\right)\left(\sum_{n=1}^{N}\mathbf{W}_n\mathbf{W}_n^T\right)^{-1} \tag{14}$$

and

$$\hat{\Sigma} = \frac{1}{N}\sum_{\mathrm{n}=1}^{\mathrm{N}}\int_{\theta_{\mathrm{n}}}(\theta_n-\gamma\mathbf{W}_n)(\theta_n-\gamma\mathrm{W}_n)^T h_{\theta}(\theta_n;\mathbf{Y}_n,\xi,\gamma,\Sigma|\mathbf{x}_n)d\theta_n \tag{15}$$

where

$$\mathrm{E}_{\mathbf{z}}(\mathbf{z}|\theta_n) = \Psi(\theta_n,\xi)\sum_{\mathbf{z}\in\Omega}\mathbf{z}\exp[\mathbf{z}'(\mathbf{b}\theta_n+\mathbf{A}\xi)] \tag{16}$$

and

$$\bar{\theta}_n = \int_{\theta_n} \theta_n h_\theta(\theta_n;\mathbf{Y}_n, \xi, \gamma, \Sigma|\mathbf{x}_n) d\theta_n . \tag{17}$$

The system of equations defined by (13), (14), and (15) is solved using an EM algorithm.

7.3.1 Quadrature and Monte Carlo Approximations

The integrals in equations (13), (14), and (15) are approximated numerically using either quadrature or Monte Carlo methods. In each case we define, Θ_p , $p=1,...,P$ a set of P D-dimensional vectors (which we call nodes) and for each node we define a corresponding weight $W_p(\gamma, \Sigma)$. The marginal item response probability (10) is then approximated using

$$f_{\mathbf{x}}(\mathbf{x};\xi, \gamma, \Sigma) = \sum_{p=1}^{P} f_{\mathbf{x}}(\mathbf{x};\xi|\Theta_p) W_p(\gamma, \Sigma) \tag{18}$$

and the marginal posterior (12) is approximated using

$$h_\Theta(\Theta_q;\mathbf{W}_n, \xi, \gamma, \Sigma|\mathbf{x}_n) = \frac{f_{\mathbf{x}}(\mathbf{x}_n;\xi|\Theta_q) W_q(\gamma, \Sigma)}{\sum_{p=1}^{P} f_{\mathbf{x}}(\mathbf{x};\xi|\Theta_p) W_p(\gamma, \Sigma)} \tag{19}$$

for q=1,...,P.

The difference between the quadrature and Monte Carlo methods lies in the way the nodes and weights are prepared. For the quadrature case we begin by choosing a fixed set of Q points, $(\Theta_{d1}, \Theta_{d1}, \ldots, \Theta_{d1Q})$, for each latent dimension and then define a set of Q^D nodes that are indexed r = 1,…,Q^D, and are given by the Cartesian coordinates

$$\Theta_r = (\Theta_{1j_1}, \Theta_{2j_2}, \ldots, \Theta_{dj_d}) \text{ with } j_1 = 1,\ldots,\mathrm{Q};\ j_2 = 1,\ldots,\mathrm{Q};\ \ldots;\ \mathrm{j_d} = 1,\ldots,\mathrm{Q}.$$

The weights are then chosen to approximate the continuous latent population density (9), that is,

$$W_p = K(2\pi)^{-d/2}|\Sigma|^{-1/2}\exp\left[-\frac{1}{2}(\Theta_p - \gamma\mathbf{W}_n)^T\Sigma^{-1}(\Theta_p - \gamma\mathbf{W}_n)\right] \tag{20}$$

where K is a scaling factor to ensure that the sum of the weights is one.

In the Monte Carlo case the nodes are drawn at random from the standard multivariate normal distribution, and at each iteration the nodes are rotated using standard methods so that they become random draws from a multivariate normal distribution with mean $\gamma\mathbf{W}_n$ and variance Σ. In the Monte Carlo case the weight for all nodes is $1/P$.

7.3.2 Latent Estimation and Prediction

The marginal item response (10) does not include parameters for the latent values θ_n and hence the estimation algorithm does not result in estimates of the latent values. For TIMSS, the expected *a-posteriori* (EAP) prediction of each student's latent achievement was produced. The EAP prediction of the latent achievement for case n is

$$\theta_n^{EAP} = \sum_{r=1}^{p} \Theta_r h_{\Theta}(\Theta_r; \mathbf{W}_n, \hat{\xi}, \hat{\gamma}, \hat{\Sigma} | \mathbf{x}_n) .^{1} \tag{21}$$

Variance estimates for these predictions were estimated using

$$\mathrm{var}(\theta_n^{EAP}) = \sum_{r=1}^{P} (\Theta_r - \theta_n^{EAP})(\Theta_r - \theta_n^{EAP})^T h_{\Theta}(\Theta_r; \mathbf{W}_n, \hat{\xi}, \hat{\gamma}, \hat{\Sigma} | \mathbf{x}_n) . \tag{22}$$

7.3.3 Drawing Plausible Values

Plausible values are random draws from the marginal posterior of the latent distribution, (12), for each student. Unlike previously described methods for drawing plausible values ConQuest does not assume normality of the marginal posterior distributions. Recall from (12) that the marginal posterior is given by

$$h_{\theta}(\theta_n; \mathbf{W}_n, \xi, \gamma, \Sigma | \mathbf{x}_n) = \frac{f_{\mathbf{x}}(\mathbf{x}_n; \xi | \theta_n) f_{\theta}(\theta_n; \mathbf{W}_n, \gamma, \Sigma)}{\int_{\theta} f(\mathbf{x}; \xi | \theta) f_{\theta}(\theta, \gamma, \Sigma) d\theta} . \tag{23}$$

The ConQuest procedure begins by drawing M vector valued random deviates, $\{\varphi_{nm}\}_{m=1}^{M}$, from the multivariate normal distribution $f_{\theta}(\theta_n, \mathbf{W}_n \gamma, \Sigma)$ for each case n. These vectors are used to approximate the integral in the denominator of (23) using the Monte Carlo integration

$$\int_{\theta} f_{\mathbf{x}}(\mathbf{x}; \xi | \theta) f_{\theta}(\theta, \gamma, \Sigma) d\theta \approx \frac{1}{M} \sum_{m=1}^{M} f_{\mathbf{x}}(\mathbf{x}; \xi | \varphi_{mn}) \equiv \Im \tag{24}$$

At the same time the values

$$p_{mn} = f_{\mathbf{x}}(\mathbf{x}_n; \xi | \varphi_{mn}) f_{\theta}(\varphi_{mn}; \mathbf{W}_n, \gamma, \Sigma) \tag{25}$$

are calculated, so that we obtain the set of pairs $\left\langle \varphi_{nm}, \frac{p_{mn}}{\Im} \right\rangle_{m=1}^{M}$, which can be used as an approximation to the posterior density (23), and the probability that φ_{nm} could be drawn from this density is given by

[1] The current version of ConQuest uses the Monte Carlo method only when producing EAP predictions and variances for those predictions.

$$q_{nj} = \frac{p_{mn}}{\sum_{m=1}^{M} p_{mn}}. \tag{26}$$

At this point L uniformly distributed random numbers, $\{\eta_i\}_{i=1}^{L}$. are generated and for each random draw the vector φ_{n1_0} that satisfies the condition

$$\sum_{s=1}^{i_0-1} q_{sn} < \eta_i \le \sum_{s=1}^{i_0} q_{sn} \tag{27}$$

is selected as a plausible vector.

7.4 SCALING STEPS

The model was fitted to the data in two steps. First the items were calibrated using the combined data from most of the countries in the population. This was called the international calibration sample. In the second stage, the model was fitted separately for each country with the item parameters fixed at the values estimated in the first step.

7.4.1 Details of the Calibration Samples

The item calibration was carried out using almost the entire sample from each of the three areas- advanced mathematics, physics, and mathematics and science literacy- where students who attempted test booklets 1A and 1B formed the mathematics and science literacy calibration sample, those who did booklets 2A-2C formed the physics sample, and the students who took booklets 3A-3C made up the advanced mathematics sample. There was a further group of students who did booklet 4, which was a selection from all three topics, who were excluded from the calibration.

Six sets of item parameters were derived from these three samples. For mathematics and science literacy, a two- dimensional run was performed for mathematics literacy and science literacy. Because these scales are quite highly correlated (about .85) it was thought better to obtain the parameters from a two-dimensional run rather than two separate unidimensional runs. For another scale, the reasoning and social utility scale[2], which is composed of a subset of the mathematics and science literacy items and includes both mathematics and science literacy items, item parameters were also estimated from the mathematics and science literacy sample. Item parameters for full advanced mathematics and physics scales were obtained by unidimensional runs from their respective samples and item parameters for a 3-subscale model for advanced mathematics and a 5-subscale model for physics were also estimated.

[2] Results for the reasoning and social utility scale were not reported in the TIMSS international report, but scores on this scale are available in the TIMSS international database (Gonzalez, Smith, and Sibberns, 1998).

Table 7.1 shows the countries which were included in the calibration for the three subject areas and the size of the sample they contributed to the calibration sample.

Table 7.1 Countries and Numbers of Students in the Population 3 Calibration Samples

Country	Mathematics and Science Literacy	Advanced Mathematics	Physics
Australia	1844	548	564
Austria	1779	599	594
Canada	4832	2381	1967
Cyprus	473	330	307
Czech Republic	1899	833	819
Denmark	*	*	*
France	1590	796	835
Germany	2182	2189	616
Greece	*	346	349
Hungary	5091	–	–
Iceland	1703	–	–
Israel	*	*	*
Italy	1578	360	*
Latvia	–	–	708
Lithuania	2887	734	–
Netherlands	1470	–	–
New Zealand	1763	–	–
Norway	2518	–	1048
Russia	2289	1402	1129
Slovenia	1387	1301	512
South Africa	2757	–	–
Sweden	2816	749	760
Switzerland	2976	1072	1039
United States	5371	2349	2678
Total Sample	49205	15989	13925

(*) Administered test but not included in calibration sample.
(–) Did not participate in assessment.

Countries that were included in the study but wholly or partly omitted from the calibration samples for various reasons are indicated by an asterisk in the table above. For example, Italy was not used in the calibration sample, but was modelled in the second step of the scaling process. A dash indicates that the country did not participate in this part of the study. The table below shows the number of countries included in the study and the calibration samples.

Table 7.2 Number of Countries in TIMSS and in Calibration Samples

	Total	Mathematics and Science Literacy	Advanced Mathematics	Physics
In TIMSS	24	23	17	18
In Calibration	22	20	15	15

7.4.2 International Scaling Results

Tables 7.3 to 7.15 display basic statistics and item parameters, along with an indicator of the fit of each individual item parameter, for the scales derived from the six calibration runs described above. Most items were dichotomous, but 3- and 4 -category items were fitted with a partial-credit model. The item parameters here are given in the logit metric and in the item-step form described in Wu, Adams, and Wilson (1997). The mean square fit statistic is an index of the fit of the data to the assumed scaling model; the statistic given here was derived by Wu (1997). Under the null hypothesis that the data and model are consistent, the expected value of these statistics is one. Values that are less than one usually indicate items with greater than average discrimination, while values that are greater than one can result from lower than average discrimination, guessing, or some other deviation from the model.

Only some questions appeared in all booklets; for example, for advanced mathematics the I cluster items were given to all students, whereas the J, K, and L cluster items were each present in only one of the three booklets – 3A, 3B and 3C. Percent correct figures were calculated by summing the total of scores from all students who provided valid responses and dividing that by the number of students multiplied by the maximum score that could be achieved for that item. This reduces to the usual percent correct for the dichotomous items.

7.4.3 Fit of the Scaling Model

Tables 7.3 to 7.6 show the results for the overall advanced mathematics and physics scales and the mathematics literacy and science literacy scales. For the advanced mathematics scale (Table 7.3) items with fit statistics greater than or equal to 1.15 are J04, J12, J18, K08, K16, L16, and L18. Item J04, in Figure 7.1, seems to fit rather poorly, with markedly lower discrimination than the other items and a downward kink for some of the higher-ability students. This item proved to have a distractor with a positive biserial for several countries. Item J12 in Figure 7.2 shows some lower discrimination though not as dramatic as for J04, and also curvature in the response for lower- performing students. Figure 7. 3 demonstrates some lack of discrimination in item K08. No items were found to have fit statistics less than .85.

Table 7.3 Item Statistics and Parameter Estimates for the International Calibration Sample - Population 3 Advanced Mathematics Scale

Item Label	Number of Respondents in International Calibration Sample	Percentage of Correct Responses	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMMI01	15987	58.7	-0.585	0.017	0.89
CSMMI02	15987	57.0	-0.652	0.017	0.89
CSMMI03	15987	61.4	-0.669	0.017	1.07
CSMMI04	15988	57.7	-0.602	0.017	0.96
CSMMI05	15975	35.3	0.529	0.018	1.04
CSMMI06	15975	47.6	-0.200	0.017	0.93
CSMMI07	15986	57.8	-0.498	0.017	0.98
CSMMI08	15987	74.9	-1.578	0.020	0.95
CSMMI09	15986	59.7	-0.745	0.018	1.00
CSMMI10	15985	58.2	-0.460	0.017	1.02
CSMMJ01	5391	53.3	-0.564	0.030	0.89
CSMMJ02	4807	35.3	0.431	0.033	1.03
CSMMJ03	5390	56.6	-0.630	0.030	0.94
CSMMJ04	5391	39.2	0.477	0.030	1.16
CSMMJ05	5394	56.1	-0.686	0.030	0.85
CSMMJ06	5392	33.5	0.597	0.030	0.96
CSMMJ07	5390	41.5	0.103	0.029	0.96
CSMMJ08	4606	67.7	-0.880	0.032	1.08
CSMMJ09	5387	22.5	1.168	0.033	1.05
CSMMJ10	5388	36.8	0.447	0.030	1.03
CSMMJ11	5393	68.8	-1.063	0.032	1.14
CSMMJ12	5393	82.5	-1.752	0.037	1.38
CSMMJ13	5392	46.2	-0.023	0.029	1.03
CSMMJ14	5392	47.4	-0.304	0.029	0.87
CSSMJ15A	5393	48.4	-0.118	0.029	0.91
CSSMJ15B	5394	7.5	2.701	0.052	0.99
CSSMJ16A	5393	67.3	-0.974	0.031	1.03
CSSMJ16B	5394	22.6	1.269	0.034	0.91
CSSMJ17	5032	27.1	0.540	0.019	1.01
CSSMJ17 (S1)			1.382	0.051	1.12
CSEMJ18	5125	14.7	0.933	0.021	1.24
CSEMJ18 (S1)			2.385	0.087	0.94
CSEMJ19	5392	34.1	0.249	0.018	1.03
CSEMJ19 (S1)			1.279	0.046	0.98
CSMMK01	5296	82.1	-2.012	0.040	1.08
CSMMK02	5296	23.3	1.022	0.033	1.03
CSMMK03	5295	63.7	-0.831	0.031	1.06
CSMMK04	5297	27.8	0.928	0.032	0.98
CSMMK05	5297	43.0	0.136	0.030	1.08
CSMMK06	5297	51.4	-0.418	0.030	0.97

Table 7.3 Item Statistics and Parameter Estimates for the International Calibration Sample - Population 3 Advanced Mathematics Scale (Continued)

Item Label	Number of Respondents in International Calibration Sample	Percentage of Correct Responses	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMMK07	5297	53.4	-0.328	0.030	1.07
CSMMK08	5297	26.8	0.917	0.032	1.15
CSMMK09	5047	42.8	0.244	0.031	1.07
CSMMK10	5296	20.5	1.352	0.035	0.97
CSMMK11	5297	50.3	-0.061	0.030	0.96
CSSMK12	5295	50.2	-0.131	0.030	1.07
CSSMK13	5294	27.5	1.084	0.033	0.92
CSSMK14	5284	10.2	1.411	0.026	0.95
CSSMK14 (S1)			2.660	0.111	0.86
CSSMK15	5285	14.9	0.931	0.021	1.02
CSSMK15 (S1)			1.911	0.068	0.91
CSEMK16	5296	51.5	-0.173	0.014	1.27
CSEMK16 (S1)			0.538	0.029	0.94
CSEMK16 (S2)			-0.651	0.033	0.93
CSEMK17	5294	27.9	0.389	0.014	1.02
CSEMK17 (S1)			1.901	0.048	1.06
CSEMK17 (S2)			-0.167	0.068	1.03
CSEMK18	5294	36.2	0.157	0.018	1.10
CSEMK18 (S1)			0.897	0.040	0.93
CSMML01	5298	69.4	-1.226	0.033	1.04
CSMML02	5298	58.7	-0.689	0.030	0.93
CSMML03	5297	40.7	0.219	0.030	0.92
CSMML04	5297	45.0	-0.027	0.030	0.96
CSMML05	5298	41.1	0.107	0.030	0.92
CSMML06	5298	31.6	0.728	0.031	0.96
CSMML07	5297	30.7	0.637	0.031	1.02
CSMML08	5298	45.8	-0.075	0.030	1.02
CSMML09	5296	56.5	-0.340	0.030	1.01
CSMML10	5296	26.2	0.918	0.032	1.06
CSMML11	5298	74.1	-1.387	0.034	1.03
CSMML12	5298	63.9	-0.856	0.031	1.08
CSSML13	5294	25.6	0.987	0.033	0.95
CSSML14	5291	48.9	-0.132	0.030	0.88
CSSML15A	5297	48.4	-0.128	0.030	1.04
CSSML15B	5295	62.3	-0.757	0.031	1.03
CSEML16	5296	35.1	0.236	0.015	1.20
CSEML16 (S1)			0.295	0.030	1.05
CSEML16 (S2)			0.078	0.041	1.15
CSEML17	5296	17.7	0.892	0.021	1.04
CSEML17 (S1)			1.779	0.064	1.10
CSEML18	5292	49.8	-0.185	0.017	1.24
CSEML18 (S1)			1.755	0.057	1.08

Figure 7.1 Empirical and Modelled Item Characteristic Curves for Advanced Mathematics Population 3 Item: CSMMJ04. Fit MNSQ=1.16

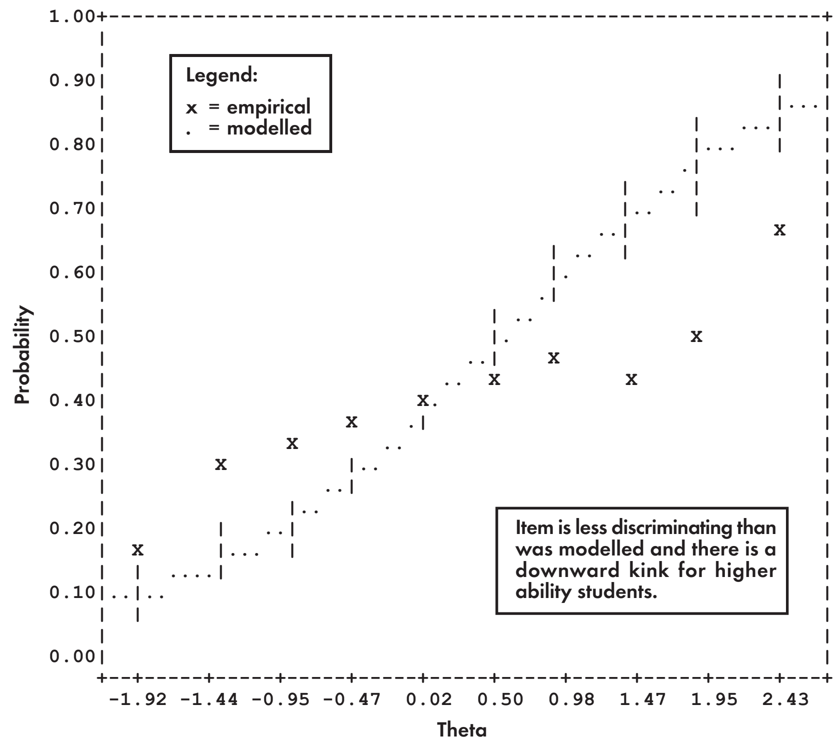


Figure 7.2 Empirical and Modelled Item Characteristic Curves for Advanced Mathematics Population 3 Item: CSMMJ12. Fit MNSQ=1.38

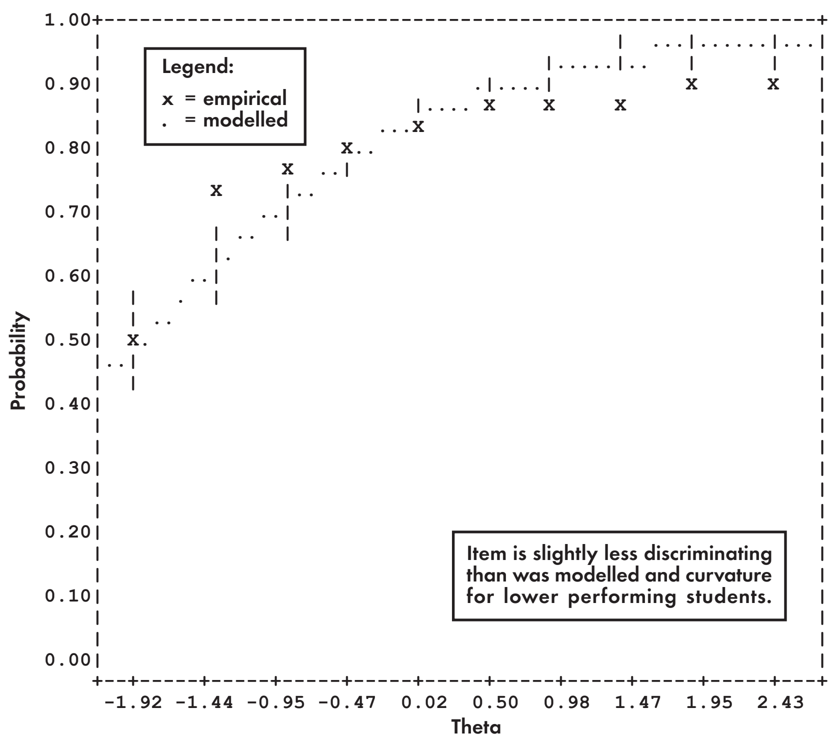
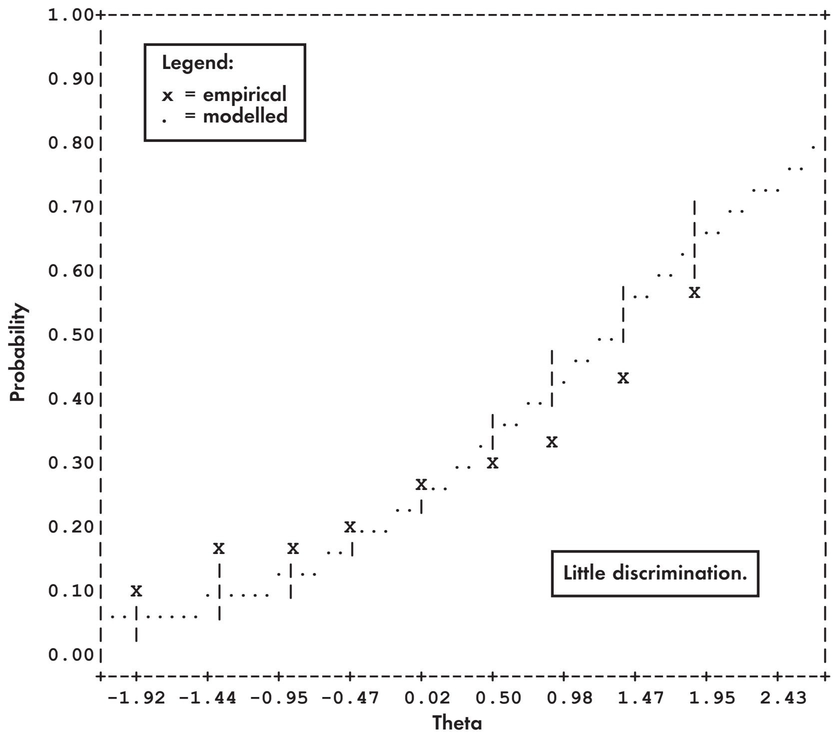


Figure 7.3 Empirical and Modelled Item Characteristic Curves for Advanced Mathematics Population 3 Item: CSMMK08. Fit MNSQ=1.15



For the physics scale (Table 7.4) items with fit statistics greater than or equal to 1.15 are F16, F17B, G09, G11 and H18. There were also four items with fit statistics less than .85. Figure 7.4 shows that item F17B is a relatively hard item, although it is not clear why it is so difficult. For item G09, Figure 7.5 shows a dip among some of the better-performing students. Further investigation showed a positive point-biserial correlation with one of the distractors in most of the countries. Four items with fit less than .85 were more discriminating than average, especially among the higher-ability students.

Table 7.4 Item Statistics and Parameter Estimates for the International Calibration Sample - Population 3 Physics Scale

Item Label	Number of Respondents in International Calibration Sample	Percentage of Correct Responses	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMPE01	13919	75.3	-1.773	0.020	1.10
CSMPE02	13921	58.9	-0.939	0.018	0.95
CSMPE03	13923	65.6	-1.408	0.019	0.95
CSMPE04	13925	84.4	-2.514	0.025	1.02
CSMPE05	13925	77.9	-2.034	0.022	1.03
CSMPE06	13925	34.6	0.083	0.019	1.07
CSMPE07	13924	45.8	-0.354	0.018	1.06
CSMPE08	13922	48.4	-0.469	0.018	1.07
CSMPE09	13919	35.0	0.083	0.019	1.00
CSMPE10	13917	45.0	-0.449	0.018	0.95
CSMPF01	4679	48.9	-0.523	0.031	1.12
CSMPF02	4679	16.8	1.097	0.039	1.11
CSMPF03	4679	39.0	-0.066	0.031	1.01
CSMPF04	4675	46.7	-0.592	0.031	0.93
CSMPF05	4675	60.3	-1.189	0.032	1.04
CSMPF06	4410	26.1	0.409	0.034	1.00
CSMPF07	4679	57.4	-0.808	0.031	1.09
CSMPF08	4679	43.3	-0.368	0.031	1.03
CSMPF09	4679	26.6	0.548	0.034	1.02
CSMPF10	4678	32.5	0.129	0.032	1.11
CSMPF11	4673	37.2	-0.050	0.031	0.95
CSSPF12	4670	15.9	0.666	0.024	0.95
CSSPF12 (S1)			1.037	0.052	1.06
CSSPF13	4671	61.6	-1.124	0.031	0.97
CSSPF14	4673	21.3	0.388	0.022	0.96
CSSPF14 (S1)			0.702	0.043	1.05
CSEPF15	4395	15.0	0.638	0.024	0.86
CSEPF15 (S1)			1.326	0.059	0.83
CSEPF16	4670	9.4	0.794	0.025	1.29
CSEPF16 (S1)			2.119	0.086	1.32
CSEPF17A	4669	25.7	0.297	0.033	1.04
CSEPF17B	4645	7.8	1.868	0.050	1.23
CSMPG01	4654	36.3	-0.123	0.032	1.11
CSMPG02	4654	65.3	-1.155	0.032	0.90
CSMPG03	4654	41.0	-0.177	0.031	1.17
CSMPG04	4654	32.7	0.148	0.032	1.09
CSMPG05	4651	37.0	0.021	0.032	0.98
CSMPG06	4652	60.1	-0.872	0.031	0.99
CSMPG07	4654	27.7	0.419	0.034	1.04
CSMPG08	4653	30.8	0.108	0.032	1.08
CSMPG09	4653	17.6	1.018	0.039	1.16
CSMPG10	4654	29.8	0.264	0.033	1.12

Table 7.4 Item Statistics and Parameter Estimates for the International Calibration Sample - Population 3 Physics Scale (Continued)

Item Label	Number of Respondents in International Calibration Sample	Percentage of Correct Responses	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSSPG11	4652	19.9	0.495	0.023	1.25
CSSPG11 (S1)			0.747	0.045	0.94
CSSPG12	4652	38.9	-0.304	0.019	0.99
CSSPG12 (S1)			0.898	0.042	1.00
CSSPG13	4652	30.2	0.087	0.032	0.95
CSSPG14	4652	24.2	0.613	0.035	0.83
CSSPG15	4652	15.6	1.292	0.042	0.80
CSSPG16	4444	35.2	0.171	0.028	1.14
CSSPG16 (S1)			-1.502	0.032	1.00
CSSPG17	4652	26.5	0.388	0.034	1.11
CSEPG18	4652	15.9	0.737	0.025	0.98
CSEPG18 (S1)			0.476	0.043	0.97
CSEPG19	4652	14.9	0.568	0.023	0.86
CSEPG19 (S1)			1.282	0.056	0.93
CSMPH01	4591	41.9	-0.086	0.032	1.12
CSMPH02	4592	52.2	-0.660	0.031	1.07
CSMPH03	4592	37.4	0.036	0.032	0.97
CSMPH04	4591	34.2	0.247	0.033	1.00
CSMPH05	4591	45.9	-0.309	0.031	1.03
CSMPH06	4591	31.4	0.394	0.034	1.00
CSMPH07	4591	35.2	-0.079	0.032	1.00
CSMPH08	4592	27.4	0.360	0.034	0.96
CSMPH09	4591	25.5	0.625	0.035	0.94
CSMPH10	4592	29.9	0.347	0.034	1.01
CSSPH12	4586	21.8	0.659	0.036	0.97
CSSPH13	4588	29.9	0.211	0.033	0.85
CSSPH14	4402	29.5	0.296	0.024	1.05
CSSPH14 (S1)			-0.478	0.033	1.01
CSSPH15	4588	25.6	0.705	0.036	0.86
CSSPH16	4585	20.6	0.322	0.021	1.04
CSSPH16 (S1)			1.711	0.064	0.83
CSEPH17	4587	15.1	0.535	0.023	1.13
CSEPH17 (S1)			1.978	0.076	1.00
CSEPH18	4587	24.7	0.378	0.023	1.17
CSEPH18 (S1)			-0.053	0.035	0.90
CSEPH19A	4412	28.0	0.258	0.022	0.93
CSEPH19A (S1)			0.308	0.039	0.93
CSEPH19B	4584	43.4	-0.276	0.032	0.92

Figure 7.4 Empirical and Modelled Item Characteristic Curves for Physics Population 3 Item: CSEPF17B. Fit MNSQ=1.23

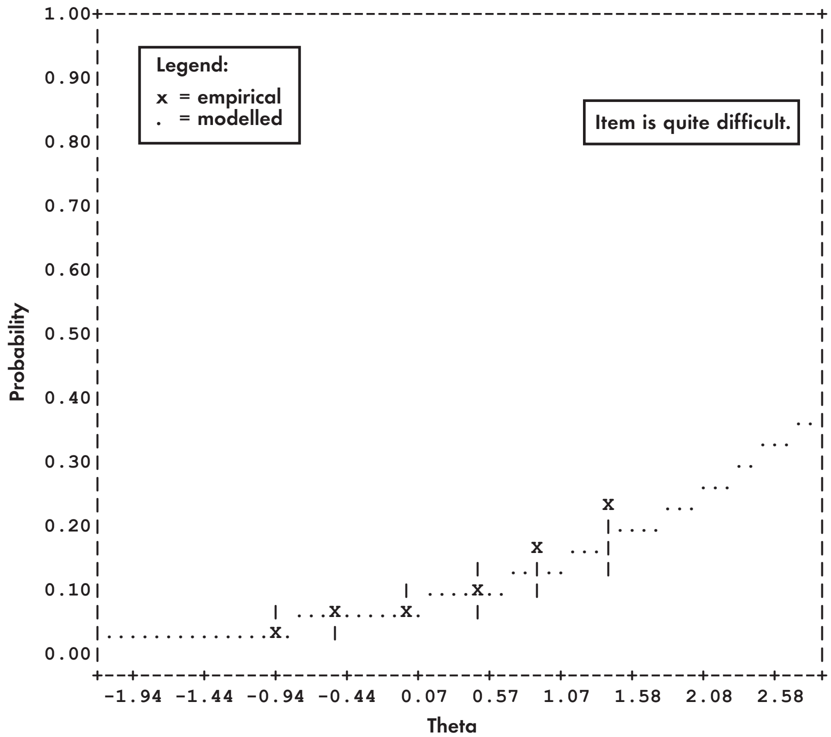
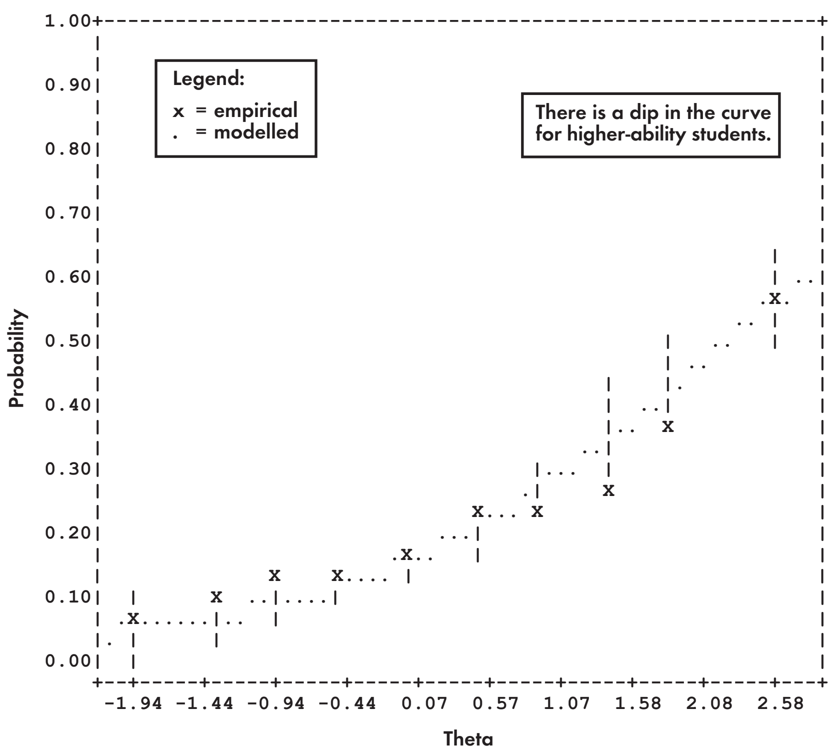


Figure 7.5 Empirical and Modelled Item Characteristic Curves for Physics Population 3 Item: CSMPG09. Fit MNSQ=1.16



For the mathematics literacy scale (Table 7.5), items with fit statistics greater than or equal to 1.15 are A04, A12, B20, B21 and B24. Figure 7.6, illustrating item A04, shows a tailing off of observed results compared with the modelled, for higher-ability students. Item B20, shown in Figure 7.7, exhibits a marked lack of discrimination, similar to J04 on the advanced mathematics scale. Both B21 and B24 in Figures 7.8 and 7.9 show similar behavior but the effect is smaller. Item A12, a partial credit item, consistently shows a low response in category 2 compared to categories 1 and 3 across the different countries. There are five items with fit statistics below .85. All of these show greater than average discrimination, but this sort of misfit was not deemed to be of concern.

Table 7.5 Item Statistics and Parameter Estimates for the International Calibration Sample - Population 3 Mathematics Literacy Scale

Item Label	Number of Respondents in International Calibration Sample	Percentage of Correct Responses	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMGA03	49191	63.8	-0.398	0.010	0.95
CSMGA04	49191	70.5	-0.797	0.011	1.18
CSMGA05	49188	48.8	0.344	0.010	0.99
CSSGA08	49170	50.8	0.254	0.006	1.04
CSSGA08 (S1)			0.538	0.012	1.09
CSEGA10	49182	33.1	1.144	0.007	1.09
CSEGA10 (S1)			-0.295	0.010	0.99
CSEGA12	49178	55.5	0.065	0.005	1.36
CSEGA12 (S1)			0.564	0.012	1.00
CSEGA12 (S2)			0.787	0.019	0.89
CSMGB14	49194	71.5	-0.859	0.011	1.05
CSMGB15	49199	58.5	-0.158	0.010	0.98
CSMGB16	49198	78.2	-1.344	0.012	1.00
CSMGB17	49197	47.4	0.384	0.010	0.86
CSMGB18	49189	37.0	0.938	0.010	1.02
CSMGB19	49196	71.6	-0.951	0.011	0.91
CSMGB20	49194	50.8	0.340	0.010	1.20
CSMGB21	44107	36.1	1.009	0.011	1.21
CSMGB22	49196	69.6	-0.747	0.011	0.95
CSMGB23	49194	53.1	0.164	0.010	1.05
CSMGB24	49193	42.0	0.725	0.010	1.17
CSSGB25	49196	36.2	0.947	0.010	0.98
CSSGB26	44094	39.7	0.629	0.006	1.09
CSSGB26 (S1)			1.816	0.021	1.12
CSMGC01	24789	68.1	-0.637	0.015	0.86
CSMGC02	24784	69.0	-0.716	0.015	0.87
CSMGC03	24781	60.5	-0.280	0.015	0.85
CSMGC04	24784	66.5	-0.563	0.015	0.92
CSMGC05	24535	70.4	-0.795	0.016	0.96
CSMGD13	24407	62.6	-0.399	0.015	0.89
CSMGD14	24404	63.9	-0.512	0.015	0.92
CSSGD15A	24405	73.7	-0.987	0.016	0.99
CSSGD15B	21915	59.0	-0.130	0.015	0.96
CSSGD16A	21916	39.2	0.841	0.015	0.93
CSSGD16B	21871	32.4	1.221	0.016	0.94
CSSGD17	24381	30.8	1.166	0.010	1.00
CSSGD17 (S1)			-0.178	0.015	1.04
CSMGC06	24790	79.4	-1.593	0.018	0.84
CSMGC07	24792	51.0	0.185	0.014	0.83
CSMGC08	24784	65.6	-0.516	0.015	0.92
CSMGC09	24790	70.7	-0.706	0.015	1.03
CSMGC11	24784	65.5	-0.613	0.015	0.81
CSSGC12	24783	24.1	1.742	0.016	0.97
CSSGC13	24788	21.5	1.909	0.017	0.90
CSMGD06	24402	63.4	-0.419	0.015	0.93
CSMGD07	24406	69.8	-0.818	0.015	0.98
CSMGD08	24406	53.6	0.066	0.014	0.93
CSMGD09	24407	69.6	-0.826	0.016	0.82
CSMGD10	24407	58.9	-0.199	0.014	0.96
CSMGD11	24406	43.4	0.553	0.014	0.88
CSMGD12	24184	28.8	1.338	0.016	0.84

Figure 7.6 **Empirical and Modelled Item Characteristic Curves for Mathematics Literacy Population 3 Item: CSMGA04 Fit MNSQ=1.18**

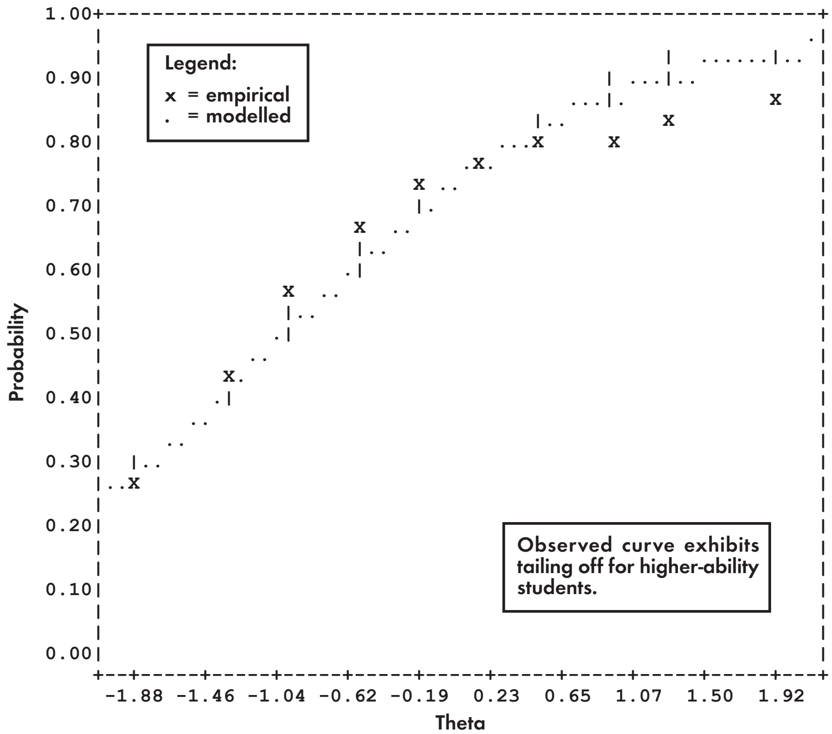


Figure 7.7 **Empirical and Modelled Item Characteristic Curves for Mathematics Literacy Population 3 Item: CSMGB20. Fit MNSQ=1.2**

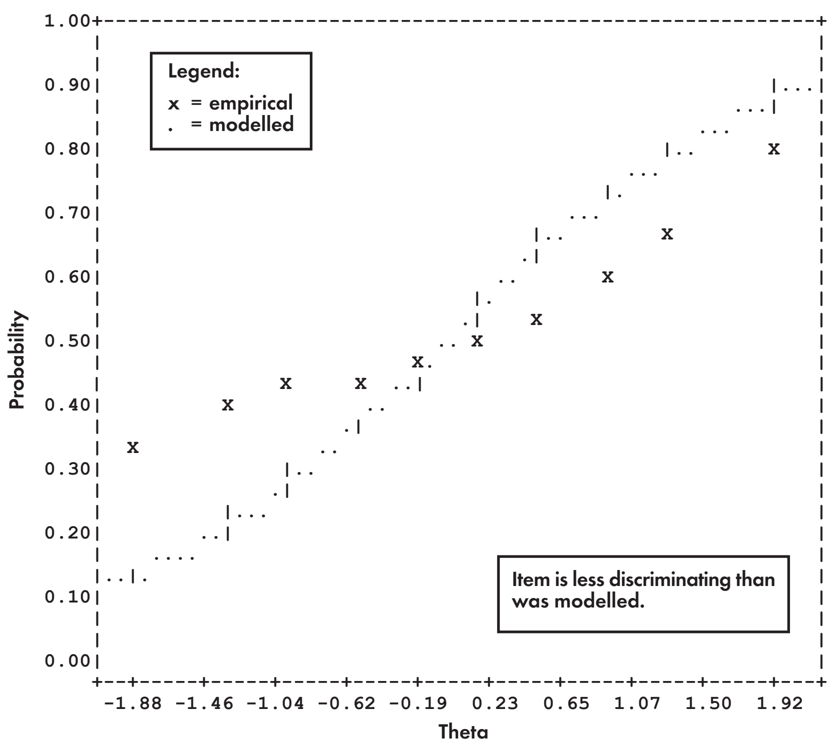


Figure 7.8 Empirical and Modelled Item Characteristic Curves for Mathematics Literacy Population 3 Item: CSMGB21. Fit MNSQ=1.21

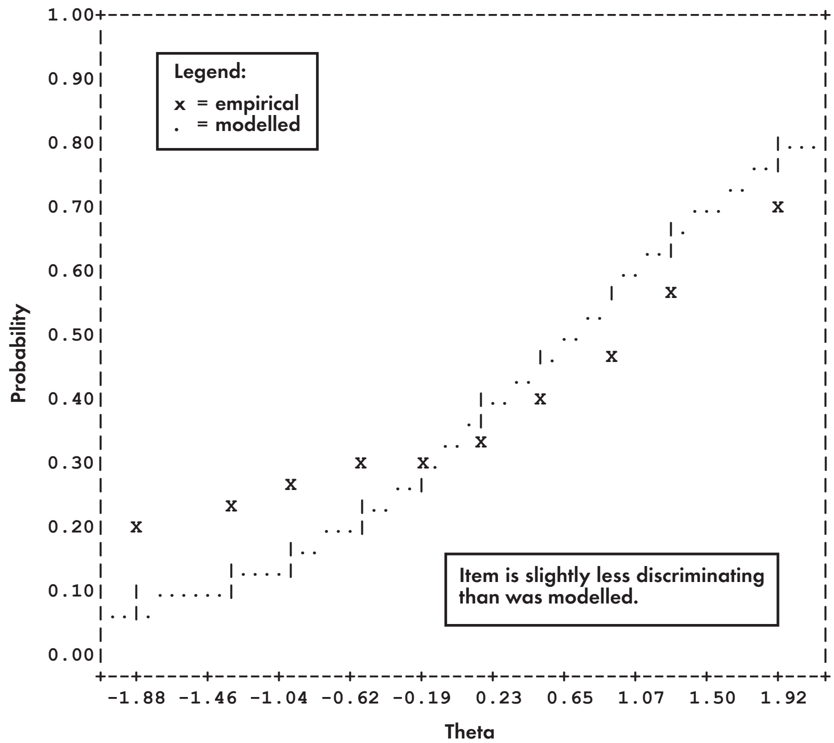
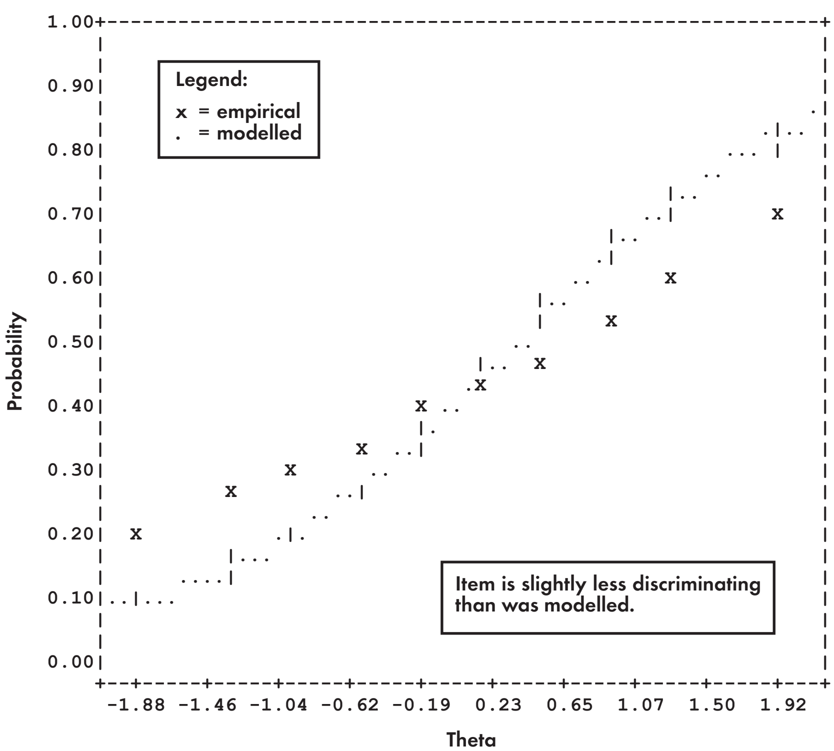


Figure 7.9 Empirical and Modelled Item Characteristic Curves for Mathematics Literacy Population 3 Item: CSMGB24. Fit MNSQ=1.17



For the science literacy scale (Table 7.6), items with fit statistics greater than or equal to 1.15 are B02, C20 and C21. Item B02 in Figure 7.10 shows evidence of less than usual discrimination. Both C20 and C21 also show slightly less than usual discrimination. There are no items with fit statistics less than .85 on this scale. This scale seemed to fit slightly better than the others.

Tables 7.7 through 7.14 present statistics for the reasoning and social utility (RSU) subscale, the three advanced mathematics subscales, and the five physics subscales. The fit statistics of the items on the subscales are very similar to the fit statistics for the overall scales, as would be expected.

Table 7.6 Item Statistics and Parameter Estimates for the International Calibration Sample - Population 3 Science Literacy Scale

Item Label	Number of Respondents in International Calibration Sample	Percentage of Correct Responses	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMGA01	49194	39.9	0.844	0.010	1.14
CSMGA02	49195	74.9	-1.060	0.011	1.00
CSSGA06A	49191	36.8	0.954	0.010	0.96
CSSGA06B	49174	43.0	0.606	0.010	0.95
CSSGA07	49186	49.4	0.296	0.006	1.04
CSSGA07 (S1)			0.404	0.011	1.04
CSEGA09B	49180	32.4	1.233	0.010	1.00
CSEGA11A	49191	72.2	-0.880	0.011	0.93
CSEGA11B	49187	60.1	-0.314	0.010	0.87
CSEGA11C	46101	43.1	0.545	0.010	0.97
CSMGB01	49196	65.0	-0.511	0.010	1.05
CSMGB02	49197	85.0	-1.625	0.013	1.19
CSMGB03	49200	60.6	-0.112	0.010	1.08
CSMGB04	47607	54.6	-0.011	0.010	1.07
CSMGB05	49198	62.6	-0.311	0.010	1.08
CSMGB06	44629	31.7	1.257	0.011	0.99
CSMGB07	49198	91.3	-2.347	0.016	1.05
CSMGB08	44109	71.4	-0.704	0.011	1.10
CSMGB09	49196	30.3	1.261	0.011	1.10
CSMGB10	49196	49.7	0.281	0.010	0.96
CSMGB11	49195	54.1	0.154	0.010	1.03
CSSGB12	49195	32.7	1.143	0.010	0.89
CSSGB13	49193	81.6	-1.451	0.012	1.03
CSMGC14	24784	66.5	-0.455	0.015	1.00
CSMGC15	24787	73.3	-0.883	0.016	0.96
CSMGC16	24787	77.3	-1.160	0.016	0.89
CSMGC17	24781	57.3	-0.041	0.014	0.93
CSSGC18	24789	32.1	1.226	0.015	0.85
CSSGC19	24788	42.0	0.590	0.008	0.96
CSSGC19 (S1)			1.210	0.021	1.07
CSSGC20	22185	30.1	0.988	0.009	1.21
CSSGC20 (S1)			1.516	0.026	0.95
CSEGC21	24781	25.4	1.515	0.011	1.15
CSEGC21 (S1)			-0.473	0.014	1.04
CSMGD01	24405	85.3	-1.803	0.020	1.02
CSSGD02	24403	57.0	-0.005	0.014	1.01
CSSGD03	24405	66.8	-0.504	0.015	0.97
CSEGD04	24397	20.2	1.860	0.017	0.87
CSEGD05A	24407	72.4	-0.838	0.015	0.98
CSEGD05B	24407	53.1	0.263	0.014	0.99

Figure 7.10 Empirical and Modelled Item Characteristic Curves for Science Literacy Population 3 Item: CSMGB02. Fit MNSQ=1.19

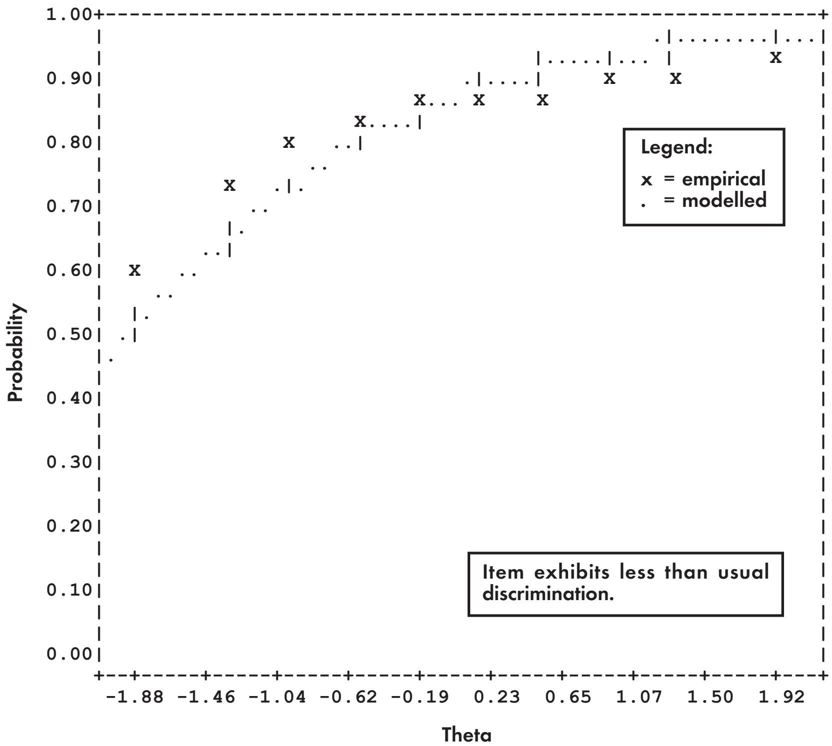


Table 7.7 Item Statistics and Parameter Estimates for the International Calibration Sample - Population 3 Reasoning and Social Utility Scale

Item Label	Number of Respondents in International Calibration Sample	Percentage of Correct Responses	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMGA01	49194	39.9	0.659	0.010	1.13
CSMGA02	49195	74.9	-1.293	0.012	1.01
CSMGA03	49191	63.8	-0.556	0.010	0.95
CSMGA04	49191	70.5	-0.945	0.011	1.12
CSMGA05	49188	48.8	0.168	0.010	1.00
CSSGA06A	49191	36.8	0.775	0.010	0.96
CSSGA06B	49174	43.0	0.415	0.010	0.95
CSSGA07	49186	49.4	0.103	0.006	1.03
CSSGA07 (S1)			0.367	0.011	1.02
CSSGA08	49170	50.8	0.083	0.006	0.93
CSSGA08 (S1)			0.582	0.012	1.07
CSEGA09B	49180	32.4	1.048	0.011	0.96
CSEGA10	49182	33.1	0.917	0.007	1.03
CSEGA10 (S1)			-0.234	0.010	0.97
CSEGA11A	49191	72.2	-1.118	0.011	0.96
CSEGA11B	49187	60.1	-0.532	0.010	0.87
CSEGA11C	46101	43.1	0.363	0.010	0.95
CSEGA12	49178	55.5	-0.802	0.005	1.14
CSEGA12 (S1)			0.610	0.012	0.95
CSEGA12 (S2)			0.805	0.019	0.87

Table 7.8 Parameter Estimates for the International Calibration Sample Population 3 Numbers and Equations Scale:

Item Label	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMMI01	-0.430	0.018	0.91
CSMMI02	-0.499	0.018	0.92
CSMMI03	-0.516	0.018	1.12
CSMMJ01	-0.406	0.030	0.88
CSMMJ02	0.627	0.034	1.01
CSMMJ03	-0.472	0.030	0.97
CSMMJ04	0.671	0.030	1.14
CSMMK01	-1.897	0.040	1.09
CSMMK02	1.205	0.033	1.00
CSSMK13	1.256	0.034	0.94
CSSMK15	1.137	0.021	1.04
CSSMK15 (S1)	1.874	0.068	0.89
CSEMK16	-0.019	0.015	1.37
CSEMK16 (S1)	0.485	0.029	0.97
CSEMK16 (S2)	-0.650	0.033	0.94
CSMML01	-1.101	0.033	1.01
CSMML02	-0.543	0.031	0.91
CSMML03	0.399	0.030	0.93
CSMML04	0.146	0.030	1.02
CSEML16	0.441	0.015	1.14
CSEML16 (S1)	0.220	0.031	1.05
CSEML16 (S2)	0.074	0.041	1.16

Table 7.9 Parameter Estimates for the International Calibration Sample Population 3 Calculus Scale

Item Label	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMMI04	-0.878	0.018	0.98
CSMMI06	-0.452	0.018	0.96
CSMMJ05	-0.966	0.031	0.87
CSMMJ06	0.399	0.031	1.03
CSMMJ14	-0.573	0.030	0.90
CSSMJ15A	-0.375	0.030	0.93
CSSMJ15B	2.600	0.053	0.92
CSSMJ17	0.373	0.020	1.14
CSSMJ17 (S1)	1.290	0.052	1.13
CSMMK03	-1.123	0.032	1.11
CSMMK04	0.722	0.033	1.01
CSMMK05	-0.109	0.030	1.13
CSMMK06	-0.692	0.030	1.05
CSEMK17	0.195	0.015	0.97
CSEMK17 (S1)	1.802	0.048	1.05
CSEMK17 (S2)	-0.165	0.068	1.00
CSMML05	-0.118	0.031	0.93
CSMML06	0.546	0.032	1.01
CSMML07	0.452	0.032	1.04

Table 7.10 Parameter Estimates for the International Calibration Sample Population 3 Geometry Scale

Item Label	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMMI07	-0.551	0.017	0.99
CSMMI08	-1.618	0.020	0.95
CSMMI09	-0.796	0.017	0.99
CSMMJ07	0.041	0.029	0.94
CSMMJ08	-0.933	0.032	1.04
CSMMJ09	1.095	0.033	1.05
CSMMJ10	0.377	0.030	1.04
CSMMJ11	-1.120	0.031	1.12
CSSMJ16A	-1.034	0.031	1.00
CSSMJ16B	1.190	0.034	0.88
CSEMJ19	0.173	0.017	0.99
CSEMJ19 (S1)	1.295	0.046	0.98
CSMMK07	-0.389	0.029	1.04
CSMMK08	0.837	0.032	1.09
CSMMK09	0.172	0.031	1.04
CSMMK10	1.264	0.035	0.92
CSSMK12	-0.199	0.029	1.03
CSSMK14	1.297	0.026	0.92
CSSMK14 (S1)	2.687	0.111	0.84
CSEMK18	0.082	0.018	0.96
CSEMK18 (S1)	0.922	0.040	0.91
CSMML08	-0.124	0.029	0.99
CSMML09	-0.385	0.030	1.00
CSMML12	-0.895	0.031	1.09
CSSML13	0.932	0.033	0.91
CSEML17	0.823	0.020	1.03
CSEML17 (S1)	1.788	0.064	1.09
CSEML18	-0.237	0.017	1.12
CSEML18 (S1)	1.771	0.057	1.06

Table 7.11 Parameter Estimates for the International Calibration Sample Population 3 Mechanics Scale

Item Label	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMPE03	-1.618	0.019	1.01
CSMPE05	-2.264	0.022	1.06
CSMPF02	0.979	0.040	0.99
CSMPF04	-0.775	0.031	0.93
CSMPF10	-0.026	0.033	1.12
CSEPF17A	0.146	0.033	0.97
CSEPF17B	1.761	0.051	1.12
CSMPG07	0.270	0.035	1.02
CSMPG08	-0.051	0.033	1.06
CSMPG09	0.891	0.039	1.12
CSSPG12	-0.466	0.019	1.01
CSSPG12 (S1)	0.836	0.042	1.00
CSSPG15	1.176	0.043	0.78
CSSPG16	0.029	0.028	1.18
CSSPG16 (S1)	-1.583	0.032	1.05
CSMPH01	-0.237	0.033	1.23
CSMPH04	0.112	0.034	0.96
CSSPH13	0.073	0.034	0.92

Table 7.12 Parameter Estimates for the International Calibration Sample Population 3 Electricity and Magnetism Scale

Item Label	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMPE04	-2.639	0.025	1.05
CSMPE06	-0.034	0.019	1.05
CSMPE09	-0.035	0.019	1.00
CSMPF06	0.296	0.035	0.96
CSMPF08	-0.487	0.031	1.01
CSSPF14	0.276	0.022	0.94
CSSPF14 (S1)	0.692	0.043	1.04
CSEPF16	0.687	0.026	1.20
CSEPF16 (S1)	2.107	0.086	1.30
CSMPG01	-0.248	0.032	1.09
CSMPG04	0.024	0.033	1.09
CSSPG17	0.269	0.034	1.10
CSEPG19	0.453	0.023	0.88
CSEPG19 (S1)	1.272	0.056	0.90
CSMPH06	0.290	0.034	1.04
CSMPH08	0.255	0.034	0.96
CSMPH10	0.242	0.034	0.95
CSSPH16	0.217	0.021	0.85
CSSPH16 (S1)	1.699	0.064	0.81
CSEPH17	0.434	0.023	1.09
CSEPH17 (S1)	1.965	0.076	0.98

Table 7.13 Parameter Estimates for the International Calibration Sample Population 3 Heat Scale

Item Label	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMPE08	-0.195	0.018	1.03
CSMPF05	-0.906	0.031	1.02
CSSPF12	0.873	0.023	0.87
CSSPF12 (S1)	1.092	0.052	1.05
CSMPG02	-0.882	0.032	0.91
CSMPG03	0.073	0.031	1.15
CSSPG11	0.708	0.022	1.09
CSSPG11 (S1)	0.799	0.045	0.94
CSMPH02	-0.386	0.031	1.06
CSMPH07	0.180	0.032	0.96
CSSPH14	0.535	0.023	1.02
CSSPH14 (S1)	-0.423	0.033	1.01

Table 7.14 Parameter Estimates for the International Calibration Sample Population 3 Wave Phenomena Scale

Item Label	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMPE01	-1.613	0.021	1.13
CSMPE10	-0.238	0.018	0.97
CSMPF01	-0.312	0.031	1.11
CSMPF11	0.179	0.032	0.95
CSSPF13	-0.938	0.032	1.02
CSMPG05	0.253	0.033	1.09
CSSPG13	0.322	0.033	1.03
CSMPH09	0.905	0.036	1.01
CSSPH12	0.940	0.036	0.96
CSEPH19A	0.546	0.023	0.98
CSEPH19A (S1)	0.221	0.039	0.93
CSEPH19B	-0.042	0.032	0.90

Table 7.15 Parameter Estimates for the International Calibration Sample Population 3 Particle, Quantum, Astrophysics, and Relativity Scale

Item Label	Difficulty Estimate in Logit Metric	Asymptotic Standard Error in Logit Metric	Mean Square Fit Statistic
CSMPE02	-0.993	0.018	0.97
CSMPE07	-0.401	0.018	1.03
CSMPF03	-0.111	0.031	0.99
CSMPF07	-0.861	0.031	1.10
CSMPF09	0.508	0.034	1.00
CSEPF15	0.612	0.024	0.90
CSEPF15 (S1)	1.307	0.060	0.82
CSMPG06	-0.934	0.032	1.03
CSMPG10	0.217	0.033	1.15
CSSPG14	0.570	0.035	0.84
CSEPG18	0.716	0.025	1.00
CSEPG18 (S1)	0.451	0.043	0.97
CSMPH03	-0.003	0.033	0.99
CSMPH05	-0.350	0.032	1.07
CSSPH15	0.677	0.036	0.87
CSEPH18	0.354	0.023	1.18
CSEPH18 (S1)	-0.077	0.035	0.91

7.4.4 The Population Model For Population 3

The population model equation (9) specifies that the latent variable θ has a distribution that is partly a function of a range of background variables. In order to derive reliable proficiency estimates, therefore, it is necessary to condition on these background variables before drawing the plausible values. A large set of background variables was used in the conditioning, including all of the questions from the student questionnaire. The information in these student variables was summarized through a principal components analysis in order to avoid multicollinearity problems and to keep the number of variables in the conditioning to a manageable level. A principal component analysis was run for each country on all students and as many components retained as explained 90 percent of the variance. Table 7.16 shows the number of components for each country. For the principal components analysis each student variable was recoded into a set of dummy variables which represented all categories of the variable as well as a missing data indicator.

For all scaling runs the variable sex was used as a conditioning variable. Additionally, preliminary national scores in mathematics and science literacy, reasoning and social utility (RSU), advanced mathematics, and physics were computed for each country using basic Rasch scaling methodology. These national scores were used in the conditioning process. As may be seen from Table 7.17, conditioning for the mathematics and science literacy scales included sex of student, the advanced mathematics national score, the physics national score, the school mean on the mathematics and science literacy national score (mathematics and science literacy and RSU combined), the principal components of the questionnaire variables, and the product of the mathematics and science literacy school mean and the principal components. Conditioning for the RSU scale was very similar, except that the RSU national score was substituted for the mathematics and science literacy national score. For advanced mathematics, the sex of student, the physics national score, the mathematics and science literacy national score (excluding RSU), the school mean on the advanced mathematics national score, the principal components, and the product of the school mean on the advanced mathematics score and the principal components. The physics conditioning was similar, and included the sex of student, the advanced mathematics national score, the mathematics and science literacy national score (excluding RSU), the school mean on the physics national score, the principal components, and the product of the physics score and the principal components.

Table 7.16 Number of Principal Components Retained in Conditioning - Population 3

Country	Retained Components
Australia	66
Austria	84
Canada	81
Cyprus	103
Czech Republic	90
Denmark	81
France	68
Germany	60
Greece	74
Hungary	103
Iceland	70
Israel	87
Italy	96
Latvia	82
Lithuania	91
Netherlands	60
New Zealand	78
Norway	81
Russia	113
Slovenia	90
South Africa	128
Sweden	79
Switzerland	91
United States	71

Table 7.17 Variables Used in Conditioning - Population 3

Variables	Mathematics and Science Literacy	RSU	Mathematics	Physics
Sex	Y	Y	Y	Y
Advanced Mathematics Score	Y	Y	N	Y
Physics Score	Y	Y	Y	N
Mathematics and Science Literacy	N	N	Y	Y
RSU Score	N	N	N	N
School Mean Advanced Mathematics	N	N	Y	N
School Mean Physics Score	N	N	N	Y
School Mean Mathematics and Science Literacy/RSU Score	Y	N	N	N
School Mean RSU Score	N	Y	N	N
Principal Components	Y	Y	Y	Y
Principal Components by School Mean Advanced Mathematics Score	N	N	Y	N
Principal Components by School Mean Physics Score	N	N	N	Y
Principal Components by School Mean Mathematics and Science Literacy	Y	N	N	N
Principal Components by School Mean RSU Score	N	Y	N	N

REFERENCES

Adams, R.J., Wilson, M.R., and Wang, W.C. (1997). The multidimensional random coefficients multinomial logit. *Applied Psychological Measurement, 21,* 1-24.

Adams, R.J., Wilson, M.R., and Wu, M.L. (1997). Multilevel item responses models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics, 22 (1),* 46-75.

Adams, R.J., Wu, M.L., and Macaskill G. (1997). Scaling methodology and procedures for the mathematics and science scales. In M.O. Martin and D.L. Kelly (Eds.), *TIMSS technical report, volume II: Implementation and analysis – primary and middle school years.* Chestnut Hill, MA: Boston College.

Gonzalez, E. J., Smith, T. A., and Sibberns, H. (Eds.). (1998). *User guide for the TIMSS international database: Final year of secondary school – 1995 assessment*. Chestnut Hill, MA.: Boston College.

Wu, M.L. (1997). *The development and application of a fit test for use with marginal estimation and generalized item response models.* Unpublished Master's dissertation, University of Melbourne.

Wu, M.L., Adams, R.J., and Wilson, M. (1997). *Conquest: Generalized item response modelling software - Manual.* Melbourne: Australian Council for Educational Research.

8 Reporting Student Achievement in Mathematics and Science Literacy, Advanced Mathematics, and Physics

Eugenio J. Gonzalez
Boston College

8.1 STANDARDIZING THE TIMSS INTERNATIONAL SCALE SCORES

The item response theory (IRT) scaling procedures described in Chapter 7 yielded imputed proficiency scores ("plausible values") in a logit metric, with the majority of scores falling in the range from -3 to +3. These scores were transformed onto an international achievement scale with mean 500 and standard deviation 100 – a scale that was more suited to reporting international results. This scale avoids negative values for student scale scores and eliminates the need for decimal points in reporting student achievement.

Since a plausible value is an imputed score that includes a random component, it is customary when using this method to draw a number of plausible values for each respondent (usually five). Each analysis is then carried out five times, once with each plausible value, and the results are averaged to get the best overall result. The variability among the five results is a measure of the error due to imputation and, where it is large, may be combined with jackknife estimates of sampling error to give a more realistic indication of the total variability of a statistic. Since the TIMSS final year of secondary school population (Population 3) showed significant variability between results from the five plausible values, it was decided to incorporate this variation in the analytic procedures.

In order to ensure that the mean of the TIMSS international achievement scale was close to the average student achievement level across countries, it was necessary to estimate the mean and standard deviation of the logit scores for all participating students. To accomplish this, the logit scores for all students from all countries were combined into a standardization sample. Each country was equally weighted. The means and standard deviations derived from this procedure are shown in Tables 8.1 through 8.12. These tables show the average logit for each of the five plausible values.

Table 8.1 Standardization Parameters of International Mathematics Literacy Scores

Scale	Mean Logit	Standard Deviation
Mathematics Literacy Plausible Value #1	0.3490	1.1086
Mathematics Literacy Plausible Value #2	0.3503	1.1012
Mathematics Literacy Plausible Value #3	0.3495	1.1027
Mathematics Literacy Plausible Value #4	0.3507	1.1038
Mathematics Literacy Plausible Value #5	0.3489	1.1040

Table 8.2 Standardization Parameters of International Science Literacy Scores

Scale	Mean Logit	Standard Deviation
Science Literacy Plausible Value #1	0.3393	0.9421
Science Literacy Plausible Value #2	0.3439	0.9407
Science Literacy Plausible Value #3	0.3425	0.9423
Science Literacy Plausible Value #4	0.3417	0.9435
Science Literacy Plausible Value #5	0.3414	0.9405

Table 8.3 Standardization Parameters of International Advanced Mathematics Scores

Scale	Mean Logit	Standard Deviation
Advanced Mathematics Plausible Value #1	-0.1156	0.8664
Advanced Mathematics Plausible Value #2	-0.1195	0.8657
Advanced Mathematics Plausible Value #3	-0.1134	0.8674
Advanced Mathematics Plausible Value #4	-0.1163	0.8684
Advanced Mathematics Plausible Value #5	-0.1191	0.8699

Table 8.4 Standardization Parameters of International Numbers and Equations Scores

Scale	Mean Logit	Standard Deviation
Numbers and Equations Plausible Value #1	0.0450	1.0782
Numbers and Equations Plausible Value #2	0.0567	1.0787
Numbers and Equations Plausible Value #3	0.0490	1.0788
Numbers and Equations Plausible Value #4	0.0552	1.0751
Numbers and Equations Plausible Value #5	0.0559	1.0817

Table 8.5 Standardization Parameters of International Calculus Scores

Scale	Mean Logit	Standard Deviation
Calculus Plausible Value #1	-0.3704	1.1983
Calculus Plausible Value #2	-0.3608	1.2005
Calculus Plausible Value #3	-0.3644	1.1984
Calculus Plausible Value #4	-0.3604	1.2015
Calculus Plausible Value #5	-0.3590	1.2062

Table 8.6 Standardization Parameters of International Geometry Scores

Scale	Mean Logit	Standard Deviation
Geometry Plausible Value #1	-0.1862	0.9357
Geometry Plausible Value #2	-0.1790	0.9334
Geometry Plausible Value #3	-0.1837	0.9345
Geometry Plausible Value #4	-0.1781	0.9327
Geometry Plausible Value #5	-0.1789	0.9371

Table 8.7 Standardization Parameters of International Physics Scores

Scale	Mean Logit	Standard Deviation
Physics Plausible Value #1	-0.5506	0.7215
Physics Plausible Value #2	-0.5457	0.7247
Physics Plausible Value #3	-0.5464	0.7240
Physics Plausible Value #4	-0.5505	0.7255
Physics Plausible Value #5	-0.5477	0.7249

Table 8.8 Standardization Parameters of International Mechanics Scores

Scale	Mean Logit	Standard Deviation
Mechanics Plausible Value #1	-0.7019	1.0645
Mechanics Plausible Value #2	-0.7052	1.0630
Mechanics Plausible Value #3	-0.7056	1.0599
Mechanics Plausible Value #4	-0.6994	1.0638
Mechanics Plausible Value #5	-0.7036	1.0636

Table 8.9 Standardization Parameters of International Electricity and Magnetism Scores

Scale	Mean Logit	Standard Deviation
Electricity and Magnetism Plausible Value #1	-0.6917	0.8441
Electricity and Magnetism Plausible Value #2	-0.6994	0.8490
Electricity and Magnetism Plausible Value #3	-0.6960	0.8472
Electricity and Magnetism Plausible Value #4	-0.6903	0.8482
Electricity and Magnetism Plausible Value #5	-0.6968	0.8455

Table 8.10 Standardization Parameters of International Heat Scores

Scale	Mean Logit	Standard Deviation
Heat Plausible Value #1	-0.3200	0.9414
Heat Plausible Value #2	-0.3243	0.9458
Heat Plausible Value #3	-0.3203	0.9432
Heat Plausible Value #4	-0.3183	0.9472
Heat Plausible Value #5	-0.3238	0.9405

Table 8.11 Standardization Parameters of International Wave Phenomena Scores

Scale	Mean Logit	Standard Deviation
Wave Phenomena Plausible Value #1	-0.3260	1.0758
Wave Phenomena Plausible Value #2	-0.3288	1.0774
Wave Phenomena Plausible Value #3	-0.3317	1.0711
Wave Phenomena Plausible Value #4	-0.3226	1.0753
Wave Phenomena Plausible Value #5	-0.3316	1.0750

Table 8.12 Standardization Parameters of International Particle, Quantum, Astrophysics and Relativity Scores

Scale	Mean Logit	Standard Deviation
Particle, Quantum, Astrophysics & Relativity Plausible Value #1	-0.6179	0.9492
Particle, Quantum, Astrophysics & Relativity Plausible Value #2	-0.6199	0.9469
Particle, Quantum, Astrophysics & Relativity Plausible Value #3	-0.6205	0.9439
Particle, Quantum, Astrophysics & Relativity Plausible Value #4	-0.6174	0.9466
Particle, Quantum, Astrophysics & Relativity Plausible Value #5	-0.6220	0.9406

Each country was weighted to contribute equally to the calculation of the international mean and standard deviation. The transformation applied to the plausible value logit scores was

$$S_{ijk} = 500 + 100 * \left(\frac{\theta_{ijk} - \bar{\theta}_j}{SD_{\theta_j}}\right)$$

where S_{ijk} is the standardized scale score with mean 500 and standard deviation 100 for student *i*, in plausible value *j*, in country *k*; θ_{ijk} is the logit score for the same student, $\bar{\theta}_j$ is the weighted average across all countries on plausible value *j*, and SD_{θ_j} is the standard deviation across all countries on plausible value *j*. Since five plausible values (logit scores) were drawn for each student, each of these was transformed so that the international mean of the result scores was 500, with standard deviation 100.

Because plausible values are actually random draws from the estimated distribution of student achievement and not actual student scores, student proficiency estimates were occasionally obtained that were unusually high or low. Where a transformed plausible value fell below 10, the value was recoded to 10, making 10 the lowest score on the transformed scale. This happened in very few cases across the countries. Where a transformed plausible value surpassed 990, the value was recoded to 990, making 990 the highest score on the transformed scale.

8.2 STANDARDIZING THE INTERNATIONAL ITEM DIFFICULTIES

To help readers of the TIMSS international reports understand the international achievement scales, TIMSS produced item difficulty maps that showed the location on the scales of several items from the subject matter content areas covered by the mathematics and science tests. In order to locate the example items on the achievement scales, the item difficulty parameter for each item had to be transformed from its original logit metric to the metric of the international achievement scales (a mean of 500 and standard deviation of 100).

The procedure for deriving the international item difficulties is described in Chapter 7. The international item difficulties obtained from the scaling procedure represent the proficiency level of a person who has a 50 percent chance of responding to the item correctly. For the item difficulty maps it was preferred that the difficulty correspond to the proficiency level of a person showing greater mastery of the item. For this reason it was decided to calibrate these item difficulties in terms of the proficiency of a person with a 65 percent chance of responding correctly.

In order to derive the item difficulties for the item difficulty maps, the original item difficulties obtained from the scaling procedure were transformed in two ways. First they were moved along the logit scale from the point where a student with that proficiency would have a 50 percent chance of responding correctly to the point where the student would have a 65 percent chance of responding correctly. This was achieved by adding the natural log of the odds of a 65 percent response rate to the original log odds, since the logit metric allows this addition to take place in a straightforward manner. Second, the new logit item difficulty was transformed into the international achievement scale. This was done five times, once with the mean and standard deviation of each plausible value (shown in Tables 8.1 through 8.12). The average of this transformation was taken as the transformed international item difficulty:

$$d'_i = \left(\frac{1}{5}\right) \times \sum_{j=1}^{5} \left(500 + 100 \times \left(\frac{d_i + \ln(0.65/0.35) - \bar{\theta}_j}{SD_{\theta_j}}\right)\right)$$

where d'_i is the item difficulty for item i transformed onto the international standardized scale metric, d'_i is the item difficulty in the original logit metric, $\bar{\theta}_j$ is the mean logit score on each plausible value for the scale to which the item is assigned, and SD_{θ_j} is the standard deviation of the plausible values. For the purpose of transforming the item difficulties, only the difficulty of the items on the overall scale was used. That is, the difficulty for an item is presented as part of one of the four overall scales reported: mathematics literacy, science literacy, advanced mathematics, or physics.

8.3 MULTIPLE COMPARISONS OF ACHIEVEMENT

An essential purpose of the TIMSS international reports is to provide fair and accurate comparisons of student achievement across the participating countries. Most of the tables in the reports summarize student achievement by means of a statistic such as a mean or percentage, and each summary statistic is accompanied by its standard error, which is a measure of the variability in the statistic resulting from the sampling process. When comparing the performance of students from two countries, standard errors can be used to assess the statistical significance of the difference between the summary statistics.

The multiple comparison charts presented in the TIMSS international report for Population 3 are designed to help the reader compare the average performance of a country with that of other participating countries of interest. The significance tests reported in these charts are based on a Bonferroni procedure for multiple comparisons that holds to 5 percent the probability of erroneously declaring the mean of one country to be different from that of another country.

If we were to take repeated samples from two populations with the same mean and test the hypothesis that the means from these two samples are significantly different at the $\alpha = .05$ level, i.e. with 95 percent confidence, then in about 5 percent of the comparisons we would expect to find significant differences between the sample means even though we know that there is no difference between the population means. In this example with one test of the difference between two means, the probability of finding significant differences in the samples when none exist in the populations (the so-called type I error) is given by $\alpha = .05$. Conversely, the probability of not making a type I error is $1 - \alpha$, which in the case of a single test is .95. However, if we wish to compare the means of three countries, this involves three tests (country A versus country B, country B versus country C, and country A versus country C). Since these are independent tests, the probability of **not** making a type I error in any of these tests is the product of the individual probabilities, which is $(1 - \alpha)(1 - \alpha)(1 - \alpha)$. With $\alpha = .05$, the overall probability of not making a type I error is only .873, which is considerably less than the probability for a single test. As the number of tests increases, the probability of not making a type I error decreases, and conversely, the probability of making a type I error increases.

Several methods can be used to correct for the increased probability of a type I error while making many simultaneous comparisons. Dunn (1961) developed a procedure that is appropriate for testing a set of a priori hypotheses while controlling the probability that the type I error will occur. When using this procedure, the researcher adjusts the value α when making multiple simultaneous comparisons to compensate for the increase in the probability of making a type I error. This is known as the Dunn-Bonferroni procedure for multiple a priori comparisons (Winer, Brown, and Michels, 1991).

In this procedure the significance level of the test of the difference between means is adjusted by dividing the significance level (α) by the number of comparisons that are planned and then looking up the appropriate quantile from the normal distribution. In deciding the number of comparisons, and hence the appropriate adjustment to the significance level for TIMSS, it was necessary to decide how the multiple comparison tables would most likely be used. One approach would have been to adjust the significance level to compensate for all possible comparisons between the countries presented in the table. This would have meant adjusting the significance level for 420 comparisons for mathematics and science literacy. In decision-making terms this would have been a very conservative procedure, however, and would have run the risk of making an error of a different kind, that of concluding that a difference between sample means is not significant when in fact there is a difference between the population means.

Since most users are likely to be interested in comparing a single country with all other countries, rather than in making all possible between-country comparisons at once, a more realistic approach would be to adjust the significance level for a number of comparisons equal to the number of countries (minus one). This was the approach adopted in TIMSS. From this perspective, for mathematics and science literacy, the number of simultaneous comparisons to be adjusted for is 20 instead 420. The number of comparisons is 15 for mathematics and also 15 for physics. As a consequence, we used the critical values shown in Table 8.13, given by the appropriate quantiles from the normal (Gaussian) distribution.

Table 8.13 Critical Values Used for the Multiple Comparison Figures in TIMSS International Report

	Alpha Level	Number of Comparisons	Critical Value
Mathematics and Science Literacy	0.05	20	3.0233
Avanced Mathematics	0.05	15	2.9353
Physics	0.05	15	2.9353

Two means were considered significantly different from each other if the absolute differences between them was greater than the critical value multiplied by the standard error of the difference. The standard error of the difference between the two means was computed as the square root of the sum of the squared standard errors of the mean:

$$se_{diff} = \sqrt{se_1^2 + se_2^2}$$

where se_1 and se_2 are the standard errors for each of the means being compared, respectively, computed using the jackknife method of variance estimation. Table 8.14 shows the means and standard errors used in the calculation of statistical significance between means for mathematics and science literacy, mathematics literacy, science literacy, advanced mathematics, and physics. By applying the Bonferroni correction, we were able to state that, for any given row or column of the multiple comparison chart, the differences between countries shown in the chart are statistically significant at the 95 percent level of confidence.

Table 8.14 Means and Standard Errors for Multiple Comparisons Figures

Country	Mathematics and Science Literacy		Mathematics Literacy		Science Literacy		Advanced Mathematics		Physics	
	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
Australia	525	9.5	522	9.3	527	9.8	525	11.6	518	6.2
Austria	519	5.4	518	5.3	520	5.6	436	7.2	435	6.4
Canada	526	2.6	519	2.8	532	2.6	509	4.3	485	3.3
Cyprus	447	2.5	446	2.5	448	3.0	518	4.3	494	5.8
Czech Republic	476	10.5	466	12.3	487	8.8	469	11.2	451	6.2
Denmark	528	3.2	547	3.3	509	3.6	522	3.4	534	4.2
France	505	4.9	523	5.1	487	5.1	557	3.9	466	3.8
Germany	496	5.4	495	5.9	497	5.1	465	5.6	522	11.9
Greece	-	-	-	-	-	-	513	6.0	486	5.6
Hungary	477	3.0	483	3.2	471	3.0	-	-	-	-
Iceland	541	1.6	534	2.0	549	1.5	-	-	-	-
Italy	475	5.3	476	5.5	475	5.3	474	9.6	.	.
Latvia (LSS)	-	-	-	-	-	-	-	-	488	21.5
Lithuania	465	5.8	469	6.1	461	5.7	516	2.6	-	-
Netherlands	559	4.9	560	4.7	558	5.3	-	-	-	-
New Zealand	525	4.7	522	4.5	529	5.2	-	-	-	-
Norway	536	4.0	528	4.1	544	4.1	-	-	581	6.5
Russian Federation	476	5.8	471	6.2	481	5.7	542	9.2	545	11.6
Slovenia	514	8.2	512	8.3	517	8.2	475	9.2	523	15.5
South Africa	352	9.3	356	8.3	349	10.5	-	-	-	-
Sweden	555	4.3	552	4.3	559	4.4	512	4.4	573	3.9
Switzerland	531	5.4	540	5.8	523	5.3	533	5.0	488	3.5
United States	471	3.1	461	3.2	480	3.3	442	5.9	423	3.3

A dash (–) indicates country did not participate in assessment.
S.E. = standard error.

8.4 ESTIMATING THE ACHIEVEMENT OF THE TOP 5 PERCENT, 10 PERCENT, AND 25 PERCENT OF STUDENTS IN THE SCHOOL-LEAVING AGE COHORT

As indicated by the test coverage indices, the samples of all final-year students in some countries represented nearly all of the students in the school-leaving age cohort, while in others it represented fewer and as low as only half of these students. For these latter countries, because of their target population, the physics and advanced mathematics samples represented a smaller fraction of the students in the school-leaving age cohort.

As described in Chapter 2, TIMSS computed an index quantifying the percentage of students in the school-leaving age cohort covered by the TIMSS samples. This index is called the TIMSS Coverage Index (TCI). Building on this index, the Mathematics TIMSS Coverage Index (MTCI) quantifies the percentage of students in the school-leaving age cohort covered by the advanced mathematics sample and the physics TIMSS Coverage Index (PTCI) quantifies the percentage of the school-leaving age cohort covered by the physics sample.

To take into account the different proportions of students in the school-leaving age cohort represented in the samples, TIMSS computed the performance in mathematics and science literacy for the top 25 percent of the students in the school-leaving age cohort, and the average performance in advanced mathematics and physics of the top 5 percent and top 10 percent of the students in the school-leaving age cohort. When computing each of these percentiles we assumed that students not tested in the subject area would have scored below the percentile in question, primarily because they were not in school, in the case of the samples of all final-year students, or because they had not taken courses in advanced mathematics or physics, in the case of the advanced mathematics and physics samples.

When computing the average performance of the students above a certain percentile, the population of students covered by the TIMSS tests had to be adjusted as follows. We assumed that students not tested would score below the percentile. For example, in the United States the TCI was 63.1 percent. This means that the US school-leaving age cohort is approximately the population covered by TIMSS (2278258.19) plus the 36.9 percent that was not covered (2278258.19 * 100 / 63.1) or approximately 3,610,552 students. Now, if we had tested all students in the school-leaving age cohort (3.6 million), then the 75th percentile of those people would have been found easily. However, we did not test 1.3 million of these students, and we assume they would have performed below the 75th percentile of all the students. Then, to find the 75th percentile all we need to do is take away the top 25 percent of the 63.1 percent which corresponds to the 60.4th percentile of the tested sample, computed as $\left(1 - \frac{25}{63.1}\right) * 100$.

Table 8.15 shows, for each assessment, the percentile that was used to select the students in the sample above the percentile points.

Table 8.15 Percentiles of Performance

Country	TCI	MTCI	PTCI	Mathematics and Science Literacy	Advanced Mathematics		Physics	
				Percentile for Top 25%	Percentile for Top 10%	Percentile for Top 5%	Percentile for Top 10%	Percentile for Top 5%
Australia	68.1%	15.7%	12.6%	63.3	36.5	68.2	20.7	60.3
Austria	75.9%	33.3%	33.1%	67.1	70.0	85.0	69.7	84.9
Canada	70.3%	15.6%	13.7%	64.4	36.1	68.0	26.8	63.4
Cyprus	47.9%	8.8%	8.8%	47.8	–	43.5	–	43.5
Czech Republic	77.6%	11.0%	11.0%	67.8	8.9	54.4	8.8	54.4
Denmark	57.7%	20.6%	3.2%	56.6	51.4	75.7	–	–
France	83.9%	19.9%	19.9%	70.2	49.8	74.9	49.8	74.9
Germany	75.3%	26.3%	8.4%	66.8	62.0	81.0	–	40.5
Greece	–	10.0%	10.0%	–	0.0	50.0	0.0	50.0
Hungary	65.3%	–	–	61.7	–	–	–	–
Iceland	54.6%	–	–	54.2	–	–	–	–
Italy	51.5%	14.1%	8.6%	51.5	29.2	64.6	–	42.0
Lithuania	42.5%	2.6%	–	41.2	–	–	–	–
Netherlands	78.0%	–	–	67.9	–	–	–	–
New Zealand	70.5%	–	–	64.5	–	–	–	–
Norway	84.0%	–	8.4%	70.2	–	–	–	40.3
Russian Federation	48.1%	2.1%	1.8%	48.0	–	–	–	–
Slovenia	87.8%	75.4%	38.6%	71.5	86.7	93.4	74.1	87.0
South Africa	48.9%	–	–	48.9	–	–	–	–
Sweden	70.6%	16.2%	16.3%	64.6	38.4	69.2	38.6	69.3
Switzerland	81.9%	14.3%	14.2%	69.5	29.9	64.9	29.4	64.7
United States	63.1%	13.7%	14.5%	60.4	27.2	63.6	30.9	65.4

A dash (–) indicates country did not participate in assessment.

8.5 REPORTING GENDER DIFFERENCES WITHIN COUNTRIES

Gender differences were reported in overall student achievement in mathematics and science literacy, mathematics literacy, science literacy, advanced mathematics, and physics, as well as in the various subject matter content areas.

The analysis of overall gender differences focused on significant differences in achievement within each country in terms of the international scale scores. These results are presented in a table with an accompanying graph indicating whether the difference between male and female achievement was statistically significant. The significance of the difference was determined by comparing the absolute value of the standardized difference between the two means with a critical value of 1.96, corresponding to a 95 percent confidence level (two-tailed test; $\alpha = 0.05$, with infinite degrees of freedom). The standardized difference between the mean for males and females (t) was computed as

$$t_k = \frac{\bar{x}_{kb} - \bar{x}_{kg}}{\sqrt{se_{kb}^2 + se_{kg}^2}}$$

where t_k is the standardized difference between two means for country k, $\bar{x}_{kb}$ and $\bar{x}_{kg}$ are the means for males and females within country k, and se_{kb} and se_{kg} are the standard errors for the males' and females' means in country k computed using the jackknife error estimation method described earlier. The above formula assumes independent samples of males and females, and was used in TIMSS due to time constraints. However, since in most countries males and females attended the same schools, the samples of males and females are not completely independent. It would have been more correct to jackknife the difference between males and females. The appropriate test is then the difference between the mean for males and the mean for females divided by the jackknife standard error of the difference. Tables 8.16 through 8.20 show, for mathematics and science literacy, advanced mathematics, and physics, the standard errors of the differences computed under the assumption of independent sampling for males and females and computed using the jackknife technique for correlated samples. No corrections for multiple comparisons were made when comparing the achievement for males and females.

Table 8.16 Standard Error of the Gender Difference Mathematics and Science Literacy

Country	Males' Mean and (S.E.)	Females' Mean and (S.E.)	Males' and Females' Difference	JRR S.E. of Difference - Correlated Samples	JRR S.E. of Difference - Independent Samples
Australia	543 (10.7)	511 (9.3)	32.0	6.8	14.2
Austria	549 (7.8)	502 (5.5)	47.0	9.4	9.6
Canada	544 (3.4)	511 (3.4)	33.0	4.6	4.8
Cyprus	456 (4.9)	439 (3.0)	18.0	6.4	5.8
Czech Republic	500 (9.9)	452 (13.8)	48.0	14.7	17.0
Denmark	554 (4.5)	507 (3.7)	47.0	5.8	5.8
France	526 (5.9)	487 (4.8)	38.0	5.2	7.6
Germany	512 (8.2)	479 (8.5)	32.0	12.3	11.8
Hungary	485 (4.5)	468 (4.5)	17.0	6.9	6.3
Iceland	565 (2.9)	522 (1.9)	43.0	3.6	3.5
Italy	492 (6.9)	461 (5.7)	31.0	7.9	8.9
Lithuania	483 (6.7)	456 (7.4)	27.0	8.7	10.0
Netherlands	584 (5.5)	533 (5.9)	51.0	7.1	8.0
New Zealand	540 (5.7)	511 (5.5)	28.0	6.0	7.9
Norway	564 (5.0)	507 (4.5)	57.0	5.8	6.8
Russian Federation	499 (5.9)	462 (6.5)	37.0	5.0	8.8
Slovenia	538 (12.6)	492 (7.1)	46.0	12.2	14.4
South Africa	366 (10.3)	341 (11.8)	25.0	11.6	15.7
Sweden	579 (5.9)	533 (3.6)	46.0	6.1	6.9
Switzerland	547 (6.0)	511 (7.5)	37.0	8.7	9.6
United States	479 (4.2)	462 (3.5)	17.0	4.7	5.5

JRR = jacknife repeated replicate method

S.E. = standard error

Table 8.17 Standard Error of the Gender Difference Mathematics Literacy

Country	Males' Mean and (S.E.)	Females' Mean and (S.E.)	Males' and Females' Difference	JRR S.E. of Difference - Correlated Samples	JRR S.E. of Difference - Independent Samples
Australia	540 (10.3)	510 (9.3)	30.0	6.7	13.9
Austria	545 (7.2)	503 (5.5)	41.0	8.5	9.0
Canada	537 (3.8)	504 (3.5)	34.0	4.9	5.2
Cyprus	454 (4.9)	439 (3.7)	15.0	7.0	6.1
Czech Republic	488 (11.3)	443 (16.8)	45.0	17.1	20.2
Denmark	575 (4.0)	523 (4.0)	52.0	5.7	5.7
France	544 (5.6)	506 (5.3)	38.0	5.1	7.7
Germany	509 (8.7)	480 (8.8)	29.0	12.3	12.4
Hungary	485 (4.9)	481 (4.8)	5.0	7.4	6.9
Iceland	558 (3.4)	514 (2.2)	44.0	3.9	4.1
Italy	490 (7.4)	464 (6.0)	26.0	8.5	9.5
Lithuania	485 (7.3)	461 (7.7)	23.0	9.3	10.6
Netherlands	585 (5.6)	533 (5.9)	53.0	7.6	8.2
New Zealand	536 (4.9)	507 (6.2)	29.0	6.4	7.9
Norway	555 (5.3)	501 (4.8)	54.0	6.2	7.1
Russian Federation	488 (6.5)	460 (6.6)	27.0	4.7	9.2
Slovenia	535 (12.7)	490 (8.0)	46.0	12.8	15.0
South Africa	365 (9.3)	348 (10.8)	17.0	11.0	14.3
Sweden	573 (5.9)	531 (3.9)	42.0	6.3	7.1
Switzerland	555 (6.4)	522 (7.4)	33.0	8.3	9.8
United States	466 (4.1)	456 (3.6)	11.0	4.4	5.5

JRR = jacknife repeated replicate method

S.E. = standard error

Table 8.18 Standard Error of the Gender Difference Science Literacy

Country	Males' Mean and (S.E.)	Females' Mean and (S.E.)	Males' and Females' Difference	JRR S.E. of Difference - Correlated Samples	JRR S.E. of Difference - Independent Samples
Australia	547 (11.5)	513 (9.4)	34.0	7.4	14.8
Austria	554 (8.7)	501 (5.8)	53.0	10.7	10.4
Canada	550 (3.6)	518 (3.8)	32.0	5.4	5.2
Cyprus	459 (5.8)	439 (3.0)	20.0	6.8	6.5
Czech Republic	512 (8.8)	460 (11.0)	51.0	12.6	14.0
Denmark	532 (5.4)	490 (4.1)	41.0	6.3	6.8
France	508 (6.7)	468 (4.8)	39.0	5.9	8.3
Germany	514 (7.9)	478 (8.5)	35.0	12.6	11.6
Hungary	484 (4.2)	455 (4.3)	29.0	6.6	6.0
Iceland	572 (2.7)	530 (2.1)	41.0	3.6	3.4
Italy	495 (6.7)	458 (5.6)	37.0	7.8	8.8
Lithuania	481 (6.4)	450 (7.3)	31.0	8.3	9.7
Netherlands	582 (5.7)	532 (6.2)	49.0	7.1	8.4
New Zealand	543 (7.1)	515 (5.2)	28.0	6.5	8.8
Norway	574 (5.1)	513 (4.5)	61.0	5.8	6.8
Russian Federation	510 (5.7)	463 (6.7)	47.0	5.8	8.8
South Africa	367 (11.5)	333 (13.0)	34.0	12.5	17.4
Sweden	585 (6.0)	534 (3.5)	51.0	6.1	6.9
Switzerland	540 (6.1)	500 (7.8)	40.0	9.4	9.9
United States	492 (4.5)	469 (3.9)	23.0	5.5	5.9
Slovenia	541 (12.7)	494 (6.4)	47.0	12.0	14.3

JRR = jacknife repeated replicate method

S.E. = standard error

Table 8.19 Standard Error of the Gender Difference Advanced Mathematics

Country	Males' Mean and (S.E.)	Females' Mean and (S.E.)	Males' and Females' Difference	JRR S.E. of Difference - Correlated Samples	JRR S.E. of Difference - Independent Samples
Australia	531 (11.4)	517 (15.1)	14.0	12.5	18.9
Austria	486 (7.3)	406 (8.6)	80.0	11.5	11.2
Canada	528 (6.4)	489 (4.4)	39.0	7.5	7.7
Cyprus	524 (4.4)	509 (6.4)	15.0	6.0	7.8
Czech Republic	524 (13.0)	432 (8.9)	92.0	10.0	15.7
Denmark	529 (4.4)	510 (4.6)	19.0	5.8	6.3
France	567 (5.1)	543 (5.1)	23.0	7.0	7.2
Germany	484 (6.5)	452 (6.6)	32.0	7.2	9.2
Greece	516 (6.6)	505 (10.2)	11.0	11.3	12.1
Italy	484 (10.6)	460 (14.1)	24.0	15.1	17.7
Lithuania	542 (3.7)	490 (5.6)	51.0	8.1	6.7
Russian Federation	568 (9.7)	515 (10.2)	53.0	10.5	14.1
Slovenia	484 (11.5)	464 (11.0)	20.0	13.5	15.9
Sweden	519 (5.9)	496 (5.2)	23.0	8.2	7.9
Switzerland	559 (5.6)	503 (5.7)	56.0	6.0	8.0
United States	457 (7.8)	426 (7.1)	31.0	8.7	10.5

JRR = jacknife repeated replicate method

S.E. = standard error

Table 8.20 Standard Error of the Gender Difference Physics

Country	Males' Mean and (S.E.)	Females' Mean and (S.E.)	Males' and Females' Difference	JRR S.E. of Difference - Correlated Samples	JRR S.E. of Difference - Independent Samples
Australia	532 (6.7)	490 (8.4)	42.0	8.2	10.8
Austria	479 (8.1)	408 (7.4)	71.0	10.4	11.0
Canada	506 (6.0)	459 (6.3)	47.0	10.5	8.7
Cyprus	509 (8.9)	470 (7.1)	40.0	12.6	11.4
Czech Republic	503 (8.8)	419 (3.9)	83.0	8.2	9.7
Denmark	542 (5.2)	500 (8.1)	42.0	10.1	9.6
France	478 (4.2)	450 (5.6)	28.0	5.8	7.0
Germany	542 (14.3)	479 (9.1)	64.0	13.5	17.0
Greece	495 (6.1)	468 (8.1)	28.0	8.2	10.1
Italy	. .	. .	.	.	.
Latvia (LSS)	509 (19.0)	467 (22.6)	42.0	7.6	29.5
Norway	594 (6.3)	544 (9.3)	51.0	8.0	11.2
Russian Federation	575 (9.9)	509 (15.3)	66.0	10.7	18.2
Slovenia	546 (16.3)	455 (18.7)	91.0	20.1	24.8
Sweden	589 (5.1)	540 (5.3)	49.0	7.3	7.4
Switzerland	529 (5.2)	446 (3.6)	83.0	5.7	6.3
United States	439 (4.3)	405 (3.1)	33.0	4.9	5.3

JRR = jacknife repeated replicate method

S.E. = standard error

8.6 PERCENT CORRECT FOR INDIVIDUAL ITEMS

To portray student achievement as fully as possible, the TIMSS international report presents many examples of the items used in the TIMSS tests, together with the percentage of students in each country responding correctly to the item. For multiple-choice items this was the weighted percentage of students that answered the item correctly. This percentage was based on the total number of students that were administered the items. Omitted and not-reached items were treated as incorrect. The percent correct for free-response items with more than one score level was computed as the weighted percentage of students that achieved the highest score possible on the item.

When the percent correct for example items was computed, student responses were classified in the following way. For multiple-choice items, the responses to item j were classified as correct (C_j) when the correct option for an item was selected, incorrect (W_j) when the incorrect option was selected, invalid (I_j) when two or more choices were made on the same question, not reached (R_j) when it was determined that the student stopped working on the test before reaching the question, and not administered (A_j) when the question was not included in the student's booklet or had been mistranslated or misprinted. For free-response items, student responses to item j were classified as correct (C_j) when the maximum number of points was obtained on the question, incorrect (W_j) when the wrong answer or an answer not worth all the points in the question was given, invalid (N_j) when, the students' response was not legible or interpretable, not reached (R_j) when it was determined that the student stopped working on the test before reaching the question, and not administered (A_j) when the question was not included in the student's booklet or had been mistranslated or misprinted. The percent correct for an item (P_j) was computed as

$$P_j = \frac{c_j}{c_j + w_j + i_j + r_j + n_j}$$

where c_j, w_j, i_j, r_j and n_j are the weighted counts of the correct, wrong, invalid, not reached, and not interpretable responses to item j, respectively.

Note that although the not-reached responses were treated as missing for the purpose of estimating the item parameters in the international IRT scaling, they were considered to be wrong answers for a student when percents correct for an item were computed.

8.7 THE TEST-CURRICULUM MATCHING ANALYSIS

TIMSS developed international tests of advanced mathematics and physics that reflect, as far as possible, the various curricula of the participating countries. The subject matter coverage of these tests was reviewed by the TIMSS Subject Matter Advisory Committee, which consists of mathematics and physics educators and practitioners from around the world, and the tests were approved for use by the National Research Coordinators (NRCs) of the participating countries. Although every effort was made in TIMSS to ensure the widest possible subject matter coverage, no test can measure all that is taught or learned in every participating country. Given that no test can cover the curriculum in every country completely, the question arises as to how well the items on the tests match the curricula of each of the participating countries. To address this issue, TIMSS asked each country to indicate which items on the tests, if any, were inappropriate to its curriculum. For each country, in turn, TIMSS took the list of remaining items, and computed the average percentage correct on these items for that country and all other countries. This allowed each country to select only those items on the tests that they would like included, and to compare the performance of their students on those items with the performance of the students in each of the other participating countries on that set of items. However, in addition to comparing the performance of all countries on the set of items chosen by each country, the Test-Curriculum Matching Analysis (TCMA) also shows each country's performance on the items chosen by each of the other countries. In these analyses, each country was able not only to see the performance of all countries on the items appropriate for its curriculum, but also to see the performance of its students on items judged appropriate for the curriculum in other countries.

Each NRC was given a questionnaire with all the items included in the TIMSS advanced mathematics and physics tests and was asked to indicate, for each item, whether it was considered an appropriate item for their curriculum. The results from these questionnaires were then used to assess the curricular coverage of the items in the tests, and what effect omitting items identified by each NRC had on the test results of all countries. It must be stressed that this analysis was not intended to replace the carefully constructed and agreed-upon tests that TIMSS used for its international comparisons and research analyses. The IRT scaling and research analyses used all items that were included in the tests and that met psychometric standards. In the TCMA analysis, items identified by NRCs were omitted from test results only in the analyses designed to illuminate and explain the international comparisons based on the entire test.

8.7.1 The Analytical Method of the TCMA[1]

The TCMA makes use of the average proportion-correct technology. The basic item-level data for a participating country were represented by the matrix D_{ikj}. This matrix contains elements d_{ikj}, which represent the scored response of student i in country k to item j. The possible values for item j are 0 or 1 for multiple-choice items, and between 0 and 3 for multiple-score items. Most of the elements of D are missing since each student took only one of four possible booklets administered in each subject. Depending on the booklet, each student took between one-seventh and three-sevenths of the total item pool (Adams and Gonzalez, 1996).

The information provided by the NRC as to whether or not an item should be omitted from these analyses was summarized in a matrix T_{kj}, where the elements t_{kj} represent the information that the NRC in country k submitted about item j (for a particular grade). The actual responses of the NRCs for an item were 0 (meaning omit this item for my country) or 1 (meaning include it). Given that multiple-score items were included

[1] The analytic method of the TCMA was developed by Albert E. Beaton, TIMSS International Study Director.

in the TIMSS tests, both matrices D_{ikj} and T_{kj} were then converted to $D_{ikj'}$ and $T_{kj'}$ matrices as described in the previous chapter. In that conversion, the score points on each item in the matrix $D_{ikj'}$ were transformed into their binary representation, and the item selection by the NRC, contained in the matrix $T_{kj'}$, was transformed into a matrix that matched the $D_{ikj'}$.

Although the procedure described here will work generally for any item selection proportion from 0 to 1, the TCMA analysis in TIMSS was limited to a binary choice of either including or excluding the item at the specific grade level. The computational procedure used for the TCMA analysis was as follows. First form the $P'_{kj'}$ matrix. The elements in matrix $P'_{kj'}$ are computed from the D_{ikj} matrix after the transformations and estimation outlined in Chapter 9 in the *TIMSS Technical Report, Volume II* (Martin and Kelly, 1997) are applied to the data. The elements of $P'_{kj'}$ are the weighted averages of the student responses in country k to item j', that is, the average of the student responses $d_{ikj'}$, estimated for some elements. Under the TIMSS design, students not administered particular items may be considered missing at random and treated as not having taken the item. Item responses coded as not reached or omitted are treated as incorrect responses.

The next step is to compute an index of text coverage. A reasonable index is the percentage of the total possible test points that were deemed appropriate by each country. This index should not be confused with the TIMSS Coverage Index (TCI) discussed in Chapter 2 and earlier in this chapter. The total possible test points in a TIMSS test are equal to C_t, and the total possible score on the items deemed appropriate in country k is computed as

$$C_k = \sum_{j'} t_{kj'}$$

The index can then be computed as the ratio of the total possible score on the items deemed appropriate in country k to the total possible test points in the TIMSS test:

$$\frac{C_k}{C_t}$$

This index indicates the proportion of score points of the test that was considered appropriate to the curriculum in the country. The index for each country is presented in Table 8.21.

Table 8.21 Index of Test Coverage Advanced Mathematics and Physics

Country	Advanced Mathematics	Physics
Australia	0.87	0.96
Austria	1.00	1.00
Canada	0.85	0.73
Cyprus	0.93	0.96
Czech Republic	0.98	0.95
Denmark	0.79	0.90
France	0.98	0.74
Germany	0.79	0.96
Russian Federation	0.82	0.47
Slovenia	0.99	0.96
Sweden	0.76	-
Switzerland	0.88	0.53
United States	1.00	1.00

After computing the index of test coverage, the next step was to compute the normalized weight matrix. To facilitate cross-national comparisons, it is useful to anchor the various national proficiency estimates in a common manner. The national proficiency estimates described in the next section have the property that, if the students in a country correctly answer all of the items deemed appropriate for that country, then the country will receive a value of 100; if the students answer all of those items incorrectly, then the country will receive a value of 0. Items not deemed appropriate to the curriculum of a country are not used in computing these values. In situations where the information in T is either 1 (include) or 0 (omit), the country values may be considered percentages of possible points attained on included items. If T contains proportions other than 0 and 1, then the country values may be greater than 100, in which case the students answered more items correctly than was expected from the values in T.

To compute such country estimates, it is necessary to construct the matrix $W_{kj'}$, with the elements $w_{kj'}$, where the matrix elements are computed as follows:

$$w_{kj} = \frac{t_{kj'}}{\sum_{j'} t_{kj'}^2}$$

where the denominator of this equation is the sum of the squares of the NRCs' judgments of the items.

The Country Comparison Matrix can be computed from $P_{kj'}$ and $W_{kj'}$ by the matrix multiplication

$$C_{kk'} = 100 * (W_{kj'} * P'_{kj'})$$

where the elements of $C_{kk'}$ indicate how the students in country k' scored on the items deemed appropriate in country k.

Another way to estimate the $C_{kk'}$ matrix directly without going through the intermediate step of computing the w_{kj} matrix is as follows:

$$C_{kk'} = \frac{\sum_{j'} t_{kj'} * p_{kj'}}{\sum_{j'} t_{kj'}^2} * 100 .$$

The estimates in the resulting Country Comparison Matrix are unbiased estimators of average student performance based on the items selected by each country for inclusion in the TCMA. The precision of estimates varies as a result of the test booklet rotation as well as the different school and student sampling plans.

8.7.2 Computing Standard Errors

The computation of standard errors for the TCMA is a continuation of the procedure for computing the standard error for the average percent correct as described in Chapter 9 of the *TIMSS Technical Report, Volume II* (Martin and Kelly, 1997). Once the $P_{kj}'^{h'}$ matrices are obtained, we then continue to compute each of the $C_{kk'}^{h'}$ matrices, which can be computed with each of the different $P_{kj}'^{h'}$ replicate matrices. This is accomplished in a straightforward manner by use of the following multiplication:

$$C_{kk'}^{h'} = \frac{\sum_{j'} t_{kj'} * p_{kj'}^{h'}}{\sum_{j'} t_{kj'}^2} * 100 .$$

The jackknifed standard errors for each of the elements in the $C_{kk'}$ matrix are then computed by applying the following formula:

$$jse_{C_{kk'}} = \sqrt{\sum_{h'} (c_{kk'} - c'^{h'}_{kk'})^2} .$$

REFERENCES

Adams, R. J. and Gonzalez, E. J. (1996). The TIMSS test design. In Martin, M.O. and Kelly, D.L. (Eds.), *TIMSS technical report, volume I: Design and development*. Chestnut Hill, MA: Boston College.

Dunn, O.J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, *56*, 52-64.

Martin, M.O., and Kelly, D.L., Eds. (1997). *TIMSS technical report, volume II: Implementation and analysis- primary and middle school years*. Chestnut Hill, MA: Boston College.

Winer, B. J., Brown, D. R., and Michels, K. M. (1991). *Statistical principles in experimental design.* New York: McGraw Hill.

Appendix A: Table of Contents for Volume I of the Technical Report

Appendix B: Characteristics of the National Samples
Population 3 (Final Year of Secondary School)

B

AUSTRALIA

Structure of Upper Secondary System

Secondary education is provided for either five or six years depending on the length of primary education in the state. Australia's secondary schools provide a comprehensive education, although students can focus on academic/pre-university studies, including humanities and art, mathematics and science, commerce, and other disciplines, or they can focus on vocationally oriented studies.

Students Tested in Mathematics and Science Literacy

Australia tested students in the final year of secondary school, Grade 12, in government, Catholic, and independent schools.

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: students in the final year of secondary school, Grade 12, enrolled in mathematics courses (varies across states) preparing them for post-secondary study, and students in Grade 12 who took such mathematics courses during Grade 11.

Physics: students in the final year of secondary school, Grade 12, enrolled in Year 12 physics.

Coverage and Exclusions

Very small schools, schools for adult education, and schools in geographically remote locations were excluded.

Sample Design

- Definitions of subpopulations:

  MO Grade 12 students having taken the following advanced mathematics courses, but not year 12 physics:

State	Course Name
New South Wales	3-unit or 4-unit math
Victoria	Change and Approximations
Queensland	Math I + Math II, or Math B+ Math C
Western Australia	Calculus
South Australia	Math 1 + Math 2
Tasmania	Math Stage 3 + special units
Northern Territory	Math 1 + Math 2
Australian Capital Terr.	Double major or major-minor in math

OP Grade 12 students having taken Year 12 Physics but not advanced mathematics

MP Grade 12 students having taken both Year 12 Physics and advanced mathematics course (defined above)

OO Grade 12 students having taken neither Year 12 Physics nor advanced mathematics course

- For planning purposes, final-year population was estimated to be 65% OO, 5% OP, 18% MO, and 12% MP.
- Sample of 132 schools selected from state sampling frames with probability proportional to school size.
- Within selected schools, lists of pupils belonging to each subpopulation were compiled and simple random samples of students were drawn from each sub-population list.

AUSTRIA

Structure of Upper Secondary System

Academic and vocational schools form the upper secondary schooling in Austria. Academic secondary school (AHS) is a four-year cycle of pre-academic general education. Students may specialize in certain areas, but generally study a whole range of subjects. At the end of the cycle, students take a matriculation examination (*Matura*) which, upon passing, enables them to enter university.

There are three variations of vocational schools in Austria. Higher-technical and vocational (BHS) is a five-year cycle in which students study a similar academic curriculum to that in the AHS, but also study theoretical subjects relevant to future professions. Students train for careers in industry, trade, business, agriculture, or human service occupations. The final examination is similar to the AHS *Matura* and enables students to continue to university or obtain certain levels of vocational qualification. The final year of this cycle is Grade 13.

Intermediate-technical and vocational schools (BMS) are basically full-time schools equivalent to the dual system of school and apprenticeship (see below). These schools provide training in apprenticed trades and general education. The cycle is one to four years, but typically lasts three to four years. Successful completion results in vocational licenses which are sometimes more extensive than the ones given by the dual system. There are also higher teacher training colleges that represent an alternative route from the ninth year (grade) onwards.

In the system of dual vocational education – Apprenticeship/Berufsschulen (BS) – apprentices in business and industry receive practical vocational training at their place of work and also attend part-time vocational schools, *Berufsschulen*. Students typically attend the *Berufsschule* one day a week where some element of general education is included. The length of the course is from two to four years, but is three years for most students. The vocational qualification licenses the recipient to work in a legally defined trade.

Students Tested in Mathematics and Science Literacy

Austria tested students in their final year of academic schools (AHS), Grade 12, their final year of higher technical and vocational (BHS), Grade 13, and their final year of medium technical and vocational (BMS), Grades 10, 11, or 12, depending on the vocational program of the student, and students in their final year of the apprenticeship (BS).

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: students in their final year of the academic or higher technical track, taking courses in advanced mathematics.

Physics: students in their final year of the academic or higher technical track, taking courses in physics.

Coverage and Exclusions

Special schools for disabled students and colleges offering programs less than 3 years were excluded. Schools that participated in the TIMSS Population 2 (seventh and eighth grade) were assessment excluded.

Sample Design

- Definitions of subpopulations:

  MO Not applicable

  OP Not applicable

  MP Students in their final year, Grade 12, of AHS

  OO Students in their final year, Grade 12, of BS, BMS, and BHS

- For planning purposes, final-year population was estimated to have 62,000 OO and 18,000 MP students.

- Samples for each subpopulation were drawn from track-specific frames. Within selected units, classrooms were selected randomly: one classroom in AHS units; two classrooms in BMS units; one classroom in BHS units; and two classrooms in BS units.

CANADA

Structure of Upper Secondary System

Secondary education in Canada is comprehensive, although students can focus on academic/pre-university studies or vocationally oriented studies. The first years of secondary school are devoted to compulsory subjects, with some optional subjects included. In the latter years, the number of compulsory subjects is reduced, permitting students to spend more time on specialized programs that prepare them for the job market, or to take specific courses they need to meet the entrance requirements of the college or university of their choice. Senior high school ends in Grade 12 in all provinces except Quebec, where it ends in Grade 11. In Ontario, some students complete secondary schooling at the end of Grade 12, whereas others continue for an extra year to complete the Ontario Academic Credits (OAC) necessary for admission to university. Students in Quebec continue from Grade 11 to either a two- or three-year training program prior to entry into tertiary education or the workplace.

Students Tested in Mathematics and Science Literacy

Canada tested students in Grade 12 in all provinces except Quebec where students in Grades 13 and 14 (depending on program) were tested. In Ontario, students completing the OAC in Grade 13 also were tested.

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: students in their final year in mathematics courses preparing them for postsecondary study (varies by province), except in Quebec where students in the two-year science program were tested.

Physics: students in their final year in physics courses preparing them for postsecondary study (varies by province), except in Quebec where students in the two-year science program were tested.

Coverage and Exclusions

Very small schools and schools in Prince Edward Island were excluded. At the time of data collection, a number of final-year students in Ontario had already graduated in the prior semester and were excluded de facto.

Sample Design

- Definitions of subpopulations:

  All of Canada except Quebec:

  MO Grade 12 students taking advanced mathematics but not physics (as defined below)

  OP Grade 12 students taking physics but not advanced mathematics (as defined below)

  MP Grade 12 students taking both advanced mathematics and physics (as defined below)

OO Grade 12 students not taking either advanced mathematics nor physics (as defined below)

Province	Physics	Advanced Mathematics
Northwest, Yukon Terr.	Physics 12	Math 12
British Columbia	Physics 12	Math 12
Alberta	Physics 30	Math 31
Saskatchawan	Physics 30	Math 31
Manitoba	Physics 300	Math 200 or 305
Ontario	OAC Physics	OAC Math
New Brunswick	Grade 12 Physics	
Nova Scotia	Physics 441 or 541	Math 441 or 541
New Foundland	Physics 3204 or 4225	Math 3201, Calculus 3105 or 4225

Quebec:

MO Not applicable

OP Not applicable

MP Students enrolled in two-year science *cegep* program

OO All other *cegep* students

- For planning purposes, final-year population was estimated to be 60-80% OO and 15-20% OP or MO; percent MP unknown prior to testing.
- 389 schools were sampled from province sampling frames with probability proportional to school size.
- Within selected schools, lists of pupils belonging to each subpopulation were compiled and simple random samples of pupils were drawn from each subpopulation list.

CYPRUS

Structure of Upper Secondary System

Academic schools (lycea) and technical schools form the upper secondary schooling in Cyprus. At the lyceum, which comprises Grades 10, 11, and 12, students can choose one of five groups of subjects – classical (arts), mathematics and science, economics, commercial/secretarial, and foreign languages.

In technical schools, also three years in duration, students can take technical courses with particular emphasis on mathematics and science. Graduates of these programs typically follow further studies in colleges or universities. Technical schools also offer vocational programs in which students in the final year follow a training program in industry for two days a week and attend school for three days a week. In the vocational section, more emphasis is given to practical skills. The aim of public technical schools is to provide industry with technicians and craftsmen in various specializations such as mechanical and automobile engineering, computers, electronics, building, graphic arts, dressmaking, gold smithery, shoe manufacturing, and many others. Cyprus' private secondary schools are oriented towards commercial and vocational education and provide a six year education program.

Students Tested in Mathematics and Science Literacy

Cyprus tested students in Grade 12 of lycea and technical schools.

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: students in their final year in the mathematics/science program of study at the lyceum.

Physics: students in their final year in the mathematics/science program of study at the lyceum.

Coverage and Exclusions

Private schools and vocational schools/programs were excluded.

Sample Design

- Definitions of subpopulations:

  MO Not applicable

  OP Not applicable

  MP Grade 12 students in mathematics/science program of study at the lyceum

  OO All other Grade 12 students

- For planning purposes, final-year population was estimated to have 5,600 OO and 1,100 MP.

- All 29 in-scope schools were sampled. All MP students in these schools were tested. In other tracks, a random sample of 1 student in 10 was drawn.

CZECH REPUBLIC

Structure of Upper Secondary System

At the time of testing there were three types of secondary schools in the Czech Republic: gymnasium, technical, and vocational. The gymnasium is a four-, six-, or eight-year general secondary school providing demanding academic training for higher education. Students are in one of three streams in the gymnasium: humanities, science, or general education. Secondary technical schools, four or five years in duration, provide a broad general education as well as specialized study in a particular field (e.g., nursing, certain technical areas, tourism, library science, accounting, etc.). Students successfully completing the gymnasium or secondary technical school, and passing the final examination (maturita), are eligible to apply to institutions of higher education. Secondary vocational schools, with programs of two, three, four, or five years duration, provide practical vocational training as well as general education, with the aim to prepare students for occupations. These professional schools specialize mostly in engineering and technical areas.

Secondary schooling ends in different years depending on the type of school and the course of study within school. In almost all secondary technical school and gymnasia, students complete their education at the end of Grade 12, although a few complete their studies in Grade 13. In vocational schools, students may end in Grades 10, 11, 12, or 13, depending on their type of vocation.

Since the time of the TIMSS testing (1995), the Czech system has been modified to reflect an extension of basic school. Beginning in 1996, Grade 9 became compulsory (until this decision was made, Grade 9 was an optional grade, attended by 14% of the age cohort in 1993/94). This means that currently all secondary technical and gymnasium students complete their education in Grade 13 and most vocational students complete their studies in Grade 12.

Students Tested in Mathematics and Science Literacy

The Czech Republic tested students in their final year of each type of school. In technical schools and gymnasia, students in Grades 12 and 13 were tested. In vocational schools, students in Grades 10, 11, 12, and 13 were tested, depending on their vocation.

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: gymnasium students in their final year of study, Grade 12 or 13.

Physics: gymnasium students in their final year of study, Grade 12 or 13.

Coverage and Exclusions

Medical schools, schools for the disabled, and dance schools were excluded.

Sample Design

- Definitions of subpopulations:

MO	Not applicable
OP	Not applicable
MP	Students in their final year of gymnasium
OO	Students in their final year of vocational and technical schools

- For planning purposes, the final-year population was estimated to be 80% OO and 20% MP

- 150 schools were sampled from track-specific frames with probability proportional to school size.

- Within selected schools, lists of classes were established and one classroom from each school was selected at random.

DENMARK

Structure of Upper Secondary System

The general upper secondary programs are comprised of the general upper secondary certificates (*Studentereksamen*), the higher preparatory exam (HF) for mature students, the higher commercial exam (HHX), and the higher technical exam (HTX). The first two programs are taught at the Gymnasium and the last two at commercial and technical schools, respectively. All programs have a duration of three years except for the HF which is two years. The aim of the first two programs is primarily to prepare students for further studies at the tertiary level. The HHX and HTX prepare pupils for higher education but qualify also as final vocational education.

Vocational upper secondary programs encompass approximately 100 different specializations including vocational education and training, training for social affairs and health officers, agricultural education, and maritime education. Vocational training in Denmark is rooted in the apprenticeship tradition, but a wide-ranging modernization has been carried out over the past 30 years. This modernization has taken into account the lack of capacity among small and medium-sized enterprises to organize and carry out such training and reflects the need for a continuous updating of such programs.

Students Tested in Mathematics and Science Literacy

Denmark tested students in Grade 12 of the general secondary and vocational schools.

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: Grade 12 mathematics and physics students in the gymnasium and mathematics students in their final year, Grade 12, of the technical or higher preparation tracks.

Physics: Grade 12 mathematics and physics students in the gymnasium and physics students in their final year, Grade 12, of the technical track.

Coverage and Exclusions

Disabled and non-native language students were excluded.

Sample Design

- Definitions of subpopulations:

  MO Grade 12 mathematics students in the academic, technical, or higher preparation tracks taking advanced mathematics

  OP Not applicable

  MP Grade 12 students in the academic track of gymnasium taking advanced mathematics and advanced physics and students in the technical track taking advanced mathematics and advanced physics

  OO All other students

- For planning purposes, final-year population was estimated to be 22% MO; 4% MP; remaining OO.
- 130 schools (all schools) sampled.
- In each school, one classroom of language arts, one of mathematics, and one of physics were drawn. Some students were tested more than once for national assessment purposes and were later identified and removed from the TIMSS sample.
- Booklet rotation was not carried out according to TIMSS procedures.
- Classrooms were not sampled according to TIMSS procedures.

FRANCE

Structure of Upper Secondary System

There are two types of upper secondary schools in France: *lycées d'enseignement général et technologique*, or upper secondary school, for Grades 10 to 12, and *lycées professionnels*, or vocational upper secondary school, which may end at Grade 11 or Grade 13.

In the *lycée d'enseignement général et technologique*, students in Grades 10, 11, and 12 are in either the general track or the technological track. In Grade 10, there are both common areas of study and optional courses in the general and technological tracks. All students at this level take mathematics and science courses. In Grade 11, the different tracks are strongly differentiated, leading to corresponding types of *baccalauréats*. The *baccalauréat général* has three main tracks: scientific (S), literary (L), and economic and social (ES). The *baccalauréat technologique* has four major tracks within it: tertiary sciences and technologies (STT), industrial sciences and technologies (STI), medical-social sciences (SMS), and laboratory sciences and technologies (STL). The type and amount of mathematics and science taken by *lycée* students is different for each of the tracks within the general and technological tracks. The final year of the general and technological tracks is Grade 12.

Vocational Grade 10 is the first year of a program leading to the *Brevet d'études professionnelles* (BEP) or to the *Certificat d'aptitude professionnelle* (CAP). Most pupils achieve a *Brevet d'études professionnelles*, which is granted after Grade 11. About 50 percent of students achieving this diploma decide to continue their studies, either by joining the technological track through a *classe d'adaptation* or by continuing in vocational secondary for an additional two years to achieve the *baccalauréat professionnel*. Their choice depends mainly on their results, but also on the area of their studies and employment prospects with a *Brevet d'études professionnelles*. The *baccalauréat* leads directly to university studies. The final year for a student in the *lycée professionnel* is either Grade 11 or Grade 13, depending on whether or not they plan to continue their studies.

Students Tested in Mathematics and Science Literacy

France tested students in the final year of preparation for the *baccalauréat* (nonrepeaters of this final year). This included students in Grade 12 preparing for the *baccalauréat général ou technologique*, and in Grade 13 for the *baccalauréat professionnel* (vocational). Also tested were students in the final year (nonrepeaters of this year) of preparation for the *Brevet d'études professionnelles* (BEP) or the *Certificat d'aptitude professionnelle* (CAP) who will not continue towards a *baccalauréat*.

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: students in their final year of the scientific track, Grade 12, preparing for the *baccalauréat général.*

Physics: students in their final year of the scientific track, Grade 12, preparing for the *baccalauréat général.*

Coverage and Exclusions

Overseas territories were excluded.

Sample Design

- Definitions of subpopulations:

  MO Not applicable

  OP Not applicable

  MP Grade 12 students enrolled in the scientific track of lycées d'enseignement général et technologique

  OO All other students

- For planning purposes, final-year population was estimated to be 23% MP; remaining OO.

- Two independent samples were drawn. The first sample consisted of 71 schools from the lycées d'enseignement général et technologique and vocational schools (lycées professionnels); within these schools a sample of students was selected from the final year for the Mathematics and Science Literacy sample. The second sample consisted of 69 lycées d'enseignement général et technologique where the scientific track is offered; within these schools a sample of students in the scientific track was drawn for the Advanced Mathematics and Physics assessments.

GERMANY

Structure of Upper Secondary System

The upper secondary education system, Grades 11 to 13, in Germany is comprised of two types of schools – gymnasia or comprehensive schools and vocational schools. Education is compulsory up to age 18. In the upper grades of gymnasium, beginning in Grade 11, students can choose specializations within a rather complicated framework that allocates approximately one-third of instruction time to languages and arts, one-fourth to social studies (civic education, history, religion or philosophy), one-third to mathematics and science, and one-twelfth to sports. Upon successful completion of the final examination at the end of Grade 12 or 13 (final year depends on the Laender) a student may attend university.

Those students interested in vocational training have a variety of options. A dual system combines general education and theoretical instruction in the specific area of occupational training in part-time schools (*Berufsschule*), and practical training in one of over 500,000 authorized companies or businesses (*Betriebe*). Usually students in the dual system attend school two days a week and work the other three days at a company in a training program. At the company, students are supervised and taught by accredited trainers according to the training regulations in effect pertaining to the occupation. In larger companies, students often receive additional instruction in company schools. There is also a broad range of full-time vocational schools, such as *Fachgymnasien*, where students are instructed in economic and technical fields and admission requirements for university-level studies are fulfilled. Other types of schools are *Fachoberschulen* that certify for further specialized scientific training at institutions of higher education as well as *Berufsfachschulen* that provide occupational training for careers in social and health services and business.

Students Tested in Mathematics and Science Literacy

Germany tested students in their final year in the academic track of upper secondary education and the vocational education programs. This corresponded to Grade 13 in the Laender of the former West Germany and to Grade 12 in the Laender of the former East Germany.

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: students in their final year, Grade 12 or 13 depending on the Laender, in advanced mathematics courses (3 to 5 periods per week).

Physics: students in their final year, Grade 12 or 13, depending on the Laender, in physics courses (3 to 5 periods per week).

Coverage and Exclusions

Waldorf schools were excluded.

Sample Design

- Definitions of subpopulations:

  MO Grade 12 or 13 (depending on Laender) students in gymnasia and comprehensive schools taking advanced mathematics (3 to 5 periods a week) but not physics

  OP Not applicable

  MP Grade 12 or 13 (depending on Laender) students in gymnasia and comprehensive schools taking advanced mathematics and physics (3 to 5 periods a week)

  OO All other students

- For planning purposes, final-year population was estimated to be 31% MO; 14% MP; 55% OO

- Schools were sampled from track-specific school sampling frames with probability proportional to school size. All final-year students in gymnasia and comprehensive schools were in sample. Within sampled full-time vocational schools, one classroom was selected at random. In sampled part-time vocational schools, two sampling algorithms were used. In about half of these schools, one classroom was drawn from the set of classes in the final year. In the other half, each school was assigned at random to one of five trades. One class of that trade was then selected at random within the school.

- Data collection was conducted in 1995 in vocational schools and in 1996 in gymnasia.

GREECE

Structure of Upper Secondary System

The upper secondary system in Greece is a three-year program, Grades 10 to 12, taken in the general (academic) *Lyceum*, in the multibranch, semi-comprehensive *Lyceum* or in the technical-vocational *Lyceum*. Some students attend vocational and technical schools that provide two years of education, ending at Grade 11. In the general *Lyceum*, students in Grades 10 and 11 take the same courses. Students in the final grade may follow one out of four option streams in order to prepare them for tertiary education entry examinations. The four possible streams are science and engineering, medical, humanities, and social science. They may follow an alternative cycle if they do not choose to continue their education at the tertiary level. In the technical-vocational and multibranch schools, a wide range of option cycles of vocational and/or general education is provided.

Students Tested in Mathematics and Science Literacy

Greece participated only in the advanced testing and therefore tested a limited portion of their final-year students in the *Lyceum*. It tested students in Grade 12 of the general (academic) *Lyceum* as well as students in Grade 12 of the multibranch *Lyceum* taking advanced courses in mathematics and/or science in preparation for university disciplines requiring mathematics and/or science.

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: students in their final year, Grade 12, of the general (academic) *Lyceum* and of the multibranch *Lyceum* taking advanced courses in mathematics and/or science in preparation for university disciplines requiring mathematics.

Physics: students in their final year, Grade 12, of the general (academic) *Lyceum* and of the multibranch *Lyceum* taking advanced courses in mathematics and/or science in preparation for university disciplines requiring physics.

Coverage and Exclusions

Greece limited testing to advanced mathematics and physics students in general lycea and multi-branch lycea. Evening classes for adults. Adults attending evening classes were not considered part of the target population.

Sample Design

- Definitions of subpopulations:

  MP   Students in science and engineering branches of general lycea and multi-branch lycea taking advanced mathematics and advanced physics

  Subpopulations not defined further: Greece tested only advanced mathematics and physics students.

- For planning purposes, the final-year population was estimated to be 14.9% MP.
- 60 academic schools (lycea) were sampled with probability proportional to school size from a national list. One class was drawn at random from the final-year classes in each school.

HUNGARY

Structure of Upper Secondary System

The upper secondary system in Hungary consists of five types of schools: a four-year academic secondary school (Grades 9 to 12), a four-year vocational secondary school (Grades 9 to 12), a three-year trade school (Grades 9 to 11), and a six-year or an eight-year academic program (Grades 7 to 12 or 5 to 12). Academic secondary schools offer general education and, for many students, lead to university. Vocational secondary schools prepare students for the work force (often technical vocations) or, alternatively, graduates may enter universities that match their vocational orientation. Trade schools and training schools emphasize practical knowledge and skills to train skilled workers. Students in the trade schools leave school after Grade 10 and spend their final year in out-of-school practice.

Students Tested in Mathematics and Science Literacy

Hungary tested students in their final year of academic secondary and vocational schools (Grade 12) and students in the final in-school year of trade school (Grade 10).

Students Tested in Advanced Mathematics and Physics

Students were not tested in advanced mathematics or physics in Hungary.

Coverage and Exclusions

Very small schools were excluded.

Sample Design

- Definitions of subpopulations:

  No subpopulations were defined: only tested in Mathematics and Science Literacy

- School sample was drawn from list ordered by probability proportional to school size.

- Within each sampled school, one classroom was randomly selected.

ICELAND

Structure of Upper Secondary System

There are four main types of upper secondary schools in Iceland:

1 Grammar schools offer a four-year academic program of study leading to matriculation (stúdentspróf), i.e., higher education entrance examination. Students who complete the course satisfactorily are entitled to apply for admission to university.

2 Industrial-vocational schools primarily offer vocational courses that prepare students for skilled trades. They also offer studies leading to a technical matriculation examination.

3 Comprehensive schools provide academic courses comparable to those of the grammar schools and vocational training comparable to that offered by industrial-vocational schools, as well as other specialized vocational training courses.

4 Specialized vocational schools offer training for specific vocations (Seamen's and navigational colleges, The Fish Processing School, marine engineering colleges, The Technical College of Iceland, fine arts colleges, agricultural colleges, The Icelandic College for Pre-school Teachers, The Icelandic College of Social Pedagogy).

At the upper secondary level, general academic education is primarily organized as a four-year course leading to matriculation, but two-year courses are also offered. The main areas of study of these two-year courses are in education, physical education, and commerce. They are organized as part of the course leading to matriculation (70 units of the 140 required) and students in these shorter courses can therefore continue on to matriculation. Such courses are usually intended as preparatory studies for other courses within the school or at specialized vocational schools.

Traditional grammar schools and upper secondary comprehensive schools are virtually the only schools offering education leading to matriculation. There are basically six courses of academic study leading to matriculation. These are studies in languages, sociology, economics, physical education, natural sciences, and physics. Additional fine arts studies, in music, for example, may lead to matriculation, as does a technical program offered as a follow-up to vocational training.

Vocational training takes place in comprehensive schools, industrial-vocational schools, and specialized vocational schools. Subjects included in vocational programs of study can be grouped as general academic subjects, theoretical vocational subjects, and practical vocational subjects. The length of the courses offered varies from one to ten semesters. Many forms of vocational training award students certification for certain types of employment. This applies especially to study in certified trades, but also to some other studies, such as the training of nurses aides and qualified skippers.

Students Tested in Mathematics and Science Literacy

Iceland tested students who were to graduate that year from an upper secondary school, that is, students in Grades 12, 13, and 14.

Students Tested in Advanced Mathematics and Physics

Students were not tested in advanced mathematics or physics.

Sample Design

- Definitions of subpopulations:

  No subpopulations were defined: only tested in Mathematics and Science Literacy

- All schools and all students in their final year of secondary school were asked to participate.

ISRAEL

Structure of Upper Secondary System

Secondary schools provide three different tracks: academic, technical and vocational, and agricultural. There are four school types: comprehensive (which cater to all three tracks); technical/vocational (vocational track); general schools (academic track); and agricultural schools (agricultural track). Programs are from two to four years and end in Grade 12. Technical education offers a range of courses, including design, computer studies, industrial automation studies, electronics, and telecommunications. Graduates of the technical track are encouraged to serve in technical units of the Israeli defense forces and to continue their studies in institutes of higher education.

Students Tested in Mathematics and Science Literacy

Israel tested students in the Hebrew public education system only. Students in their final year of secondary school, Grade 12, were tested, in all three tracks.

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: students in advanced mathematics courses in Comprehensive and General schools.

Physics: students in physics courses in Comprehensive and General schools.

Coverage and Exclusions

Only schools in the Hebrew public school system were included in the sample. The Jewish Orthodox Independent Education System, special education, and non-Jewish schools were excluded.

Sample Design

- Definitions of subpopulations:

  MO Students in general and comprehensive schools taking advanced mathematics but not physics

  OP Not applicable

  MP Students in general and comprehensive schools taking advanced mathematics and physics

  OO All other students

- For planning purposes, final-year population was estimated to be 70% OO; 20% MO, and 10% MP. The estimated number of students in final year was 67,000. There were estimated to be about 20,000 students in 270 general schools; about 18,000 students in 190 comprehensive schools, and 29,000 students in 330 vocational/agricultural schools.

- The sampling plan was as follows: 40 vocational/agricultural schools, 30 comprehensive schools, and 30 general schools were to be sampled with probability proportional to school size. In each school, one Language class was to be selected for the Mathematics and Science Literacy assessment; advanced mathematics and physics students, if any, were to be removed from sample. In the general and comprehensive schools, one mathematics class was to be drawn and MPs removed from mathematics testing. All MP students would be placed on a list and assigned one of nine test booklets.

- Problems were encountered with test booklet rotation, school tracking forms, classroom tracking forms, and student sampling. It was not possible to compute sampling weights.

ITALY

Structure of Upper Secondary System

There are four upper secondary school types lasting three, four, or five years: classical schools, art schools, technical schools, and vocational schools. Classical schools include the *Liceo Classico*, which prepares humanities students for university; the *Liceo Scientifico*, which prepares mathematics and science students for university; the *Instituto Magistrale* for primary teacher education; the *Scuola Magistrale* for preprimary teacher education; and the *Liceo Linguistico* which prepares language students for university. Art schools, including the *Liceo Artistico* and the *Instituti d'Arte*, train students in the visual arts and lead to university or fine arts academies.

Technical schools, *Instituti Technici*, provide a five-year program to prepare students for professional, technical, or administrative occupations in the agricultural, industrial, or commercial sector. These schools give students access to university. Vocational schools provide a three-year program to train students to become qualified first-level technicians. Students may study an additional two years at *Instituti Professionali* and obtain a "professional maturity" designation, giving access to university.

Students Tested in Mathematics and Science Literacy

Italy tested students in all types of schools in their final year of secondary school. The final grade of school depended on the focus of study within school type. Classical studies: *Liceo Classico* (Grade 13); *Liceo Scientifico* (Grade 13); *Instituto Magistrale* (Grade 12); and *Scuola Magistrale* (Grade 11). Artistic studies: *Liceo Artistico* (Grade 12); *Instituto d'art* (Grade 12); and *Scuola d'art* (Grade 11). Vocational studies: *Instituto Professionale* (Grade 11). Technical studies: *Instituti Technici* (Grade 13).

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: students in their final year of *Liceo Scientifico* (classical schools), Grade 11, 12, or 13, depending on the student's program of study, and *Instituti Technici* (technical schools), Grade 13.

Physics: students in their final year of *Liceo Scientifico* (classical schools), Grade 11, 12, or 13, depending on the student's program of study, and *Instituti Technici* (technical schools), Grade 13.

Coverage and Exclusions

Four geographic regions did not participate. Private schools were excluded.

Sample Design

- Definitions of subpopulations:

  MO Students in *Instituti Technici* taking advanced mathematics but not physics

  OP Students in *Instituti Technici* taking physics but not advanced mathematics

MP Students in *Instituti Technici* and *Liceo Scientifico* taking advanced mathematics and physics

OO All other students

- For planning purposes, final-year population was estimated to have 78,000 MP (17% of 462,000 final-year students in *Liceo Scientifici*); 51,000 MO (11% of final-year students) in *Instituti Tecnici*.

- Schools were sampled with probability proportional to school size; 22 Classical (non-science) schools, 6 Art schools, 64 Technical schools; 25 Classical Scientific schools, and 33 Vocational schools were sampled. In each school, one classroom was selected from the final year.

LATVIA

Structure of Upper Secondary System

After basic education, Latvian students may attend secondary school (Grades 10 to 12), where they enter a three-year academic program to prepare for further studies in higher education or enter a vocational school for two to four years. In the academic secondary program, compulsory subjects include Latvian language and literature, mathematics, a foreign language, world history, Latvian history, and physical education. Optional subjects include the study of a second foreign language, economics, geography, computer science, physics, chemistry, biology, music, nature and society, and others. Vocational schools prepare students for independent technical work in various fields and include technical schools, medical schools, agricultural schools, teacher-training schools, and art schools. Vocational schools include instruction in theory and practice in the vocation of choice and some general education instruction.

Students Tested in Mathematics and Science Literacy

Latvia did not test students in mathematics and science literacy.

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: Latvia did not test students in advanced mathematics.

Physics: students in Grade 12, enrolled in physics courses, in Latvian-speaking academic secondary schools.

Coverage and Exclusions

Latvia tested only physics students in academic secondary schools. Only Latvian-speaking students were included.

Sample Design

- Definitions of subpopulations:

  Subpopulations were not defined; Latvia tested only students taking advanced physics (500 lessons in 3 years) or ordinary physics (300 lessons in 3 years)

- 50 schools offering Advanced Physics were sampled. All students taking Advanced Physics plus one class of Ordinary Physics were sampled.

LITHUANIA

Structure of Upper Secondary System

Upper secondary education in Lithuania includes four-year gymnasia, three-year secondary schools, and two-, three-, or four-year programs in vocational schools. The gymnasium is a four-year educational institution which offers general education at a more advanced level than that in the secondary schools. Traditionally, gymnasia are split into two programs: (1) humanities and (2) mathematics and science. Vocational schools provide general secondary education and training in a profession. There are also "youth schools" for students in basic or secondary school who are, for social reasons, unable to attend general schools. The youth schools provide a one- or two-year program after which students may reenter either the general or vocational schools.

Students Tested in Mathematics and Science Literacy

Lithuania tested students in Grade 12 in vocational, gymnasia, and secondary schools where Lithuanian is the language of instruction.

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: students in their final year, Grade 12, of the mathematics and science gymnasia and students in secondary schools offering enhanced curriculum in mathematics.

Physics: Lithuania did not test students in physics.

Coverage and Exclusions

Schools not under the authority of the Ministry of Education or the Ministry of Science were excluded. Only Lithuanian-speaking students were tested.

Sample Design

- Definitions of subpopulations:

  MO Students taking optional Math 5 and students taking advanced mathematics in specialized schools

  OO All other students

  Subpopulations were not defined further; Lithuania tested only in Advanced Mathematics and Mathematics and Science Literacy

- Used the same school sample as TIMSS Population 2, plus a supplementary sample of 44 vocational schools and all students taking advanced mathematics in specialized schools.

NETHERLANDS

Structure of Upper Secondary System

Secondary education in the Netherlands is four to six years in duration. Students may follow one of four main tracks: pre-university education (VWO); senior general secondary education (HAVO); junior general secondary education (MAVO); or junior secondary vocational education (VBO).

VWO is a six-year program that leads to university or colleges of higher professional education. HAVO is a five-year program designed to prepare students for higher professional education. MAVO is a four-year program after which students may go on to the fourth year of HAVO, take a short or long senior secondary vocational education course (KMBO or MBO), join an apprenticeship course (LLW), or enter the labor market. VBO is a four-year course of prevocational education specializing in technical, home economics, commercial, trade, and agricultural studies. This can lead to a KMBO or MBO course, an apprenticeship course (LLW), or the labor market. As of 1993, a common core curriculum is taught in the first three grades of VBO, MAVO, HAVO, and VWO. The core curriculum includes 15 subjects, among which are mathematics, combined physics and chemistry, biology, and geography (including earth science). This was the structure of the Netherlands' education system at the time of testing (1995). As of August 1997, the MBO, KMBO, and LLW programs are designated as Senior Vocational Education, offering short and long courses on a full-time or part-time basis.

Students Tested in Mathematics and Science Literacy

The Netherlands tested students in the final year, Grade 12, of the six-year VWO (pre-university) program, students in the final year, Grade 11, of the five-year HAVO (senior general secondary) program, and students in the second year, Grade 12, of a two- to four-year MBO or KMBO (senior secondary vocational) program. These latter students would have completed a four-year MAVO program or a four-year VBO program after primary school before beginning the KMBO or MBO program. Students in the LLW (apprenticeship) programs were excluded.

Students Tested in Advanced Mathematics and Physics

The Netherlands did not test students in advanced mathematics or physics.

Coverage and Exclusions

Students in the LWW (apprenticeships) were excluded.

Sample Design

- Definitions of subpopulations:

  MO Grade 12 students in VWO track taking Math A or Math B but not physics

  OP Grade 12 students in VWO track taking advanced physics but not advanced mathematics

MP Grade 12 students in VWO track taking advanced physics and Math A or Math B

OO All other students

Note: Subpopulations were defined, although the Netherlands only tested students in Mathematics and Science Literacy

- For planning purposes, final-year population was estimated to have 35,000 students in VWO (17% OO, 46% OP, 21% MO, and 16% MP); about 44,000 students in HAVO; and close to 91,000 students in MBO students (all OO)
- Schools were sampled from track-specific frames with probability proportional to school size. Within selected schools, lists of classes were compiled and one classroom was selected at random from the available tracks.

NEW ZEALAND

Structure of Upper Secondary System

Secondary education in New Zealand is offered in comprehensive schools from Grades 8 to 12 (Years 9 to 13). At the lower secondary level, students are required to take a number of compulsory subjects in combination with some optional subjects. The diversity of subjects from which students may choose increases in Grades 11 and 12 (Years 12 and 13). Senior students may also be studying subjects at both senior class levels. For example, a student in Grade 12 may take all Grade 12 subjects, or a combination of Grade 11 and Grade 12 subjects.

There are three national awards which students may choose to study for at secondary school, although not all students choose to participate in national examinations. The first, School Certificate, is the national award undertaken by students at the end of their third year of secondary schooling (Grade 10). The second award, Sixth Form Certificate, is undertaken by most students in their fourth year of secondary schooling (Grade 11). Both certificates can be awarded in single subjects, and a candidate may enter in up to six subjects in one year for each award. The third award, University Bursaries/ Entrance Scholarship, is undertaken by the majority of students at the end of Grade 12 (Year 13). Students may elect to sit for examinations in up to five subjects. In addition, students who have completed a five-year course of study are awarded a Higher School Certificate. A student's performance in, for example, School Certificate mathematics and/or science, often determines his/her participation in these national examinations. While participation in national examinations provides an indication of subject choice, it does not, however, include the range of non-assessed courses or school-developed courses undertaken by many students in the senior school.

Students Tested in Mathematics and Science Literacy

New Zealand tested students in Grade 12 and students in Grade 11 who were not returning to school for Grade 12.

Students Tested in Advanced Mathematics and Physics

Students were not tested in advanced mathematics or physics.

Sample Design

- Definitions of subpopulations:

  Subpopulations were not defined: New Zealand tested only for Mathematics and Science Literacy

- For planning purposes, it was estimated that there are 41,000 students in Grade 11 (Form 6, half of whom it was expected would not be returning to school for Grade 12) and 27,000 in Grade 12 (Form 7), for a total target population size of 48,000.

- 79 schools were drawn with probability proportional to school size. Then, 10 Form 6 and 26 Form 7 students were drawn at random from their grade. Form 6 students were asked a screening question to identify who would be coming back the next year (that is, they are not in their final year of secondary school and thus not eligible for testing) and they were eliminated from the estimation procedure.

NORWAY

Structure of Upper Secondary System

Upper secondary education normally covers the 16-19 year age group or the period from the tenth to the twelfth year of education and training, including general and vocational education as well as apprenticeship training.

Under the system for students tested for TIMSS in 1995, general and vocational studies existed side by side in the same school. There were ten areas of study: General (Academic) Studies; Commercial and Clerical Subjects; Physical Education; Craft and Aesthetic Subjects; Home Economics; Technical and Industrial Subjects; Fishing Trade Subjects; Agricultural and Rural Subjects; Maritime Subjects; and Social Studies and Health. The first three areas of study, as well as the music branch within the area of study of Aesthetic Subjects, met the requirements for admission to universities and other higher educational institutions.

This structure was rather complicated, with a varied set of offerings ranging from general schooling to vocational areas of study with special one-, two-, and three-year programs for more than 200 vocational areas.

Beginning in 1994, a simple, comprehensive system for upper secondary school was introduced. All young people between the ages of 16 and 19 have a legal right to three years of upper secondary education, qualifying them for an occupation and/or higher education.

The following three-year programs of study are offered: General and Business Studies; Music, Drama, and Dance Studies; Sports and Physical Education (all three studies qualifying for higher education); Health and Social Studies; Arts, Crafts, and Design Studies; Agriculture, Fishing, and Forestry Studies; Hotel, Cooking, Waiting, and Food Processing Trades; Building and Construction Trades; Service and Technical Building Trades; Electrical Trades; Engineering and Mechanical Trades; Chemical and Processing Trades; Carpentry. (The last ten programs normally qualify students for an occupation.) It has now become much easier for those with a vocational occupation to meet the requirements for entry to higher education. The number of courses in the second and third years are significantly reduced in the new reform.

Students Tested in Mathematics and Science Literacy

Norway tested students in Grade 12 within all areas of study.

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: Norway did not test students in advanced mathematics.

Physics: students in their final year, Grade 12, of the three-year physics course in the General (Academic) Studies area. The three-year course in physics includes a foundation course in general science and two physics courses, normally taken in the second and third year.

Coverage and Exclusions

Sample Design

- Definitions of subpopulations:

  MO Not defined

  OP Grade 12 students in general (academic) program taking three-year (3FY) physics course

  MP Not defined

  OO All other students

- For planning purposes, final-year population was estimated to have 27,500 students in General (academic) studies (68% OO; 15%MO; 1%OP; 16%MP), 9300 students in HK branch (100% OO), 8800 in HI branch (100% OO), and 5800 HS and other branches (100% OO).

- A first probability proportional to school size sample of 80 schools offering three-year physics course was drawn and a three-year physics classroom was selected at random; three physics test booklets and two literacy test booklets were rotated. A second (independent) sample of 60 schools was drawn from the complete list of schools and a class of language arts was selected for the mathematics and science literacy assessment. Duplicate students (if any) were identified and excluded from a second testing.

RUSSIAN FEDERATION

Structure of Upper Secondary System

The upper secondary education system in the Russian Federation is a two- to four-year program following compulsory education. Students in upper secondary school join either the general secondary program (usually two years) or vocational program (two to four years). General secondary includes general schools, schools specializing in specific disciplines, gymnasia, lycea, boarding schools, and schools for children with special needs. There are two possibilities for vocational education: initial vocational education provided in so-called professional-technical schools and secondary vocational education provided in the secondary specialized educational establishments (SSZY, technicums, colleges, etc.). All students in upper secondary education have mathematics and science as compulsory subjects. Graduates from both general secondary and vocational secondary programs may continue their education in universities or other higher educational institutions after passing the entrance examinations.

Students Tested in Mathematics and Science Literacy

The Russian Federation tested students in the final year, Grade 11, of general secondary schools.

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: students in their final year, Grade 11, in general secondary schools in advanced mathematics courses or advanced mathematics and physics courses.

Physics: students in their final year, Grade 11, in general secondary schools in advanced physics courses or advanced mathematics and physics courses.

Coverage and Exclusions

Vocational schools and non-Russian-speaking students excluded.

Sample Design

- Definitions of subpopulations:

  MO — Grade 11 students in general secondary schools in advanced mathematics courses but not in physics courses

  OP — Grade 11 students in general secondary schools in physics courses but not in advanced mathematics courses

  MP — Grade 11 students in general secondary schools in advanced mathematics and physics courses

  OO — All other Grade 11 students in general secondary schools

- For planning purposes, it was estimated that 1,500,000 students graduated in 1993, one-third from the vocational program and two-thirds from general secondary. Of those, about 20% take advanced mathematics or physics.
- A first-stage sample of regions was drawn; then, school frames were set up within each selected region. The sample of schools used for the TIMSS Population 2 assessment was used for the Mathematics and Science Literacy assessment, whenever a selected school included the final-year of secondary school. This sample comprised 165 schools (of the 175 drawn for Population 2). 15 students were selected from the final year. A supplementary probability-proportional-to-size (PPS) sample of 132 schools offering advanced physics and advanced mathematics was selected for the Advanced Mathematics and Physics testing. Local lists of student categorization were compiled (excluding OO students covered by first sample) and up to 15 students were drawn from each list.

SLOVENIA

Structure of Upper Secondary System

There are three types of secondary schools in Slovenia: the four-year gymnasium, the four-year technical and professional school, and the two- or three-year vocational school. Students may write an entrance examination to enter tertiary education after completing any four-year upper secondary school. Gymnasia are in principle comprehensive, but some offer a science-heavy curriculum while others emphasize humanities and languages. All students must study mathematics, physics, chemistry, biology, two foreign languages, and a social sciences program of psychology, sociology, and philosophy. As of 1995, students sit for a five-subject externally assessed baccalaureate examination to enter university. The examination includes Slovenian, mathematics, a foreign language, and two subjects chosen by the student. The technical and professional baccalaureate features the same required subjects as the gymnasia, but students choose from economics, electronics, engineering, or similar subjects for the final two sessions. Vocational schools offer programs from two to four years in duration, and usually involve practical work experience as well as classroom time. All vocational schools end with a final examination that may differ from school to school.

Students Tested in Mathematics and Science Literacy

Students in Grade 12 in gymnasia and in technical secondary schools, as well as students in Grade 11 in vocational schools were tested.

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: students in their final year of gymnasia and technical and professional schools, Grade 12, were tested (all take advanced mathematics).

Physics: students in their final year of gymnasia, Grade 12, taking the physics matura exam were tested.

Coverage and Exclusions

Two-year vocational schools were excluded.

Sample Design

- Definitions of subpopulations:

  MO Grade 12 gymnasium students not taking physics (all take advanced mathematics)

  OP Not applicable

  MP Grade 12 gymnasium students taking physics

  OO Grade 12 students in vocational and technical schools

- For planning purposes, it was estimated that there were 11,300 final-year gymnasium students (9,100 MO and 2,200 MP) and 14,000 final-year vocational and technical students (all OO).

- Original plan was to select 70 gymnasia and 50 technical schools with probability proportional to school size. One class would be drawn at random from the technical schools and 15 MO and 15 MP students would be drawn from the list of grade 12 gymnasium students. Low response rates required that all schools be contacted.

SOUTH AFRICA

Structure of Upper Secondary System

Senior secondary school in South Africa covers Grades 10 to 12. The majority of South African secondary schools are comprehensive. During the first year of senior secondary school (Grade 10), students select six subjects, including the required English and Afrikaans, defining the focus of their studies. Mathematics and science are optional subjects. There are a limited number of schools that provide commercial or technical subjects and a few that provide specialization in the arts. Because of the previous absence of compulsory schooling in South Africa, there is a wide range of entry ages in South African schools, a problem compounded by large numbers of students repeating classes and high drop-out rates.

Students Tested in Mathematics and Science Literacy

Students in Grade 12 were tested in South Africa.

Students Tested in Advanced Mathematics and Physics

South Africa did not test students in advanced mathematics or physics.

Sample Design

- Definitions of subpopulations:

  No subpopulations were defined: South Africa only tested in Mathematics and Science Literacy

- For planning purposes, it was estimated that there were 450,000 students enrolled in Grade 12.

- Schools were stratified by province and sampled within province. Schools in the TIMSS Population 2 (grades 7 and 8) sample were included whenever such schools had students in the final year of secondary school. Supplementary samples were drawn from the provinces of Eastern Cape and North West. The combined samples from Population 2 and the supplements of Eastern Cape and North West comprised 185 schools; of those, 140 offered Grade 12. In many instances, exact measures of size were not known, or were not recorded at the time of data collection. Approximate (provincial averages) sizes had to be used for those undocumented schools.

SWEDEN

Structure of Upper Secondary System

Since 1970, upper secondary school was divided into 47 different lines (*linjer*) and some 400 specialized courses (*specialkurser*). The duration of the lines was two or three years (2-*åriga linjer* and 3-*åriga linjer*, respectively). Thirty-six of the lines were practical/ vocational, and 30 of these were of two years duration. Out of the 11 lines for students preparing for university, 5 were of two years duration. The lines were further divided into branches or variants. A new system of upper secondary education was implemented in the early 1990s and was fully up and running by 1996. The new upper secondary system in Sweden is organized into 16 national study programs of three years duration. Students may also follow a specially designed program or an individual program. All 16 national tracks enable students to attend university, although two tracks, Natural Science and Social Science, are specially-geared towards preparing students for university. All programs include eight core subjects: Swedish, English, civics, religious education, mathematics, general science, physical and health education, and arts activities. At the time of TIMSS testing, some schools were still on the former system where students were in upper secondary for two years, while other schools had switched to the new system of a three-year course.

Students Tested in Mathematics and Science Literacy

In schools where the new three-year upper secondary system was implemented, students in Grade 12 were tested. In schools with the former two- or three-year system, students in the final year, Grade 11 or 12, respectively, were tested.

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: students in the final year, Grade 12, of the Natural Science or Technology lines.

Physics: students in the final year, Grade 12, of the Natural Science or Technology lines.

Sample Design

- Definitions of subpopulations:

  MO Not applicable

  OP Not applicable

  MP Grade 12 students in Natural Science or Technology line

  OO All other students

- For planning purposes, it was estimated that there were 98,000 students enrolled in the final year, with 16,600 OO (17%) and 81,340 MP (83%).

- To obtain reasonable measures of size, some pseudo-schools were formed by joining the smallest schools.

- 150 schools sampled with probability proportional to school size from a complete list of schools. This translated into 157 schools including the schools making up the pseudo-schools. In each sampled school, one class of MP students was drawn at random and a systematic sample of 20 OO students was also drawn.

SWITZERLAND

Structure of Upper Secondary System

Upper secondary education in Switzerland is divided into four major types that last between two to five years, depending on the type and canton. The four types are: *Maturitätsschule* (gymnasium); general education; vocational training; and teacher training. Each major track is differentiated into a number of tracks with narrower definitions. The *Maturitätsschule* is designed to prepare students for university entrance. Typically, students enter at age 15/16, for a total of four years. The school leaving certificate gives them access to higher education. There are five types of *Maturitätsschule*: Type A (emphasis on Greek and Latin); Type B (Latin and modern languages); Type C (mathematics and science); Type D (modern languages); and Type E (economics). *Maturitätsschulen* are governed by federal regulation. The final grade in this type of school could be Grade 12, 12.5, or 13, depending on the canton.

General education schools provide general education to prepare students for certain non-university professions (such as paramedical and social fields). These programs are two or three years in duration and comprise about 3 percent of the in-school population. The upper secondary teacher training program is a five-year program that begins after compulsory education and can lead to university studies.

Vocational training is mostly in the form of apprenticeship, consisting of two basic elements: practical training on the job in an enterprise (3.5 to 4 days per week), and theoretical and general instruction in a vocational school (1 to 1.5 days per week). Vocational training is regulated by federal law and provides recognized apprenticeships of two to four years duration in approximately 280 vocations in the industrial, handicraft, and service sectors. Some students do go on to specialized tertiary institutes in the corresponding vocational field. The final year of vocational training varies by occupation.

Students Tested in Mathematics and Science Literacy

Students in their final year of gymnasium, general education, teacher training, and vocational training were tested. This corresponded to Grade 11 or 12 in gymnasium (final year depends on the canton); Grade 12 in the general track; Grade 12 in the teacher-training track; and Grade 11, 12, or 13 in vocational track (final year varies by occupation).

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: students in their final year, Grade 12 or 13, of *Maturitätsschule* (gymnasium), in schools and programs (A-E) with federal recognition.

Physics: students in their final year, Grade 12 or 13, of *Maturitätsschule* (gymnasium), in schools and programs (A-E) with federal recognition.

Coverage and Exclusions

Montessori schools and special needs students.

Sample Design

- Definitions of subpopulations:

  MO Not applicable

  OP Not applicable

  MP Students in last year of scientific track of academic schools taking 7 lessons of mathematics and at least 3 lessons of physics per week or students in the last year of other tracks in academic schools taking 4 lessons of mathematics and 2 of physics.

  OO All other students

- For planning purposes, it was estimated that there were 3,300 students (20% MP) in the scientific track and 8,000 students (about 2% MP) in other academic tracks. Total enrollment was estimated to be 71,800.

- Schools were sampled from region-by-track specific frames with probability proportional to school size. Within selected schools, lists of classes were established and one classroom was selected at random. Some schools were included with certainty.

UNITED STATES

Structure of Upper Secondary System

Secondary education in the United States is comprehensive and lasts from Grade 9 to 12 or 10 to 12. Students attend high schools that offer a wide variety of courses. Each student chooses or is guided in the selection of an individually unique set of courses based on their personal interests, future aspirations, or ability. Students who choose a higher proportion of courses which prepare them for university study are generally said to be in a college preparatory or "academic" school program. Those who choose a higher proportion of vocational courses are in a vocational/technical or "vocational" school program. Those whose choice of courses combines general academic and vocational coursework are in general academic or "general" school programs.

Students Tested in Mathematics and Science Literacy

Students in Grade 12 were tested in the United States.

Students Tested in Advanced Mathematics and Physics

Advanced Mathematics: students in Grade 12 who had taken Advanced Placement Calculus, Calculus, or Pre-Calculus.

Physics: students in Grade 12 who had taken Advanced Placement Physics or Physics.

Coverage and Exclusions

Sample Design

- Definitions of subpopulations:

  MO Students having taken Advanced Placement Calculus, Calculus, or Pre-Calculus during grades 9 - 12 but not physics

  OP Students having taken Advanced Placement Physics or Physics during grades 9 - 12 but not advanced mathematics

  MP Students having taken Advanced Placement Calculus, Calculus, or Pre-Calculus and Advanced Placement Physics or Physics during grades 9 - 12

  OO Students not having taken Advanced Placement Calculus, Calculus, or Pre-Calculus, or Advanced Placement Physics or Physics during grades 9 - 12

- Target population was 72.9% OO; 9.0% OP; 6.7% MO; and 11.3% MP

- A three-stage stratified design was used. Geographic regions were the primary sampling units: 48 non-certainty and 11 certainty primary sampling units were drawn for the first-stage sample. Then, 250 schools were selected with probability proportional to school size from stratified lists within the selected primary sampling units. Finally, students were selected from school-level subpopulation lists.

Appendix C: Sampling and Imputation Standard Errors by Gender Tables

Table C.1 Sampling and Imputation Standard Errors
Mathematics and Science Literacy Scale
Males in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	775	71070	543	10.7	1.2	10.7	99.8	5.8
Austria	878	26595	549	7.7	1.4	7.8	84.9	4.5
Canada	2672	122763	544	3.3	0.8	3.4	84.0	2.2
Cyprus	251	2045	456	4.6	1.8	4.9	78.0	3.9
Czech Republic	1115	70435	500	9.9	0.9	9.9	92.9	3.5
Denmark	1206	16777	554	4.3	1.0	4.5	80.0	3.0
France	813	298449	526	5.8	1.2	5.9	75.1	3.6
Germany	1071	523599	512	7.9	2.1	8.2	85.6	4.0
Hungary	2370	56115	485	4.4	0.8	4.5	90.6	3.0
Iceland	800	1085	565	2.7	0.9	2.9	76.5	2.0
Italy	776	174400	492	6.8	1.0	6.9	85.6	4.8
Lithuania	948	7657	483	6.7	0.5	6.7	75.8	3.3
Netherlands	745	75222	584	5.4	0.8	5.5	78.1	4.2
New Zealand	852	18434	540	5.5	1.6	5.7	96.6	3.3
Norway	1190	22142	564	5.0	0.7	5.0	89.1	3.1
Russian Federation	841	392477	499	5.7	1.5	5.9	81.0	3.3
Slovenia	828	13336	538	12.5	1.1	12.6	84.1	8.3
South Africa	1315	178372	366	10.3	0.6	10.3	87.8	8.4
Sweden	1462	34787	579	5.8	1.2	5.9	96.0	2.8
Switzerland	1660	36160	547	5.9	1.2	6.0	86.7	3.4
United States	2839	1133206	479	4.1	0.9	4.2	92.6	2.4

S.D. = standard deviation
S.E. = standard error

Table C.2 Sampling and Imputation Standard Errors Mathematics and Science Literacy Scale Females in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	1166	99777	511	9.3	0.2	9.3	88.8	5.4
Austria	1041	42428	502	5.4	1.0	5.5	71.3	3.2
Canada	2533	137972	511	3.2	1.0	3.4	79.3	1.9
Cyprus	280	2491	439	2.9	0.8	3.0	67.3	2.9
Czech Republic	1052	67024	452	13.7	1.0	13.8	83.8	3.7
Denmark	1446	20329	507	3.6	0.8	3.7	75.8	2.5
France	759	331891	487	4.7	1.1	4.8	68.0	2.6
Germany	1108	416372	479	8.4	1.5	8.5	88.9	4.6
Hungary	2542	51285	468	4.4	0.6	4.5	76.1	2.3
Iceland	887	1199	522	1.8	0.7	1.9	71.9	1.3
Italy	840	206434	461	5.7	0.6	5.7	78.3	5.0
Lithuania	1938	14489	456	7.4	0.5	7.4	81.3	3.5
Netherlands	725	70694	533	5.7	1.3	5.9	82.4	4.7
New Zealand	911	19115	511	5.4	1.1	5.5	85.0	3.1
Norway	1328	21664	507	4.5	0.4	4.5	76.2	2.6
Russian Federation	1448	638710	462	6.5	0.7	6.5	81.2	3.6
Slovenia	735	12601	492	7.1	0.5	7.1	72.6	3.8
South Africa	1370	187962	341	11.8	0.9	11.8	87.1	13.6
Sweden	1606	36457	533	3.6	0.3	3.6	80.3	2.2
Switzerland	1623	28240	511	7.4	1.4	7.5	84.7	2.9
United States	2968	1145052	462	3.5	0.2	3.5	84.7	3.0

S.D. = standard deviation
S.E. = standard error

Table C.3 Sampling and Imputation Standard Errors - Mathematic Literacy Scale Males in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	775	71070	540	10.3	0.5	10.3	102.8	6.1
Austria	878	26595	545	7.1	1.2	7.2	82.2	4.1
Canada	2672	122763	537	3.7	1.1	3.8	90.7	2.7
Cyprus	251	2045	454	4.5	1.8	4.9	77.8	4.0
Czech Republic	1115	70435	488	11.2	1.1	11.3	101.3	4.0
Denmark	1206	16777	575	3.9	0.8	4.0	83.8	3.8
France	813	298449	544	5.6	1.0	5.6	79.1	3.6
Germany	1071	523599	509	8.3	2.8	8.7	91.0	4.4
Hungary	2370	56115	485	4.8	0.9	4.9	98.9	3.0
Iceland	800	1085	558	3.4	0.6	3.4	86.5	2.4
Italy	776	174400	490	7.2	1.5	7.4	89.6	5.0
Lithuania	948	7657	485	7.3	0.7	7.3	80.3	4.2
Netherlands	745	75222	585	5.5	1.0	5.6	81.6	3.8
New Zealand	852	18434	536	4.5	1.8	4.9	100.9	3.0
Norway	1190	22142	555	5.2	0.9	5.3	95.0	2.9
Russian Federation	841	392477	488	6.4	1.2	6.5	85.6	3.5
Slovenia	828	13336	535	12.6	1.1	12.7	87.2	8.9
South Africa	1315	178372	365	9.3	0.9	9.3	83.3	8.2
Sweden	1462	34787	573	5.9	1.0	5.9	103.2	3.0
Switzerland	1660	36160	555	6.3	1.2	6.4	87.9	3.6
United States	2839	1133206	466	4.0	0.8	4.1	94.2	2.6

S.D. = standard deviation
S.E. = standard error

Table C.4 Sampling and Imputation Standard Errors - Mathematic Literacy Scale Females in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	1166	99777	510	9.3	0.7	9.3	90.9	5.1
Austria	1041	42428	503	5.4	1.4	5.5	73.5	2.9
Canada	2533	137972	504	3.3	1.2	3.5	86.7	2.6
Cyprus	280	2491	439	3.5	1.0	3.7	67.6	2.9
Czech Republic	1052	67024	443	16.7	1.9	16.8	91.8	3.6
Denmark	1446	20329	523	3.9	0.9	4.0	81.9	2.6
France	759	331891	506	5.1	1.3	5.3	74.9	2.8
Germany	1108	416372	480	8.7	1.6	8.8	93.9	4.5
Hungary	2542	51285	481	4.8	0.6	4.8	84.6	2.3
Iceland	887	1199	514	2.1	0.7	2.2	83.5	1.2
Italy	840	206434	464	5.9	1.0	6.0	83.7	5.2
Lithuania	1938	14489	461	7.7	0.8	7.7	85.8	3.6
Netherlands	725	70694	533	5.8	1.3	5.9	90.4	4.4
New Zealand	911	19115	507	6.0	1.6	6.2	93.2	3.0
Norway	1328	21664	501	4.8	0.3	4.8	84.5	2.5
Russian Federation	1448	638710	460	6.5	1.2	6.6	83.7	3.9
Slovenia	735	12601	490	8.0	0.5	8.0	79.4	4.6
South Africa	1370	187962	348	10.7	1.0	10.8	79.9	13.3
Sweden	1606	36457	531	3.8	0.4	3.9	89.4	2.4
Switzerland	1623	28240	522	7.2	1.4	7.4	85.9	2.9
United States	2968	1145052	456	3.5	0.7	3.6	87.6	2.6

S.D. = standard deviation
S.E. = standard error

Table C.5 Sampling and Imputation Standard Errors - Science Literacy Scale Males in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	775	71070	547	11.3	2.1	11.5	103.8	5.6
Austria	878	26595	554	8.5	1.8	8.7	93.6	5.0
Canada	2672	122763	550	3.4	1.3	3.6	86.1	2.2
Cyprus	251	2045	459	5.4	2.1	5.8	88.5	4.6
Czech Republic	1115	70435	512	8.8	0.7	8.8	91.2	3.2
Denmark	1206	16777	532	5.2	1.4	5.4	87.1	3.3
France	813	298449	508	6.4	2.0	6.7	81.1	3.4
Germany	1071	523599	514	7.7	1.6	7.9	86.8	3.9
Hungary	2370	56115	484	4.1	0.8	4.2	91.1	3.0
Iceland	800	1085	572	2.4	1.2	2.7	75.5	1.9
Italy	776	174400	495	6.6	1.2	6.7	88.5	4.9
Lithuania	948	7657	481	6.4	0.5	6.4	79.3	2.9
Netherlands	745	75222	582	5.6	1.3	5.7	81.9	4.9
New Zealand	852	18434	543	6.8	1.9	7.1	100.0	4.7
Norway	1190	22142	574	5.0	0.9	5.1	92.7	3.6
Russian Federation	841	392477	510	5.4	1.9	5.7	86.2	3.7
Slovenia	828	13336	541	12.7	1.1	12.7	87.2	7.8
South Africa	1315	178372	367	11.4	1.0	11.5	98.2	8.5
Sweden	1462	34787	585	5.8	1.4	6.0	95.4	2.8
Switzerland	1660	36160	540	5.9	1.4	6.1	92.5	3.3
United States	2839	1133206	492	4.4	1.1	4.5	97.6	2.7

S.D. = standard deviation
S.E. = standard error

Table C.6 Sampling and Imputation Standard Errors - Science Literacy Scale Females in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	1166	99777	513	9.4	0.6	9.4	93.9	5.9
Austria	1041	42428	501	5.6	1.4	5.8	75.5	3.4
Canada	2533	137972	518	3.7	0.9	3.8	80.3	2.2
Cyprus	280	2491	439	3.0	0.6	3.0	76.4	3.6
Czech Republic	1052	67024	460	10.9	0.7	11.0	83.5	3.6
Denmark	1446	20329	490	4.0	1.0	4.1	82.1	2.8
France	759	331891	468	4.7	1.0	4.8	71.4	2.4
Germany	1108	416372	478	8.4	1.4	8.5	90.7	4.7
Hungary	2542	51285	455	4.2	0.8	4.3	77.9	2.3
Iceland	887	1199	530	1.9	0.8	2.1	68.8	1.8
Italy	840	206434	458	5.6	0.4	5.6	81.5	4.6
Lithuania	1938	14489	450	7.3	0.5	7.3	84.4	3.6
Netherlands	725	70694	532	6.0	1.6	6.2	82.0	5.2
New Zealand	911	19115	515	5.1	0.7	5.2	86.5	3.8
Norway	1328	21664	513	4.5	0.5	4.5	78.7	2.7
Russian Federation	1448	638710	463	6.6	1.0	6.7	89.0	3.2
Slovenia	735	12601	494	6.3	1.0	6.4	71.9	3.4
South Africa	1370	187962	333	12.9	1.5	13.0	99.9	13.5
Sweden	1606	36457	534	3.4	0.5	3.5	79.1	2.2
Switzerland	1623	28240	500	7.7	1.6	7.8	90.4	3.4
United States	2968	1145052	469	3.8	0.7	3.9	88.6	3.5

S.D. = standard deviation
S.E. = standard error

Table C.7 Sampling and Imputation Standard Errors - Advanced Mathematics Scale Males in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	360	21889	531	11.1	2.2	11.4	107.7	9.0
Austria	299	11492	486	6.9	2.1	7.3	76.1	5.4
Canada	1474	30953	528	6.2	1.5	6.4	102.7	2.9
Cyprus	237	509	524	4.2	1.4	4.4	90.4	3.9
Czech Republic	451	7877	524	12.9	1.0	13.0	106.2	12.0
Denmark	853	8193	529	4.2	1.3	4.4	76.0	2.3
France	665	93959	567	4.9	1.5	5.1	69.8	2.6
Germany	832	109689	484	6.4	0.7	6.5	86.2	4.1
Greece	316	10121	516	6.5	1.1	6.6	111.2	7.5
Italy	258	63223	484	10.6	0.8	10.6	94.2	8.7
Lithuania	372	691	542	3.4	1.5	3.7	84.1	3.8
Russian Federation	908	22259	568	9.7	1.1	9.7	110.6	4.4
Slovenia	746	11199	484	11.5	0.9	11.5	96.9	5.4
Sweden	644	11309	519	5.7	1.8	5.9	88.2	3.6
Switzerland	766	6002	559	5.1	2.4	5.6	93.3	3.9
United States	1417	254188	457	7.6	1.9	7.8	95.9	4.8

S.D. = standard deviation
S.E. = standard error

Table C.8 Sampling and Imputation Standard Errors - Advanced Mathematics Scale Females in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	285	17609	517	14.7	3.4	15.1	110.4	9.3
Austria	464	18837	406	8.6	0.7	8.6	86.7	6.6
Canada	1298	27551	489	4.3	0.8	4.4	88.9	2.7
Cyprus	151	322	509	6.0	2.3	6.4	76.6	4.9
Czech Republic	650	11569	432	8.6	2.1	8.9	88.6	6.4
Denmark	479	4860	510	4.5	1.1	4.6	67.6	3.4
France	390	55043	543	5.0	0.9	5.1	67.2	2.9
Germany	1418	147800	452	6.5	1.0	6.6	81.2	3.9
Greece	138	4455	505	9.6	3.5	10.2	88.0	8.5
Italy	140	41254	460	14.0	1.8	14.1	94.9	13.1
Lithuania	362	669	490	5.5	0.8	5.6	78.0	6.8
Russian Federation	730	20599	515	10.1	1.5	10.2	106.0	8.0
Slovenia	775	11340	464	10.9	1.3	11.0	89.1	3.5
Sweden	357	5099	496	4.5	2.7	5.2	77.6	4.5
Switzerland	623	5216	503	5.4	1.5	5.7	76.6	4.9
United States	1368	242664	426	6.9	1.7	7.1	97.6	5.6

S.D. = standard deviation
S.E. = standard error

Table C.9 Sampling and Imputation Standard Errors - Numbers and Equations
Males in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	360	21889	523	9.7	1.5	9.9	98.4	9.1
Austria	299	11492	455	5.9	1.7	6.2	77.1	5.1
Canada	1474	30953	526	5.4	1.5	5.6	88.3	2.7
Cyprus	237	509	518	6.4	1.2	6.5	97.5	3.4
Czech Republic	451	7877	510	11.1	1.7	11.3	99.0	7.6
Denmark	853	8193	507	3.5	0.9	3.6	63.9	2.2
France	665	93959	551	5.3	1.0	5.4	58.6	3.3
Germany	832	109689	475	6.0	1.6	6.2	79.9	4.4
Greece	316	10121	540	8.7	2.6	9.1	121.4	9.9
Italy	258	63223	472	10.6	0.5	10.6	100.9	8.8
Lithuania	372	691	568	2.8	1.0	3.0	82.7	4.8
Russian Federation	908	22259	576	9.5	1.8	9.6	104.1	3.8
Slovenia	746	11199	503	12.9	1.6	13.0	111.8	7.6
Sweden	644	11309	529	6.2	1.3	6.4	91.4	3.9
Switzerland	766	6002	536	5.4	1.9	5.7	92.5	4.6
United States	1417	254188	470	5.8	1.9	6.1	84.8	4.0

S.D. = standard deviation
S.E. = standard error

Table C.10 Sampling and Imputation Standard Errors - Numbers and Equations
Females in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	285	17609	511	10.8	3.0	11.2	97.1	8.2
Austria	464	18837	385	9.2	1.1	9.3	89.6	6.9
Canada	1298	27551	496	4.2	1.6	4.5	76.5	3.6
Cyprus	151	322	497	6.4	2.8	7.0	83.3	5.2
Czech Republic	650	11569	427	10.4	1.6	10.5	92.6	7.1
Denmark	479	4860	498	3.2	1.3	3.5	59.3	3.1
France	390	55043	544	3.7	1.0	3.9	51.8	3.1
Germany	1418	147800	446	5.1	0.5	5.1	76.4	3.8
Greece	138	4455	537	10.1	2.5	10.4	90.3	7.2
Italy	140	41254	441	14.1	0.5	14.1	103.8	15.0
Lithuania	362	669	526	4.4	3.0	5.4	81.0	3.9
Russian Federation	730	20599	533	9.4	3.1	9.8	105.0	7.6
Slovenia	775	11340	480	10.8	0.8	10.8	97.5	3.5
Sweden	357	5099	511	5.0	2.5	5.6	80.1	5.4
Switzerland	623	5216	488	5.6	1.2	5.7	73.9	5.6
United States	1368	242664	447	6.7	1.8	6.9	87.0	5.5

S.D. = standard deviation
S.E. = standard error

Table C.11 Sampling and Imputation Standard Errors - Calculus Scale Males in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	360	21889	533	13.6	1.5	13.6	102.5	14.0
Austria	299	11492	486	6.5	2.2	6.9	73.1	7.4
Canada	1474	30953	521	5.4	0.9	5.5	97.2	3.8
Cyprus	237	509	559	4.9	1.0	5.0	101.2	5.2
Czech Republic	451	7877	488	10.9	0.9	11.0	102.0	9.8
Denmark	853	8193	517	4.1	1.0	4.3	87.0	2.9
France	665	93959	569	4.2	0.8	4.3	65.2	2.8
Germany	832	109689	471	5.6	0.8	5.6	87.6	4.4
Greece	316	10121	540	7.8	2.7	8.2	104.4	6.3
Italy	258	63223	520	11.4	0.9	11.4	106.8	8.8
Lithuania	372	691	518	3.9	1.8	4.3	79.4	3.3
Russian Federation	908	22259	560	8.9	0.8	8.9	105.9	5.6
Slovenia	746	11199	479	8.2	0.9	8.2	73.7	3.8
Sweden	644	11309	484	5.8	1.4	6.0	91.8	3.7
Switzerland	766	6002	536	6.3	2.4	6.8	99.2	4.1
United States	1417	254188	460	4.6	2.6	5.3	95.6	4.5

S.D. = standard deviation
S.E. = standard error

Table C.12 Sampling and Imputation Standard Errors - Calculus Scale Females in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	285	17609	525	11.9	2.7	12.2	97.4	9.4
Austria	464	18837	412	7.3	0.8	7.3	83.4	4.8
Canada	1298	27551	484	4.2	2.6	4.9	83.4	4.2
Cyprus	151	322	562	7.7	2.3	8.0	98.2	4.8
Czech Republic	650	11569	417	8.2	1.6	8.3	82.2	4.7
Denmark	479	4860	491	5.2	1.3	5.4	82.8	3.6
France	390	55043	544	3.9	1.2	4.1	59.3	2.5
Germany	1418	147800	442	5.0	1.1	5.2	81.3	2.9
Greece	138	4455	536	11.3	4.0	12.0	83.0	10.4
Italy	140	41254	521	13.4	0.8	13.5	108.1	7.4
Lithuania	362	669	478	4.4	1.7	4.8	71.2	4.9
Russian Federation	730	20599	512	10.6	2.5	10.9	101.2	9.4
Slovenia	775	11340	463	7.9	0.8	7.9	67.1	2.4
Sweden	357	5099	472	4.5	2.0	4.9	78.0	4.2
Switzerland	623	5216	486	6.1	1.0	6.2	85.7	5.7
United States	1368	242664	439	5.9	1.6	6.1	97.9	6.5

S.D. = standard deviation
S.E. = standard error

Table C.13 Sampling and Imputation Standard Errors - Geometry Scale Males in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	360	21889	505	14.0	1.5	14.1	122.6	14.1
Austria	299	11492	509	7.5	1.8	7.7	82.1	4.6
Canada	1474	30953	516	4.7	2.4	5.3	97.1	2.8
Cyprus	237	509	520	5.1	1.0	5.2	104.2	5.6
Czech Republic	451	7877	543	12.0	1.4	12.1	105.2	11.8
Denmark	853	8193	531	4.0	1.3	4.2	74.6	2.6
France	665	93959	555	5.3	2.1	5.7	77.3	2.7
Germany	832	109689	498	6.9	1.4	7.0	77.0	4.4
Greece	316	10121	505	7.0	2.8	7.5	119.4	6.3
Italy	258	63223	485	10.4	0.9	10.4	102.5	9.0
Lithuania	372	691	539	3.2	1.7	3.6	82.8	2.9
Russian Federation	908	22259	570	8.8	0.9	8.9	102.8	4.5
Slovenia	746	11199	482	9.5	1.5	9.6	86.0	4.4
Sweden	644	11309	500	5.4	1.1	5.5	85.0	3.4
Switzerland	766	6002	569	3.2	1.9	3.8	86.8	4.3
United States	1417	254188	439	5.4	2.0	5.8	94.2	5.3

S.D. = standard deviation
S.E. = standard error

Table C.14 Sampling and Imputation Standard Errors - Geometry Scale Females in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	285	17609	485	13.6	2.8	13.8	121.5	8.1
Austria	464	18837	433	9.5	1.3	9.6	95.6	8.3
Canada	1298	27551	482	4.2	2.0	4.6	88.9	2.4
Cyprus	151	322	512	8.1	2.8	8.5	90.8	7.8
Czech Republic	650	11569	461	7.2	0.8	7.2	85.3	5.8
Denmark	479	4860	519	3.6	1.7	4.0	63.9	3.6
France	390	55043	529	4.5	1.6	4.8	71.6	3.6
Germany	1418	147800	480	5.5	0.7	5.6	72.7	4.2
Greece	138	4455	485	14.6	4.8	15.4	102.4	13.6
Italy	140	41254	472	14.5	0.4	14.5	105.1	14.7
Lithuania	362	669	491	4.9	3.1	5.8	74.3	5.6
Russian Federation	730	20599	525	9.9	3.4	10.5	99.2	7.4
Slovenia	775	11340	469	8.9	0.5	8.9	80.0	2.9
Sweden	357	5099	476	4.7	2.0	5.1	75.9	5.5
Switzerland	623	5216	522	5.6	1.8	5.9	80.1	5.9
United States	1368	242664	408	6.8	1.3	7.0	96.1	6.3

S.D. = standard deviation
S.E. = standard error

Table C.15 Sampling and Imputation Standard Errors - Physics Scale Males in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	417	20973	532	6.5	1.4	6.7	82.3	5.6
Austria	306	11559	479	8.1	0.7	8.1	82.5	5.7
Canada	1440	28901	506	5.0	3.2	6.0	89.7	4.2
Cyprus	230	523	509	8.8	1.0	8.9	107.8	7.9
Czech Republic	426	7460	503	8.7	1.4	8.8	83.1	5.4
Denmark	480	1547	542	4.7	2.3	5.2	86.9	4.4
France	670	90387	478	3.9	1.5	4.2	67.0	4.4
Germany	487	59545	542	14.3	1.5	14.3	92.7	6.9
Greece	311	10015	495	5.9	1.4	6.1	89.7	5.0
Latvia (LSS)	374	495	509	18.9	1.5	19.0	99.5	11.5
Norway	781	3221	594	5.8	2.3	6.3	87.6	2.5
Russian Federation	714	17949	575	9.8	1.5	9.9	102.8	3.8
Slovenia	566	8274	546	16.1	2.1	16.3	99.4	10.8
Sweden	651	11056	589	4.8	1.7	5.1	94.1	3.7
Switzerland	727	5662	529	4.9	1.5	5.2	85.8	4.0
United States	1617	270205	439	4.2	1.1	4.3	61.5	5.0

S.D. = standard deviation
S.E. = standard error

Table C.16 Sampling and Imputation Standard Errors - Physics Scale Females in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	244	10647	490	7.3	4.2	8.4	74.6	5.3
Austria	457	18721	408	7.3	1.0	7.4	71.5	5.9
Canada	913	22098	459	6.0	1.8	6.3	74.9	3.9
Cyprus	137	311	470	6.9	1.7	7.1	96.0	7.9
Czech Republic	661	11968	419	3.9	0.4	3.9	63.2	5.1
Denmark	117	394	500	6.0	5.4	8.1	73.9	6.8
France	417	57604	450	5.0	2.3	5.6	60.8	3.2
Germany	222	26328	479	9.1	1.1	9.1	80.3	5.3
Greece	148	4652	468	7.9	1.8	8.1	78.6	6.9
Latvia (LSS)	334	484	467	22.6	1.9	22.6	96.6	11.4
Norway	267	1148	544	9.0	2.4	9.3	88.3	4.5
Russian Federation	519	15026	509	14.8	3.8	15.3	108.1	9.1
Slovenia	160	3162	455	18.3	3.9	18.7	105.6	6.4
Sweden	361	5402	540	5.1	1.4	5.3	77.8	4.8
Switzerland	632	5495	446	3.5	0.8	3.6	68.8	2.9
United States	1497	252579	405	2.8	1.4	3.1	53.0	1.8

S.D. = standard deviation
S.E. = standard error

Table C.17 Sampling and Imputation Standard Errors - Mechanics Scale Males in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	417	20973	524	7.7	1.1	7.8	84.9	6.8
Austria	306	11559	459	6.6	0.9	6.6	74.1	5.4
Canada	1440	28901	499	6.5	1.1	6.6	92.1	4.9
Cyprus	230	523	551	9.6	0.5	9.6	120.1	7.6
Czech Republic	426	7460	514	8.3	1.6	8.4	80.8	6.7
Denmark	480	1547	540	5.5	0.9	5.5	87.9	4.2
France	670	90387	470	5.2	1.9	5.6	76.4	4.5
Germany	487	59545	515	9.5	1.5	9.6	84.2	7.7
Greece	311	10015	525	6.9	1.5	7.0	91.6	5.4
Latvia (LSS)	374	495	509	15.1	1.3	15.2	93.4	13.1
Norway	781	3221	589	5.5	2.6	6.1	84.0	3.5
Russian Federation	714	17949	563	7.4	0.2	7.4	84.3	4.7
Slovenia	566	8274	576	17.4	1.9	17.5	111.3	12.6
Sweden	651	11056	586	4.5	0.6	4.6	76.0	3.1
Switzerland	727	5662	519	5.3	0.6	5.3	82.4	4.5
United States	1617	270205	446	3.2	1.5	3.5	54.1	3.7

S.D. = standard deviation
S.E. = standard error

Table C.18 Sampling and Imputation Standard Errors - Mechanics Scale Females in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	244	10647	474	6.6	1.6	6.8	83.4	8.8
Austria	457	18721	399	6.2	0.8	6.3	70.3	5.0
Canada	913	22098	440	5.4	1.8	5.7	72.8	3.7
Cyprus	137	311	496	10.3	0.5	10.3	102.5	8.6
Czech Republic	661	11968	440	4.7	1.0	4.8	66.1	5.3
Denmark	117	394	483	9.2	4.3	10.2	80.4	5.4
France	417	57604	437	3.8	3.9	5.5	68.0	4.6
Germany	222	26328	453	10.4	1.8	10.6	88.0	6.8
Greece	148	4652	489	6.7	2.7	7.2	83.6	7.3
Latvia (LSS)	334	484	468	19.6	2.2	19.8	84.5	6.1
Norway	267	1148	523	8.7	2.4	9.0	86.4	6.7
Russian Federation	519	15026	507	12.2	1.1	12.3	90.3	10.5
Slovenia	160	3162	487	21.6	2.4	21.7	122.6	10.4
Sweden	361	5402	517	4.0	1.7	4.4	68.7	4.4
Switzerland	632	5495	444	3.5	0.3	3.5	70.9	2.7
United States	1497	252579	393	2.6	1.0	2.8	51.4	2.0

S.D. = standard deviation
S.E. = standard error

Table C.19 Sampling and Imputation Standard Errors - Electricity and Magnetism Scale Males in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	417	20973	525	6.6	1.2	6.7	90.5	6.3
Austria	306	11559	468	9.0	1.3	9.1	99.8	6.6
Canada	1440	28901	497	6.0	1.8	6.2	88.3	5.1
Cyprus	230	523	507	8.4	1.1	8.5	118.7	8.2
Czech Republic	426	7460	501	8.5	1.9	8.7	79.5	6.8
Denmark	480	1547	515	4.3	1.2	4.5	84.0	5.6
France	670	90387	495	4.0	1.2	4.2	61.2	3.5
Germany	487	59545	522	11.8	2.6	12.1	93.7	6.3
Greece	311	10015	522	6.2	1.8	6.5	109.6	5.4
Latvia (LSS)	374	495	496	16.6	2.5	16.8	96.9	10.2
Norway	781	3221	570	5.9	2.1	6.2	93.1	3.7
Russian Federation	714	17949	575	7.6	1.1	7.7	101.5	3.5
Slovenia	566	8274	522	16.4	2.1	16.6	107.3	13.1
Sweden	651	11056	579	4.7	1.0	4.8	92.8	3.8
Switzerland	727	5662	507	7.0	1.2	7.1	95.8	4.4
United States	1617	270205	430	3.4	0.8	3.5	59.2	3.4

S.D. = standard deviation
S.E. = standard error

Table C.20 Sampling and Imputation Standard Errors - Electricity and Magnetism Scale Females in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	244	10647	488	7.8	2.6	8.3	86.9	7.9
Austria	457	18721	409	6.8	1.5	6.9	82.5	4.7
Canada	913	22098	468	6.3	1.6	6.5	73.1	4.5
Cyprus	137	311	494	7.3	1.2	7.4	106.9	9.5
Czech Republic	661	11968	443	3.1	1.1	3.3	62.7	5.0
Denmark	117	394	498	6.9	3.8	7.8	68.9	4.7
France	417	57604	491	4.8	1.9	5.2	57.8	3.8
Germany	222	26328	491	7.4	1.9	7.7	79.3	6.2
Greece	148	4652	515	10.3	3.9	11.0	94.2	8.0
Latvia (LSS)	334	484	474	17.9	3.9	18.4	90.1	6.7
Norway	267	1148	549	9.7	2.4	10.0	90.5	5.9
Russian Federation	519	15026	519	12.5	3.2	12.9	105.7	8.1
Slovenia	160	3162	470	13.7	2.0	13.8	108.4	12.2
Sweden	361	5402	551	4.4	1.7	4.7	74.1	4.8
Switzerland	632	5495	452	4.3	1.4	4.5	83.4	3.8
United States	1497	252579	409	3.5	0.8	3.6	56.0	2.0

S.D. = standard deviation
S.E. = standard error

Table C.21 Sampling and Imputation Standard Errors - Heat Scale Males in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	417	20973	524	4.9	1.2	5.0	84.1	5.6
Austria	306	11559	485	7.9	1.2	8.0	92.0	7.3
Canada	1440	28901	520	4.6	2.3	5.2	86.9	4.3
Cyprus	230	523	484	9.7	0.7	9.8	151.3	9.4
Czech Republic	426	7460	513	6.3	1.8	6.6	82.3	5.3
Denmark	480	1547	517	5.1	1.3	5.3	102.1	7.4
France	670	90387	496	3.9	0.9	4.0	69.4	4.3
Germany	487	59545	513	5.6	2.9	6.3	95.2	4.9
Greece	311	10015	490	7.9	1.8	8.1	122.3	7.0
Latvia (LSS)	374	495	523	17.8	1.6	17.8	110.9	12.2
Norway	781	3221	545	3.3	2.8	4.4	68.2	2.7
Russian Federation	714	17949	555	7.3	1.7	7.5	99.5	5.7
Slovenia	566	8274	538	12.9	2.0	13.1	107.7	7.2
Sweden	651	11056	529	5.7	0.7	5.8	85.6	2.9
Switzerland	727	5662	538	4.2	0.9	4.3	88.5	3.8
United States	1617	270205	480	4.2	0.7	4.2	63.2	4.5

S.D. = standard deviation
S.E. = standard error

Table C.22 Sampling and Imputation Standard Errors - Heat Scale Females in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	244	10647	503	5.9	2.1	6.2	77.2	5.8
Austria	457	18721	420	6.7	0.9	6.8	85.4	4.8
Canada	913	22098	492	7.8	2.2	8.1	80.4	7.1
Cyprus	137	311	461	11.1	1.1	11.2	134.8	11.2
Czech Republic	661	11968	472	4.4	0.7	4.5	74.4	7.3
Denmark	117	394	487	8.2	5.1	9.6	89.9	8.9
France	417	57604	487	4.7	3.3	5.7	65.1	3.5
Germany	222	26328	461	9.7	4.3	10.6	93.1	9.9
Greece	148	4652	460	10.4	1.7	10.5	104.2	8.6
Latvia (LSS)	334	484	484	23.2	2.8	23.4	111.1	8.2
Norway	267	1148	511	6.7	2.0	7.0	69.4	7.5
Russian Federation	519	15026	501	14.7	1.9	14.8	104.5	10.6
Slovenia	160	3162	470	18.4	3.6	18.7	129.9	13.4
Sweden	361	5402	507	4.7	2.6	5.4	71.1	4.6
Switzerland	632	5495	480	5.5	1.4	5.7	80.8	4.1
United States	1497	252579	474	2.6	0.9	2.7	52.7	1.7

S.D. = standard deviation
S.E. = standard error

Table C.23 Sampling and Imputation Standard Errors - Wave Phenomena Scale Males in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	417	20973	529	8.8	1.8	9.0	100.3	9.1
Austria	306	11559	506	7.2	0.9	7.3	80.3	4.5
Canada	1440	28901	497	4.0	1.4	4.3	83.7	3.5
Cyprus	230	523	519	10.4	0.6	10.4	125.6	9.8
Czech Republic	426	7460	491	7.1	1.3	7.2	74.1	3.7
Denmark	480	1547	547	6.1	1.6	6.3	100.9	6.6
France	670	90387	475	5.3	1.7	5.6	73.0	3.1
Germany	487	59545	551	12.4	2.4	12.7	95.8	6.8
Greece	311	10015	457	7.4	0.8	7.4	97.8	6.5
Latvia (LSS)	374	495	515	17.2	1.8	17.3	92.6	12.3
Norway	781	3221	575	4.4	2.2	4.9	85.6	2.9
Russian Federation	714	17949	539	7.6	2.1	7.9	101.0	4.2
Slovenia	566	8274	538	11.8	1.3	11.9	106.5	8.2
Sweden	651	11056	576	6.1	0.8	6.1	111.3	4.2
Switzerland	727	5662	533	4.6	1.3	4.8	85.4	4.8
United States	1617	270205	460	2.4	1.0	2.6	54.7	2.3

S.D. = standard deviation
S.E. = standard error

Table C.24 Sampling and Imputation Standard Errors - Wave Phenomena Scale Females in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	244	10647	498	6.1	3.8	7.2	89.4	5.0
Austria	457	18721	444	9.7	0.7	9.7	82.3	10.9
Canada	913	22098	476	6.2	1.8	6.4	73.5	4.6
Cyprus	137	311	486	8.3	1.1	8.4	104.5	7.9
Czech Republic	661	11968	419	4.8	0.6	4.9	62.9	3.4
Denmark	117	394	493	7.9	6.1	10.0	79.7	6.8
France	417	57604	448	3.7	2.7	4.6	70.8	3.8
Germany	222	26328	485	9.7	2.9	10.1	87.7	6.4
Greece	148	4652	444	6.9	2.0	7.2	82.0	5.8
Latvia (LSS)	334	484	480	16.1	1.5	16.2	86.0	11.7
Norway	267	1148	519	9.6	3.6	10.2	91.0	6.2
Russian Federation	519	15026	487	12.2	2.4	12.4	104.7	9.7
Slovenia	160	3162	446	13.0	3.2	13.4	110.7	10.5
Sweden	361	5402	528	5.1	2.9	5.9	91.3	4.7
Switzerland	632	5495	460	4.3	1.0	4.4	76.6	2.4
United States	1497	252579	442	3.0	0.5	3.0	50.1	1.8

S.D. = standard deviation
S.E. = standard error

Table C.25 Sampling and Imputation Standard Errors
Particle, Quantum, Astrophysics, and Relativity
Males in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	417	20973	533	6.6	1.1	6.7	86.1	6.2
Austria	306	11559	505	9.9	1.2	9.9	82.5	5.5
Canada	1440	28901	513	5.9	0.6	6.0	80.3	4.5
Cyprus	230	523	450	7.7	0.4	7.7	130.3	6.5
Czech Republic	426	7460	498	6.8	1.6	6.9	87.9	5.4
Denmark	480	1547	546	5.9	1.3	6.0	84.8	5.5
France	670	90387	485	4.0	1.6	4.3	61.0	3.6
Germany	487	59545	561	15.1	2.7	15.3	108.3	9.7
Greece	311	10015	456	6.4	0.7	6.4	96.9	7.3
Latvia (LSS)	374	495	505	16.5	1.1	16.6	94.2	10.3
Norway	781	3221	585	4.4	2.3	5.0	80.7	3.2
Russian Federation	714	17949	561	7.8	1.3	7.9	91.9	4.7
Slovenia	566	8274	528	18.6	2.6	18.7	110.5	12.6
Sweden	651	11056	570	3.2	0.5	3.3	76.8	3.1
Switzerland	727	5662	519	5.7	0.9	5.8	87.0	5.9
United States	1617	270205	466	3.4	1.2	3.6	51.4	4.5

S.D. = standard deviation
S.E. = standard error

Table C.26 Sampling and Imputation Standard Errors
Particle, Quantum, Astrophysics, and Relativity
Males in Their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	244	10647	497	7.5	2.4	7.8	87.0	5.4
Austria	457	18721	465	5.9	1.6	6.1	79.1	4.6
Canada	913	22098	471	4.9	1.4	5.1	73.6	4.5
Cyprus	137	311	411	9.9	0.4	9.9	126.7	11.0
Czech Republic	661	11968	425	4.5	1.1	4.6	74.2	3.5
Denmark	117	394	529	6.4	3.6	7.4	73.7	7.4
France	417	57604	457	3.8	1.5	4.1	57.7	6.0
Germany	222	26328	508	13.2	2.8	13.5	98.4	8.1
Greece	148	4652	426	5.5	1.6	5.7	79.4	8.3
Latvia (LSS)	334	484	470	20.6	3.0	20.8	92.9	10.5
Norway	267	1148	549	9.8	1.7	9.9	87.1	10.2
Russian Federation	519	15026	520	13.8	2.1	13.9	100.4	9.5
Slovenia	160	3162	458	13.5	4.1	14.1	102.0	8.9
Sweden	361	5402	538	5.9	2.0	6.2	73.2	4.7
Switzerland	632	5495	457	4.2	1.5	4.4	66.8	2.8
United States	1497	252579	446	2.1	0.9	2.3	44.7	1.8

S.D. = standard deviation
S.E. = standard error

Acknowledgments

TIMSS was truly a collaborative effort among hundreds of individuals around the world. Staff from the national research centers, the international management, advisors, and funding agencies worked closely to design and implement the most ambitious study of international comparative achievement ever undertaken. TIMSS would not have been possible without the tireless efforts of all involved. Below, the individuals and organizations are acknowledged for their contributions. Given that implementing TIMSS has spanned more than seven years and involved so many people and organizations, this list may not pay heed to all who contributed throughout the life of the project. Any omission is inadvertent. TIMSS also acknowledges the students, teachers, and school principals who contributed their time and effort to the study.

MANAGEMENT AND OPERATIONS

Since 1993, TIMSS has been directed by the International Study Center at Boston College in the United States. Prior to this, the study was coordinated by the International Coordinating Center at the University of British Columbia in Canada. Although the study was directed centrally by the International Study Center and its staff members implemented various parts of TIMSS, important activities also were carried out in centers around the world. The data were processed centrally by the IEA Data Processing Center in Hamburg, Germany. Statistics Canada was responsible for collecting and evaluating the sampling documentation from each country and for calculating the sampling weights. The Australian Council for Educational Research conducted the scaling of the achievement data.

International Study Center (1993-)

Albert E. Beaton, International Study Director
Michael O. Martin, Deputy International Study Director
Ina V.S. Mullis, Co-Deputy International Study Director
Eugenio J. Gonzalez, Director of Operations and Data Analysis
Dana L. Kelly, Research Associate
Teresa A. Smith, Senior Research Associate
Kathleen O'Connor, Research Associate
Cheryl L. Flaherty, Research Associate (former)
Rachel L. Saks, Research Associate
Maryellen Harmon, Performance Assessment Coordinator
Kelvin Gregory, Research Associate
Robert Jin, Computer Programmer
Ce Shen, Computer Programmer
William J. Crowley, Fiscal Administrator
Thomas M. Hoffmann, Publications Coordinator
José Rafael Nieto, Data Graphics Coordinator

International Study Center (Continued)

Ann G.A. Tan, Conference Coordinator
Mary C. Howard, Office Supervisor
Diane Joyce, Secretary
Joanne E. McCourt, Secretary
Sarah Andrews, Secretary
Kathleen A. Haley, Graduate Assistant (former)
Craig D. Hoyle, Graduate Assistant

International Coordinating Center (1991-93)

David F. Robitaille, International Coordinator
Robert A. Garden, Deputy International Coordinator
Barry Anderson, Director of Operations
Beverley Maxwell, Director of Data Management

Statistics Canada

Pierre Foy, Senior Methodologist
Suzelle Giroux, Senior Methodologist
Jean Dumais, Senior Methodologist
Nancy Darcovich, Senior Methodologist
Marc Joncas, Senior Methodologist
Laurie Reedman, Junior Methodologist
Claudio Perez, Junior Methodologist

IEA Data Processing Center

Jens Brockmann, Research Assistant
Michael Bruneforth, Senior Researcher (former)
Jedidiah Harris, Research Assistant
Dirk Hastedt, Senior Researcher
Svenja Moeller, Research Assistant
Knut Schwippert, Senior Researcher
Heiko Sibberns, Senior Researcher
Jockel Wolff, Research Assistant

Australian Council for Educational Research

Raymond J. Adams, Principal Research Fellow
Margaret Wu, Research Fellow
Nikolai Volodin, Research Fellow
David Roberts, Research Officer
Greg Macaskill, Research Officer

IEA Secretariat

Tjeerd Plomp, Chairperson
Hans Wagemaker, Executive Director
Barbara Malak-Minkiewicz, Manager Membership Relations
Leendert Dijkhuizen, Financial Officer
Karin Baddane, Secretary

FUNDING AGENCIES

Funding for the International Study Center was provided by the National Center for Education Statistics of the U.S. Department of Education, the U.S. National Science Foundation, and the International Association for the Evaluation for Educational Achievement. Eugene Owen and Lois Peak of the National Center for Education Statistics and Larry Suter of the National Science Foundation each played a crucial role in making TIMSS possible and for ensuring the quality of the study. Funding for the International Coordinating Center was provided by the Applied Research Branch of the Strategic Policy Group of the Canadian Ministry of Human Resources Development. This initial source of funding was vital in initiating the TIMSS project. Tjeerd Plomp, Chair of the IEA and of the TIMSS Steering Committee, has been a constant source of support throughout TIMSS. It should be noted that each country provided its own funding for the implementation of the study at the national level.

NATIONAL RESEARCH COORDINATORS

The TIMSS National Research Coordinators and their staff had the enormous task of implementing the TIMSS design in their countries. This required obtaining funding for the project; participating in the development of the instruments and procedures; conducting field tests; participating in and conducting training sessions; translating the instruments and procedural manuals into the local language; selecting the sample of schools and students; working with the schools to arrange for the testing; arranging for data collection, coding, and data entry; preparing the data files for submission to the IEA Data Processing Center; contributing to the development of the international reports; and preparing national reports. The way in which the national centers operated and the resources that were available varied considerably across the TIMSS countries. In some countries, the tasks were conducted centrally, while in others, various components were subcontracted to other organizations. In some countries, resources were more than adequate, while in others, the national centers were operating with limited resources. Of course, across the life of the project, some NRCs have changed. This list attempts to include all past NRCs who served for a significant period of time as well as all the present NRCs. All of the TIMSS National Research Coordinators and their staff members are to be commended for their professionalism and their dedication in conducting all aspects of TIMSS.

NATIONAL RESEARCH COORDINATORS

Argentina

Carlos Mansilla
Universidad del Chaco
Av. Italia 350
3500 Resistencia
Chaco, Argentina

Australia

Jan Lokan
Raymond Adams *
Australian Council for Educational Research
19 Prospect Hill
Private Bag 55
Camberwell, Victoria 3124
Australia

Austria

Guenter Haider
Austrian IEA Research Centre
Universität Salzburg
Akademiestraße 26/2
A-5020 Salzburg, Austria

Belgium (Flemish)

Christiane Brusselmans-Dehairs
Rijksuniversiteit Ghent
Vakgroep Onderwijskunde &
The Ministry of Education
Henri Dunantlaan 2
B-9000 Ghent, Belgium

Belgium (French)

Georges Henry
Christian Monseur
Universite de Liège
B32 Sart-Tilman
4000 Liège 1, Belgium

Bulgaria

Kiril Bankov
Foundation for Research, Communication,
Education and Informatics
Tzarigradsko Shausse 125, Bl. 5
1113 Sofia, Bulgaria

Canada

Alan Taylor
Applied Research & Evaluation Services
University of British Columbia
2125 Main Mall
Vancouver, B.C. V6T 1Z4
Canada

Colombia

Carlos Jairo Diaz
Universidad del Valle
Facultad de Ciencias
Multitaller de Materiales Didacticos
Ciudad Universitaria Meléndez
Apartado Aereo 25360
Cali, Colombia

Cyprus

Constantinos Papanastasiou
Department of Education
University of Cyprus
Kallipoleos 75
P.O. Box 537
Nicosia 133, Cyprus

Czech Republic

Jana Strakova
Vladislav Tomasek
Institute for Information on Education
Senovazne Nam. 26
111 21 Praha 1, Czech Republic

*Past National Research Coordinator.

Denmark

Peter Weng
Peter Allerup
Borge Prien*
The Danish National Institute for Educational Research
28 Hermodsgade
Dk-2200 Copenhagen N, Denmark

England

Wendy Keys
Derek Foxman*
National Foundation for Educational Research
The Mere, Upton Park
Slough, Berkshire SL1 2DQ
England

France

Anne Servant
Ministère de l'Education Nationale
142, rue du Bac
75007 Paris, France

Josette Le Coq*
Centre International d'Etudes Pédagogiques (CIEP)
1 Avenue Léon Journault
93211 Sèvres, France

Germany

Rainer Lehmann
Humboldt-Universitaet zu Berlin
Institut Fuer Allgemeine Erziehungswissenschaft
Geschwister-Scholl-Str. 6
10099 Berlin, Germany

Juergen Baumert
Wilfried Bos
Rainer Waterman
Max-Planck Institute for Human Development and Education
Lentzeallee 94
14191 Berlin, Germany

Manfred Lehrke
Universität Kiel
IPN Olshausen Str. 62
24098 Kiel, Germany

Greece

Georgia Kontogiannopoulou-Polydorides
Department of Education (Nipiagogon)
University of Athens
Navarinou 13A, Neochimio
Athens 10680, Greece

Joseph Solomon
Department of Education
University of Patras
Patras 26500, Greece

Hong Kong

Frederick Leung
Nancy Law
The University of Hong Kong
Department of Curriculum Studies
Pokfulam Road, Hong Kong

Hungary

Péter Vari
National Institute of Public Education
Centre for Evaluation Studies
Dorottya U. 8, P.O. Box 120
1051 Budapest, Hungary

Iceland

Einar Gudmundsson
Institute for Educational Research
Department of Educational Testing and Measurement
Surdgata 39
101 Reykjavik, Iceland

Indonesia

Jahja Umar
Ministry of Education and Culture
Examination Development Center
Jalan Gunung Sahari - 4
Jakarta 10000, Indonesia

Ireland

Deirdre Stuart
Michael Martin*
Educational Research Centre
St. Patrick's College
Drumcondra
Dublin 9, Ireland

*Past National Research Coordinator.

Iran, Islamic Republic

Ali Reza Kiamanesh
Ministry of Education
Center for Educational Research
Iranshahr Shomali Avenue
Teheran 15875, Iran

Israel

Pinchas Tamir
The Hebrew University
Israel Science Teaching Center
Jerusalem 91904, Israel

Ruth Zuzovsky
Tel Aviv University
School of Education
Ramat Aviv
PO Box 39040
Tel Aviv 69978, Israel

Italy

Anna Maria Caputo
Ministero della Pubblica Istruzione
Centro Europeo dell'Educazione
Villa Falconieri
00044 Frascati, Italy

Japan

Masao Miyake
Eizo Nagasaki
National Institute for Educational Research
6-5-22 Shimomeguro
Meguro-Ku, Tokyo 153, Japan

Korea

Jingyu Kim
Hyung Im*
National Board of Educational Evaluation
Evaluation Research Division
Chungdam-2 Dong 15-1, Kangnam-Ku
Seoul 135-102, Korea

Kuwait

Mansour Hussein
Ministry of Education
P. O. Box 7
Safat 13001, Kuwait

Latvia

Andrejs Geske
University of Latvia
Faculty of Education & Psychology
Jurmalas Gatve 74/76, Rm. 204a
Riga, Lv-1083, Latvia

Lithuania

Algirdas Zabulionis
University of Vilnius
Faculty of Mathematics
Naugarduko 24
2006 Vilnius, Lithuania

Mexico

Fernando Córdova Calderón
Director de Evaluación de Politicas y
Sistemas Educativos
Netzahualcoyotl #127 2ndo Piso
Colonia Centro
Mexico 1, D.F., Mexico

Netherlands

Wilmad Kuiper
Klaas Bos
Anja Knuver
University of Twente
Faculty of Educational Science
and Technology
Department of Curriculum
P.O. Box 217
7500 AE Enschede, Netherlands

New Zealand

Megan Chamberlain
Steve May
Hans Wagemaker*
Ministry of Education
Research and International Section
P.O. Box 1666
45-47 Pipitea Street
Wellington, New Zealand

*Past National Research Coordinator.

Norway

Svein Lie
University of Oslo
SLS Postboks 1099
Blindern 0316
Oslo 3, Norway

Gard Brekke
Alf Andersensv 13
3670 Notodden, Norway

Philippines

Milagros Ibe
University of the Philippines
Institute for Science and Mathematics
Education Development
Diliman, Quezon City
Philippines

Ester Ogena
Science Education Institute
Department of Science and Technology
Bicutan, Taquig
Metro Manila 1604, Philippines

Portugal

Gertrudes Amaro
Ministerio da Educacao
Instituto de Inovação Educacional
Rua Artilharia Um 105
1070 Lisboa, Portugal

Romania

Gabriela Noveanu
Institute for Educational Sciences
Evaluation and Forecasting Division
Str. Stirbei Voda 37
70732-Bucharest, Romania

Russian Federation

Galina Kovalyova
The Russian Academy of Education
Institute of General Secondary School
Ul. Pogodinskaya 8
Moscow 119905, Russian Federation

Scotland

Brian Semple
Scottish Office, Education &
Industry Department
Victoria Quay
Edinburgh, E86 6QQ
Scotland

Singapore

Wong Cheow Cher
Chan Siew Eng*
Research and Evaluation Branch
Block A Belvedere Building
Ministry of Education
Kay Siang Road
Singapore 248922

Slovak Republic

Maria Berova
Vladimir Burjan*
SPU-National Institute for Education
Pluhova 8
P.O. Box 26
830 00 Bratislava
Slovak Republic

Slovenia

Marjan Setinc
Barbara Japelj
Pedagoski Institut Pri Univerzi v Ljubljana
Gerbiceva 62, P.O. Box 76
61111 Ljubljana, Slovenia

South Africa

Sarah Howie
Derek Gray*
Human Sciences Research Council
134 Pretorius Street
Private Bag X41
Pretoria 0001, South Africa

Spain

José Antonio Lopez Varona
Instituto Nacional de Calidad y Evaluación
C/San Fernando del Jarama No. 14
28071 Madrid, Spain

*Past National Research Coordinator.

Sweden

Ingemar Wedman
Anna Hofslagare
Kjell Gisselberg*
Umeå University
Department of Educational Measurement
S-901 87 Umeå, Sweden

Switzerland

Erich Ramseier
Amt Für Bildungsforschung der Erziehungs-
direktion des Kantons Bern
Sulgeneck Straße 70
Ch-3005 Bern, Switzerland

Thailand

Suwaporn Semheng
Institute for the Promotion of Teaching Science
and Technology
924 Sukhumvit Road
Bangkok 10110, Thailand

United States

William Schmidt
Michigan State University
Department of Educational Psychology
463 Erikson Hall
East Lansing, MI 48824-1034
United States

*Past National Research Coordinator.

TIMSS ADVISORY COMMITTEES

The TIMSS International Study Center was supported in its work by several advisory committees. The TIMSS International Steering Committee provided guidance to the International Study Director on policy issues and general direction of the study. The TIMSS Technical Advisory Committee provided guidance on issues related to design, sampling, instrument construction, analysis, and reporting, ensuring that the TIMSS methodologies and procedures were technically sound. The Subject Matter Advisory Committee ensured that current thinking in mathematics and science education were addressed by TIMSS, and was instrumental in the development of the TIMSS tests. The Free-Response Item Coding Committee developed the coding rubrics for the free-response items. The Performance Assessment Committee worked with the Performance Assessment Coordinator to develop the TIMSS performance assessment. The Quality Assurance Committee helped to develop the quality assurance program.

International Steering Committee

Tjeerd Plomp (Chair), the Netherlands
Lars Ingelstam, Sweden
Daniel Levine, United States
Senta Raizen, United States
David Robitaille, Canada
Toshio Sawada, Japan
William Schmidt, United States
Benny Suprapto Brotosiswojo, Indonesia

Technical Advisory Committee

Raymond Adams, Australia
Pierre Foy, Canada
Andreas Schleicher, Germany
William Schmidt, United States
Trevor Williams, United States

Sampling Referee

Keith Rust, United States

Subject Area Coordinators

Robert Garden, New Zealand (Mathematics)
Graham Orpwood, Canada (Science)

Special Mathematics Consultant

Chancey Jones

Subject Matter Advisory Committee

Svein Lie (Chair), Norway
Antoine Bodin, France
Peter Fensham, Australia
Robert Garden, New Zealand
Geoffrey Howson, England
Curtis McKnight, United States
Graham Orpwood, Canada
Senta Raizen, United States
David Robitaille, Canada
Pinchas Tamir, Israel
Alan Taylor, Canada
Ken Travers, United States
Theo Wubbels, the Netherlands

Free-Response Item Coding Committee

Svein Lie (Chair), Norway
Vladimir Burjan, Slovak Republic
Kjell Gisselberg, Sweden
Galina Kovalyova, Russian Federation
Nancy Law, Hong Kong
Josette Le Coq, France
Jan Lokan, Australia
Curtis McKnight, United States
Graham Orpwood, Canada
Senta Raizen, United States
Alan Taylor, Canada
Peter Weng, Denmark
Algirdas Zabulionis, Lithuania

Performance Assessment Committee

Derek Foxman, England
Robert Garden, New Zealand
Per Morten Kind, Norway
Svein Lie, Norway
Jan Lokan, Australia
Graham Orpwood, Canada

Quality Control Committee

Jules Goodison, United States
Hans Pelgrum, The Netherlands
Ken Ross, Australia

Editorial Committee

David F. Robitaille (Chair), Canada
Albert Beaton, International Study Director
Paul Black, England
Svein Lie, Norway
Rev. Ben Nebres, Philippines
Judith Torney-Purta, United States
Ken Travers, United States
Theo Wubbels, the Netherlands