PROGRESS IN INTERNATIONAL READING LITERACY STUDY

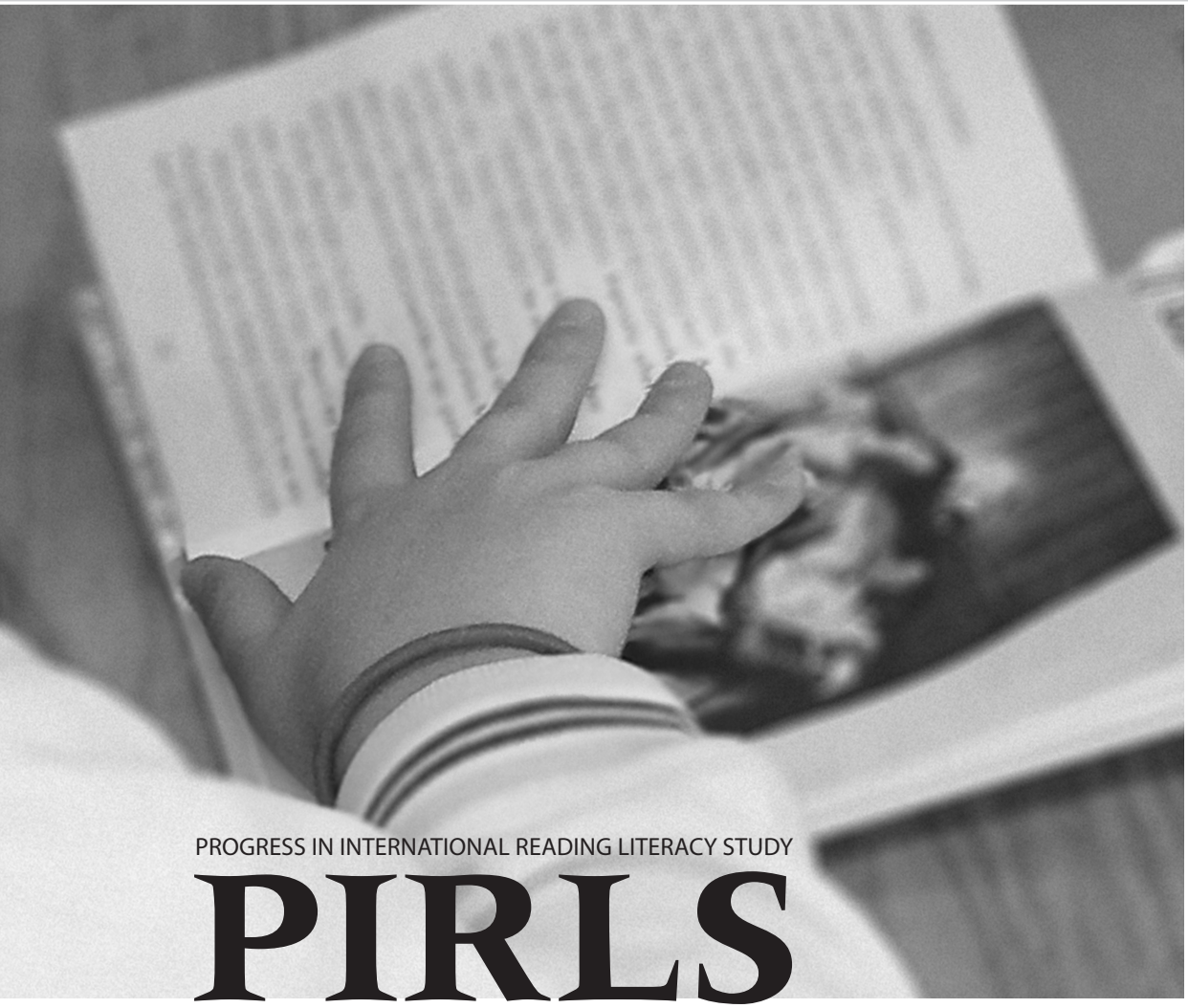
PIRLS

Edited by
Pierre Foy
Ann M. Kennedy

PIRLS 2006 User Guide

for the International Database

TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College



PROGRESS IN INTERNATIONAL READING LITERACY STUDY

PIRLS

PIRLS 2006 User Guide
for the International Database

Contributors

Michael O. Martin
Ina V.S. Mullis
Marc Joncas
Juliane Barth
Oliver Neuschmidt
Olaf Zuehlke
Milena Taneva
Kathleen L. Trong
Eugenio J. Gonzalez
Joseph Galia

Edited by

Pierre Foy
Ann M. Kennedy

TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

PIRLS 2006 User Guide for the International Database /
Edited by Pierre Foy and Ann M. Kennedy

Publisher: TIMSS & PIRLS International Study Center,
Lynch School of Education, Boston College

Library of Congress
Catalog Card Number: 1-889938-45-9

ISBN: 2007925070

For more information about PIRLS contact:

TIMSS & PIRLS International Study Center
Lynch School of Education
Boston College
Chestnut Hill, MA 02467
United States

tel: +1-617-552-1600
fax: +1-617-552-1203

email: pirls@bc.edu

http://pirls.bc.edu

Printed and bound in the United States.

Table of Contents

Chapter 1

Overview of the PIRLS 2006 International Database

1.1 Introduction

To support and promote secondary analyses aimed at improving reading education in the primary grades, the PIRLS 2006 international database makes available to researchers, analysts, and other users the data collected and processed by IEA's PIRLS 2006 project. This database comprises student reading achievement data as well as student, parent, teacher, school, and curricular background data for 40 countries, including Belgium with two educational systems and Canada with five provinces (45 participants in total). The database includes data from over 210,000 students and their parents, about 6,750 teachers and school principals, and the National Research Coordinators of each country. All participating countries gave the IEA permission to release their national data.

As the recognized pioneer of international assessments, IEA has been conducting comparative studies of students' academic achievement for almost 50 years. IEA's Progress in International Reading Literacy Study (PIRLS) provides internationally comparative data about students' reading achievement in primary school (the fourth grade in most participating countries). The fourth grade is an important transition point in children's development as readers, because most of them should have learned to read, and are now reading to learn. PIRLS has roots in earlier IEA studies, including the reading component of IEA's six-subject study in 1973 (Thorndike, 1973; Walker, 1976) and IEA's Reading Literacy Study conducted in 1991 in 32 countries (Elley, 1992, 1994) and again in 2001 in nine countries to provide trends (Martin, Mullis, Gonzalez, & Kennedy, 2003).

PIRLS 2006 was an ambitious and demanding study, involving complex procedures for assessing students' reading achievement, drawing student samples, and analyzing and reporting the data. In order to work effectively with the PIRLS

data it is necessary to have an understanding of the characteristics of the study, which are described fully in the *PIRLS 2006 Technical Report* (Martin, Mullis, & Kennedy, 2007). It is intended, therefore, that this User Guide be used in conjunction with the technical report. Whereas the User Guide describes the organization and content of the database, the technical report provides the rationale for the techniques used and for the variables created.

1.2 Overview of the User Guide Contents

This User Guide describes the content and format of the data in the PIRLS 2006 international database. Chapter 2 details aspects of the study that determine the structure and content of the database, and describes the database variables that were constructed to analyze and report the results of the study. Chapter 3 introduces the IEA's International Database (IDB) Analyzer software (IEA, 2007) and presents examples of analyses of the PIRLS 2006 data. Chapters 4 and 5 present similar examples using the SPSS (2005) and SAS (2002) statistical software systems, respectively, with corresponding analysis programs provided on DVD.

The User Guide is accompanied by four supplementary volumes: Supplement 1 comprises the international version of each of the PIRLS 2006 background questionnaires, Supplement 2 describes any national adaptations to the questions in each questionnaire, Supplement 3 describes in detail how indices and other derived variables were constructed for reporting the PIRLS data, and Supplement 4 describes the sampling stratification variables for each country.

This User Guide also includes a DVD that contains the PIRLS 2006 international database with all student reading achievement and background questionnaires data files. The DVD also includes support materials. The following is a list of the DVD folder names and a description of their contents:

DATA:	All data files in SAS and SPSS formats
PROGRAMS:	All SAS and SPSS programs and macros
CODEBOOKS:	Codebook files with descriptions of all variables in the PIRLS data files
ALMANACS:	Data almanacs containing summary statistics for all items and background variables
ITEMS:	Item information file and PDF versions of the released reading passages, their corresponding items, and scoring guides for constructed-response items

CURRICULUM:	Curriculum Questionnaire data file
REPORTS:	PDF versions of the *PIRLS 2006 Assessment Framework and Specifications, PIRLS 2006 International Report*, *PIRLS 2006 Technical Report*, *PIRLS 2006 Encyclopedia*, and this User Guide with its Supplements
IDB ANALYZER:	Executable file for installing the IEA's IDB Analyzer

The file names within the DVD generally follow the DOS file naming convention: file names with up to eight characters, followed by a three-character extension (as in FILENAME.EXT). Files with the same names are complementary to each other, and the extension identifies their function or type.

The extensions used for the files contained in the DVD are the following:

.SAS	SAS programs and macros
.SPS	SPSS programs and macros
.EXP	SAS Export files
.SAV	SPSS System files
.DOC	Almanacs in MSWord format
.PDF	Almanacs, codebooks and reports in PDF format
.SDB	Codebooks in standard Dbase format (readable in Excel)
.XLS	Curriculum Questionnaire data file and item information file in Excel format
.EXE	IDB Analyzer Setup executable file

1.3 The PIRLS 2006 Assessment Frameworks and Specifications

The *PIRLS 2006 Assessment Framework and Specifications* (Mullis, Kennedy, Martin, & Sainsbury, 2006) contains a detailed description of the PIRLS 2006 assessment of reading comprehension. Based on the *PIRLS 2001 Framework and Specifications* (Campbell, Kelly, Mullis, Martin, & Sainsbury, 2001), the framework defines the two major aspects of students' reading literacy—purposes for reading and processes of comprehension. Reading for literary experience and reading to acquire and use information are the two major purposes that account for the majority of reading experiences of young children. Readers make meaning of texts in a variety of ways, depending not only on the purpose for reading, but also on the difficulty of the text and the reader's prior knowledge. PIRLS looks at

four processes of comprehension: focus on and retrieve explicitly stated information; make straightforward inferences; interpret and integrate ideas and information; and examine and evaluate content, language, and textual elements. These processes are the basis for developing comprehension questions in the reading assessment.

A third major aspect of students' reading literacy is reading behaviors and attitudes. The framework defines the relationships among the home, school, and national and community environments that help to shape the development of reading literacy among young children. Because measuring trends in students' reading literacy is an important focus of PIRLS, the PIRLS 2006 contextual framework was similar to that used in 2001.

The PIRLS 2006 Achievement Booklets

The PIRLS 2006 assessment design, also elaborated in the *PIRLS 2006 Framework and Specifications*, builds on PIRLS 2001, in which there were four literary and four informational test blocks. The decision to report reading achievement scale scores by process as well as by purpose, in combination with the desire to include a range of texts within each reading purpose, made it necessary to increase PIRLS 2006 to include five literary and five informational test blocks. Each of the 10 test blocks included a reading passage and its accompanying questions.

So as not to overburden the young children participating in PIRLS, and in line with the practice in 2001, the testing time was limited to 80 minutes (two passages) per student, with an additional 15-30 minutes allotted for a student questionnaire. With 10 reading passages in total, but just 2 to be given to any one student, passages and their accompanying items were assigned to student test booklets according to a matrix sampling plan. The 10 passages were distributed across 13 booklets, two per booklet, so that passages were paired together in a booklet in as many different ways as possible. Each student booklet consisted of two 40-minute blocks of passages and items, which were accompanied by the Student Questionnaire. So as to present at least some passages in a more natural, authentic setting, two blocks (one literary and one informational) were presented in colorized, magazine-type format, with the questions appearing in a separate booklet. This booklet, Booklet 13, is referred to as the PIRLS "Reader."[1]

[1] The PIRLS 2006 test booklet design is described in Mullis, I.V.S., Kennedy, A.M., Martin, M.O., & Sainsbury, M. (2006). *PIRLS 2006 assessment framework and specifications*, 2nd edition. Chestnut Hill, MA: Boston College.

Student Achievement Scores

Because the test booklet completed by each student contained only a subset of the items in the whole assessment item pool, each student essentially responded to just a part of the assessment, which posed a challenge in terms of determining individual student achievement scores. As described in Foy, Galia, and Li (2007), PIRLS used a sophisticated psychometric scaling technique (known as item response theory scaling with conditioning and multiple imputation) to derive estimates for each student of the scores they would have received had they completed the entire assessment. These imputed student achievement scores were then used in analyzing and reporting the data.

Each student record in the PIRLS 2006 international database contains imputed scores in reading overall, as well as for the two reading purposes subscales and two reading processes subscales. Because each imputed score is a prediction based on limited information, it almost certainly includes some small amount of error. To allow analysts to judge the effect of the imputation on their analyses, the PIRLS database provides five separate imputed estimates (known as plausible values) for each score. Accordingly, an analysis may be replicated as many as five times, using a different score each time, and the results compared to judge the impact of the imputation process. The database also includes software that enables analysts using the SAS or SPSS software systems to combine the results of the five replications into a single result and to compute standard errors that incorporate both sampling and imputation error.

The PIRLS reading achievement scales were designed to provide reliable measures of student achievement on the metric of the scale established originally with the 2001 assessment. Treating equally all the countries that participated in 2001, the PIRLS scale average over those countries was set at 500 and the standard deviation at 100. Subsequently, the results from 2001 and 2006 were linked so that the scores from the two assessments were on the same scale and may be compared for analytic purposes.

PIRLS 2006 Background Questionnaires

In order to measure trends and collect baseline information about key factors related to students' home and school environments, PIRLS 2006 administered questionnaires to students, parents, teachers, and school principals. Additionally, PIRLS 2006 included a curriculum questionnaire that provided information about the reading curriculum. Based on the contexts for learning to read, as defined in

the PIRLS framework, the information from the five questionnaires complements the fourth-grade students' reading achievement results.

Student Questionnaire

Each student in the selected class completed a *Student Questionnaire*. The questionnaire included questions about home resources, languages spoken in the home, students' reading habits both inside and outside of school, students' reading self-concept and their attitudes towards reading, classroom instructional practices related to teaching reading, and school safety.

Learning to Read Survey (Home)

The parents or guardians of each student completed a *Learning to Read Survey*. The questionnaire asked about preparations for primary schooling, including attendance in preschool and literacy-centered activities in the home before the child began school, such as reading books, singing songs, or writing letters or words. Parents answered questions about home resources in addition to information about their highest level of education and employment situations.

Teacher Questionnaire

Teachers of the assessed classes responded to the *Teacher Questionnaire*. The questionnaire focused on reading activities and materials used for reading instruction and the assessment of students' performance in reading. Teachers were asked to refer specifically to the class of students selected for the PIRLS assessment. Teachers also answered questions about their professional preparation and experience in teaching reading.

School Questionnaire

The principal of each school sampled for PIRLS completed a *School Questionnaire*. Principals answered questions about the emphasis on the reading curriculum in the school, the availability and use of materials to teach reading, and whether the school provided programs and services that involve the students and their families. Additionally, the questionnaire asked school principals general questions about their school's demographic characteristics, resources, and environment.

Curriculum Questionnaire

The National Research Coordinator within each country was responsible for completing the *Curriculum Questionnaire*. Questions primarily centered on the defined national or regional curriculum in fourth grade, including what it prescribed and how it is disseminated. NRCs also answered questions about requirements for teachers and how teachers are informed about the reading curriculum. An addendum to the curriculum questionnaire asked about country-level policies regarding entry to primary school as they related to the student tested in PIRLS 2006.

Development of the PIRLS 2006 background questionnaires is described in the *PIRLS 2006 Technical Report* (see Kennedy, 2007). The international version of each of the PIRLS 2006 background questionnaires is provided in Supplement 1 of this User Guide, along with a summary table of questionnaire items that indicates whether an item also was asked in PIRLS 2001. Documentation of the national adaptations to each questionnaire is provided in Supplement 2. The indices and variables other than indices derived from the student, home, teacher, and school questionnaires, are listed with detailed descriptions and analysis notes in Supplement 3.

PIRLS 2006 Student Populations

As explained in the PIRLS 2006 framework, PIRLS 2006 had as its target population students enrolled in the fourth grade of formal schooling, counting from the first year of primary school defined by UNESCO's International Standard Classification for Education (UNESCO, 2006). Accordingly, the fourth year of formal schooling should be the fourth grade in most countries. To avoid testing very young children, however, PIRLS has a policy that the average age of children in the grade tested should not be below 9.5 years old.

In each country, representative samples of students were selected using a two-stage sampling design. Although countries could, with prior approval, adapt the sampling design to local circumstances, in general countries selected at least 150 schools at the first stage using probability-proportional-to-size sampling. Countries could incorporate in their sampling design important reporting variables (for example, urbanicity or school type) as stratification variables. At the second stage, one or two classes were randomly sampled in each school. Generally, this resulted in a sample size of at least 4,000 students per country. Some countries opted to include more schools and classes, enabling additional

analyses, which resulted in larger sample sizes (see Joncas, 2007, for a detailed description of the sample design and implementation).

Exhibit 1.1 lists all the countries that have participated in PIRLS in 2001 and 2006. Forty countries, including Belgium with two educational systems and Canada with five Canadian provinces, participated in the 2006 PIRLS assessment. Of these, 26 countries and two provinces had trend data from PIRLS 2001.[2] Participating in PIRLS for the first time in 2006 were 13 countries and three provinces.

[2] Although Kuwait participated in PIRLS 2001, the data were not considered suitable for measuring trends, and so Kuwait does not appear in any trend exhibits.

Exhibit 1.1 Participation in PIRLS

Country	ISO Code	Numeric Code	2006	2001
Argentina	ARG	032		•
Austria	AUT	040	•	
Belgium (Flemish)	BFL	956	•	
Belgium (French)	BFR	957	•	
Belize	BLZ	084		•
Bulgaria	BGR	100	•	•
Canada, Alberta	CAB	9134	•	
Canada, British Columbia	CBC	9135	•	
Canada, Nova Scotia	CNS	9136	•	
Canada, Ontario	COT	9132	•	•
Canada, Quebec	CQU	9133	•	•
Chinese Taipei	TWN	158	•	
Colombia	COL	170		•
Cyprus	CYP	196		•
Czech Republic	CZE	203		•
Denmark	DNK	208	•	
England	ENG	926	•	•
France	FRA	250	•	•
Georgia	FEO	268	•	
Germany	DEU	276	•	•
Greece	GRC	300		•
Hong Kong SAR	HKG	344	•	•
Hungary	HUN	348	•	•
Iceland	ISL	352	•	•

Exhibit 1.1 Participation in PIRLS (continued)

Country	ISO Code	Numeric Code	2006	2001
Iceland (5th grade)*	IS5	9352	•	
Indonesia	IDN	360	•	
Iran, Islamic Rep. of	IRN	364	•	•
Israel	ISR	376	•	•
Italy	ITA	380	•	•
Kuwait	KWT	414	•	•
Latvia	LVA	428	•	•
Lithuania	LTU	440	•	•
Luxembourg	LUX	442	•	
Macedonia, Rep. of	MKD	807	•	•
Moldova, Rep. of	MDA	498	•	•
Morocco	MAR	504	•	•
Netherlands	NLD	528	•	•
New Zealand	NZL	554	•	•
Norway	NOR	578	•	•
Norway (5th grade)*	NO5	9578	•	
Poland	POL	616	•	
Qatar	QAT	634	•	
Romania	ROM	642	•	•
Russian Federation	RUS	643	•	•
Scotland	SCO	927	•	•
Singapore	SGP	702	•	•
Slovak Republic	SVK	703	•	•
Slovenia	SVN	705	•	•
South Africa	ZAF	710	•	
Spain	ESP	724	•	
Sweden	SWE	752	•	•
Trinidad and Tobago	TTO	780	•	
Turkey	TUR	792		•
United States	USA	840	•	

* For their own purposes as an additional effort, Iceland and Norway administered PIRLS 2006 to small samples of their fifth grade students. These data are included in the international database as separate files.

Assessment Dates

PIRLS 2006 was administered near the end of the school year in each country. In countries in the Southern Hemisphere (where the school year typically ends in November or December) the assessment was conducted in October or November 2005. In the Northern Hemisphere, the school year typically ends in June; so in most of these countries the assessment was conducted between March and June 2006.

1.4 Study Management and Organization

PIRLS is a major undertaking of IEA, and together with the Trends in International Mathematics and Science Study (TIMSS), comprises the core of IEA's regular cycles of studies. The PIRLS assessment at the fourth grade complements TIMSS, which regularly assesses mathematics and science achievement at fourth and eighth grades.

The TIMSS & PIRLS International Study Center at Boston College has responsibility for the overall direction and management of the TIMSS and PIRLS projects. Headed by Drs. Ina V.S. Mullis and Michael O. Martin, the study center is located in the Lynch School of Education. Dr. Ann M. Kennedy is the PIRLS Project Coordinator. In carrying out the project, the TIMSS & PIRLS International Study Center worked closely with the IEA Secretariat in Amsterdam, which provided guidance overall and was responsible for verification of all translations produced by the participating countries. The IEA Data Processing and Research Center in Hamburg was responsible for processing and verifying the data submitted by the participants; Statistics Canada in Ottawa was responsible for school and student sampling activities; and Educational Testing Service (ETS) in Princeton, New Jersey consulted on psychometric methodology and provided software for scaling the achievement data.

Dr. Marian Sainsbury from the National Foundation for Educational Research in England (NFER) was the PIRLS 2006 Reading Coordinator and Dr. Patricia Donahue from ETS was a special reading assessment consultant. Together with the Reading Development Group, a panel of internationally recognized experts in reading research, instruction, and assessment, they provided excellent guidance throughout PIRLS 2006.

To work with the international team and coordinate within-country activities, each participating country designated an individual to be the PIRLS National Research Coordinator (NRC). The NRCs have the complicated and challenging

task of implementing the PIRLS study in their countries in accordance with the PIRLS guidelines and procedures. The quality of the PIRLS assessment and data depends on the work of the NRCs and their colleagues in carrying out the very complex sampling, data collection, and scoring tasks involved. In addition, the Questionnaire Development Group, comprised of NRCs, provided advice on questionnaire development.

Continuing the tradition of truly exemplary work established in PIRLS 2001, the PIRLS 2006 NRCs (often the same NRCs as in 2001), performed their many tasks with dedication, competence, energy, and goodwill, and have been commended by the IEA Secretariat, the TIMSS & PIRLS International Study Center, the IEA Data Processing and Research Center, and Statistics Canada for their commitment to the project and the high quality of their work.

References

Campbell, J.R., Kelly, D.L., Mullis, I.V.S., Martin, M.O., & Sainsbury, M. (2001). *Framework and specifications for PIRLS assessment 2001* (2nd ed.). Chestnut Hill, MA: Boston College.

Elley, W.B. (1992). *How in the world do students read?* The Hague, Netherlands: IEA.

Elley, W.B. (Ed.). (1994). *The IEA study of reading literacy: Achievement and instruction in thirty-two school systems*. Oxford, England: Elsevier Science Ltd.

Foy, P., Galia, J., and Li, I. (2007). Scaling methods and procedures for the PIRLS 2006 reading achievement scales. In M.O. Martin, I.V.S. Mullis, and A.M. Kennedy (Eds.), *PIRLS 2006 technical report* (pp. 149-172). Chestnut Hill, MA: Boston College

IEA. (2007). *International database analyzer* (version 1.4.0.5). Hamburg, Germany: IEA Data Processing and Research Center.

Joncas, M. (2007). PIRLS 2006 Sampling weights and participation rates. In M.O. Martin, I.V.S. Mullis, and A.M. Kennedy (Eds.), *PIRLS 2006 technical report* (pp. 105-130). Chestnut Hill, MA: Boston College.

Kennedy, A.M., Mullis, I.V.S, Martin, M.O., & Trong, K.L. (2007). *PIRLS 2006 encyclopedia: A guide to reading education in the forty PIRLS 2006 countries*. Chestnut Hill, MA: Boston College.

Kennedy, A.M. (2007). Developing the PIRLS 2006 background questionnaires. In M.O. Martin, I.V.S. Mullis and A.M. Kennedy (Eds.), *PIRLS 2006 technical report* (pp. 23-34). Chestnut Hill, MA: Boston College.

Martin, M.O., Mullis, I.V.S., Gonzalez, E.J.,& Kennedy, A.M.(2003). *Trends in children's reading literacy achievement 1991-2001: IEA's repeat in nine countries of the 1991 reading literacy study*. Chestnut Hill, MA: Boston College.

Martin, M.O., Mullis, I.V.S., & Kennedy, A.M. (Eds.). (2007). *PIRLS 2006 technical report.* Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Kennedy, A.M. (2003). *PIRLS 2001 international report: IEA's study of reading literacy achievement in 35 countries*. Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M. , & Sainsbury, M. (2006). *PIRLS 2006 assessment frameworks and specifications* (2nd ed.). Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M., & Foy, P. (2007). *PIRLS 2006 international report: IEA's progress in international reading literacy study in primary schools in 40 countries*. Chestnut Hill, MA: Boston College.

SAS Institute (2002). *SAS system for Windows* (version 9.1). Cary, NC: SAS Institute.

SPSS Inc. (2005). *SPSS for Windows* (version 14.0). Chicago, IL: SPSS Inc.

Thorndike, R. L. (1973). *Reading comprehension education in fifteen countries. International studies in evaluation III*. Stockholm: Almqvist and Wiksell.

Walker, D.A. (1976). *The IEA six subject survey: An empirical study of education in twenty-one countries*. New York: John Wiley & Sons Inc.

Chapter 2

The PIRLS 2006 Data Files

2.1 Overview

This chapter discusses the aspects of the study that determine the structure and content of the database. Details of the data files for achievement items, achievement scores, and background variables are provided with special attention to the file and variable naming conventions for each file type.

2.2 PIRLS 2006 Achievement Block and Booklet Design

As shown in Exhibit 2.1, half of the 10 test blocks were devoted to measuring literary purposes (labeled L1-L5), and the other half were focused on acquiring and using information (labeled I1-I5). Two of the literary blocks ("Lump of Clay" and "Flowers") and two of the informational blocks ("Antarctica" and "Leonardo") were used in the PIRLS 2001 assessment. The remaining six blocks were used for the first time in 2006.

Exhibit 2.1 PIRLS 2006 Achievement Test Blocks

Literary Block Number	Literary Block Title	Informational Block Number	Informational Block Title
L1	Lump of Clay (2001)	I1	Antarctica (2001)
L2	Flowers (2001)	I2	Leonardo (2001)
L3	Shiny Straw (2006)	I3	Day Hiking (2006)
L4	Fly Eagle (2006)	I4	Sharks (2006)
L5	Unbelievable Night (2006)	I5	Searching for Food (2006)

The 10 achievement blocks were distributed across 13 booklets, two per booklet, so that passages were paired together in a booklet in as many different ways as possible. Accordingly, each student booklet consisted of two 40-minute blocks of passages and items. The PIRLS 2006 assessment included 126 items across the 10 assessment blocks, comprising a total of 167 score points. The numbers of multiple-choice and constructed-response items by reading purpose are presented in Exhibit 2.2. The two question formats—constructed response and multiple choice—were evenly represented in the total number of items, with 64 multiple-choice items and 62 constructed-response items in the assessment. The total number of items and score points were distributed equally between the two purposes for reading.

Exhibit 2.2 PIRLS 2006 Assessment Item Specifications

	Number of Multiple-choice Items	Number of Constructed-response Items			Total Number of Items	Total Number of Score Points
		1 pt.	2 pts.	3 pts.		
Literary	34	13	13	4	64	85
Informational	30	15	14	3	62	82
Total	64	28	27	7	126	167

Exhibit 2.3 presents the portion of the assessment devoted to each of the four processes of reading comprehension. The distribution of actual score points across the processes approximates the distribution established in the PIRLS 2006 framework. Retrieval and straightforward inferencing processes combines items from the focus on and retrieve explicitly stated information and make straightforward inferences comprehension processes. Similarly, interpreting, integrating, and evaluating processes are based on items from the interpret and integrate ideas and information and examine and evaluate content, language, and textual elements processes.

Exhibit 2.3 Distribution of Score Points Across Reading Processes

PIRLS 2006 Processes of Reading Comprehension	Number of Score Points	Percentage of Total Score Points
Focus on and retrieve explicitly stated information	36	22
Make straightforward inferences	47	28
Interpret and integrate ideas and information	61	37
Examine and evaluate content, language, and textual elements	23	14
Total	167	100

Item Release Policy

As in the previous PIRLS assessment in 2001, the design for PIRLS 2006 and beyond (2011, 2016, etc.) provides for retaining some of the passages and items for the measurement of trend and releasing some passages into the public domain. In PIRLS 2006, two of the assessment blocks for each reading purpose were released after the assessment results for 2006 were published, one that originated in 2001 and one from 2006 (The PIRLS Reader), for a total of four released passages ("Lump of Clay", "Antarctica", "Unbelievable Night", and "Searching for Food"). For PIRLS 2011, new passages and items will be developed to take the place of the released achievement blocks. The DVD that accompanies this user guide includes a PDF version of the released passages and items with the corresponding scoring guides for constructed-response items. The DVD also includes an item information file that identifies items from the released and the retained blocks.

2.3 PIRLS 2006 Data Files

This section describes the content and format of the database files. Data files are provided for each country that participated in PIRLS and for which internationally comparable data are available.

Data files include:

- Student achievement data files with responses to the items on the PIRLS test and the student reading scores (plausible values) for each of the five PIRLS reading achievement scales, as well as within-country scoring reliability scores.
- For each background questionnaire, a background data file with information from students, their parents or guardians, their reading teachers, and the principals of their schools. The student background data file also includes the student reading scores (plausible values) on the assessment.

The files and the variables they contain are described in the following sections.

Data File Naming Convention

Before describing the various data files, this section presents the file naming convention used in the PIRLS database. The filenames of the data files included in the database consist of an eight-character string followed by a three-character file extension (e.g., ASGBGRR2), and use the following conventions:

- The first character of a file name is always "A".
- The second character indicates the source, or level, of the information in a file. The letter "C" indicates a school-level data file, the letter "T" a teacher-level data file, the letter "S" a student-level data file and the letter "U" is reserved for the curriculum questionnaire data file. For example, the file A<u>S</u>GBGRR2 is a student-level data file.
- The third character indicates the type of data in a file. The letter "A" is used for student achievement data, the letter "H" for home questionnaire data, the letter "G" for all other background questionnaire data, the letter "R" for scoring reliability data, and "T" for student-teacher linkage data. For example, the file AS<u>G</u>BGRR2 is a student background questionnaire data file.

- Characters four through six identify the country using a three-character alphanumeric country abbreviation based on the ISO 3166 coding scheme. Exhibit 1.1in Chapter 1 lists the codes of all participating countries. The curriculum questionnaire data file uses the three-letter combination PRL. The example student background data file ASGBGRR2 is for Bulgaria.
- The seventh and eighth characters indicate the study cycle. The combination "R2" is used in PIRLS 2006, whereas "R1" was used in the PIRLS 2001 international database.
- The three-character file extensions used for the data files are .EXP for SAS export data, .SAV for SPSS data.

For each file type, a separate data file is provided for each participating country. Countries for which data are not available for a particular type have an empty file. There is a single curriculum questionnaire data file in Excel format, with all contributing countries' data included.

PIRLS Student Achievement Data File (ASA)

The student achievement data file contains the student responses to the individual achievement items in the PIRLS reading test. The student achievement data file is best suited for performing item-level analyses. Achievement scores (plausible values) for each of the PIRLS 2006 reading scales are available in the student achievement data file, as well as in the background questionnaire data file. Thus, it is usually not necessary to link these two data file types.

Students who participated in PIRLS were administered one of 13 test booklets, each with a series of questions. Some of these questions were multiple-choice items and some were constructed-response items. The student achievement data files contain the actual responses to the multiple-choice questions and the scores assigned to the constructed-response items.

Achievement Item Variable Naming Conventions

The achievement item variable names are based on an 8-character alphanumeric code (e.g., R021U01M), which adheres to the following rules:

- The first character is always "R" to denote Reading.
- The second and third characters indicate the assessment year when the item was first used in PIRLS. The code "01" is used for items introduced in PIRLS 2001 and "02" is used for items introduced in PIRLS 2006.

- The fourth character is always "1" and indicates the population for which the item is intended, which is the fourth grade of primary schooling in PIRLS.
- The fifth character indicates the passage where the item is located. The following codes are used:

  A Antarctica: Land of Ice

  C The Little Lump of Clay

  E Fly, Eagle, Fly

  F Flowers on the Roof

  K Sharks

  L Leonardo da Vinci

  N Discover the Fun of Day Hiking

  S Searching for Food

  U An Unbelievable Night

  Y Shiny Straw
- The sixth and seventh characters are a sequential number that identifies the sequential position of the item within the passage.
- The eighth character indicates the item type. The letter "C" is used for constructed-response items and the letter "M" for multiple-choice items. Therefore, the example item R021U01M is a multiple-choice reading item that is the first item in the block Unbelievable Night, which was first used in PIRLS 2006.

Achievement Item Response Code Values

A series of conventions also were adopted to code the data included in the data files. This section describes these conventions for the achievement items.

The values assigned to each of the achievement item variables also depend on the item format. For multiple-choice items, numerical values from 1 through 4 are used to correspond to the response options A through D, respectively. For these items, the correct response is included as part of the variable label in the

codebook file and a program is included as part of the database to score these items. For the constructed-response items, numerical values ranging from 0 to 3 are used to correspond to the score points obtained from the students' responses.

Codes for Missing Values

A subset of the values for each variable type is reserved for specific codes related to different categories of missing data. We recommend that you read the following section with particular care since the way in which you make use of these missing codes may have serious consequences for your analyses.

Omitted Response Codes (SAS: . ; SPSS: sysmis, 9)

"Omitted" response codes are used for items that a student should have answered but did not. An omitted response code also is given when an item is left blank or when two or more response options are checked for a multiple-choice item.

Not Administered Response Codes (SAS: .A; SPSS: 8)

Special codes were given to items that were "Not Administered" to distinguish these cases from data that are missing due to non-response. In general, the not administered code is used when an item was not administered as part of the questionnaire or test instruments either by design, such as in the case of the rotated test items, or unintentionally, such as when an item was left out of the instrument or misprinted. The not administered code is used in the following cases:

- Achievement item not assigned to the student: all students participating in PIRLS received only one of the 13 test booklets. All variables corresponding to items that were not given to a student are coded as "Not Administered."
- Student absent from session: When a student was not present for a particular testing session, all variables relevant to that session are coded as "Not Administered."
- Item left out or misprinted: When a particular item (or a whole page) was misprinted, or not available to the respondent, the corresponding variables are coded as "Not Administered."

- Achievement items omitted or mistranslated: Any item identified during translation verification or item review as having a translation error, such that the nature of the question was altered, is coded as "Not Administered."

Not Reached Response Codes (ASCII: 6 ; SAS: .R ; SPSS: 6)

Items left blank at the end of a passage are considered to be not reached. For most purposes, PIRLS treats the not-reached items as incorrect responses, except during the item calibration step of the IRT scaling, when not-reached items are considered to have not been administered (see Foy, Galia, & Li, 2007).

PIRLS 2006 Achievement Scores in the Student Files

Achievement scales were produced for each of the two reading purposes (reading for literary experience and reading for information) and for two processes of comprehension (retrieving and straightforward inferencing, and interpreting, integrating, and evaluating), as well as for reading overall. A detailed description of the PIRLS 2006 scaling is available in the *PIRLS 2006 Technical Report* (Foy, Galia, & Li, 2007). For each of the five achievement scales (see Exhibit 2.4), the PIRLS 2006 database provides five separate estimates of each student's score. The five score estimates are known as "plausible values," and the variability between them encapsulates the uncertainty inherent in the score estimation process.

The plausible values are the best measures of student achievement in reading available in the PIRLS 2006 international database, and should be used as the outcome measure in any study of student achievement. Plausible values may readily be analyzed using the IDB Analyzer and other software described in this User Guide. As shown in Exhibit 2.4 the first letter of the variable name follows the convention used with other background and derived variables and files included in the database.

Exhibit 2.4 Plausible Values Variables for PIRLS 2006 Reading Achievement Scales

Variable	Scale
ASRREA01-05	Reading Overall
ASRLIT01-05	Reading for Literary Purposes
ASRINF01-05	Reading for Informational Purposes
ASRRSI01-05	Retrieving and Straightforward Inferencing Processes
ASRIIE01-05	Interpreting, Integrating, and Evaluating Processes

In addition to the plausible values for reading overall and by purpose and process, the PIRLS database includes three interim achievement scores that were computed as part of the data processing effort and included in the achievement data files.

Raw Scores

After the achievement items were scored (1 for correct, 0 for incorrect for multiple choice items; 0, 1, 2, or 3 points in the case of the constructed-response items), raw scores were computed by adding the number of points obtained by each student over all the items in the student's test book. Because the raw score is dependent on the number of items in the student's test book, and since this number varies from test book to test book, the raw scores are not comparable across booklets, and so are of limited utility. Their main value in the database is as a validity check for analysts who wish to apply a different scoring approach to the PIRLS items.

Standardized Raw Scores

Because of the difficulty in making any comparisons across the test booklets using only the number of raw score points obtained on a set of items, raw scores were standardized by booklet to provide a simple score that could be used in comparisons across booklets in the same survey cycle in preliminary analyses. The standardized score was computed so that the weighted mean

score within each booklet in a country was equal to 50, and the weighted standard deviation was equal to 10.

National Rasch Scores

The national Rasch scores were computed to facilitate preliminary item analyses that were conducted prior to the PIRLS 2006 IRT scaling. Their main purpose was to provide a preliminary measure of overall reading achievement that could be used as a criterion variable in studies of item discrimination. The national Rasch scores were standardized to have a mean score of 150 and a standard deviation of 10 within each country. Because each country has the same mean score and dispersion, these scores are not useful for international comparisons.

International Benchmarks of Achievement

To help users of the PIRLS achievement results understand what performance on the reading achievement scale means in terms of the types of texts they could read and the types of items they could answer successfully, PIRLS identified four points on the reading scale to serve as international benchmarks. As shown in Exhibit 2.5, the benchmark scores are 625, 550, 475, and 400, corresponding to the Advanced International Benchmark, the High International Benchmark, the Intermediate International Benchmark, and the Low International Benchmark, respectively. PIRLS used a technique known as scale anchoring to summarize and describe student achievement at these four points on the scale. The *PIRLS 2006 International Report* (Mullis, Martin, Kennedy, & Foy, 2007) presents the results of this scale anchoring, and reports the percentage of students in each country reaching each of the international benchmarks.

Exhibit 2.5 PIRLS 2006 International Benchmarks for Reading Achievement

Scale Score	International Benchmark
625	Advanced International Benchmark
550	High International Benchmark
475	Intermediate International Benchmark
400	Low International Benchmark

To facilitate analysts in using the international benchmarks in secondary analyses, the PIRLS international database contains a set of variables indicating which international benchmark the student reached. There is a benchmark variable for each plausible value (ASRIBM01-05).

The following codes are used for the benchmark variables:

Code 1	Student performed below the Low International Benchmark
Code 2	Student performed at or above the Low International Benchmark, but below the Intermediate International Benchmark
Code 3	Student performed at or above the Intermediate International Benchmark but below the High International Benchmark
Code 4	Student performed at or above the High International Benchmark but below the Advanced International Benchmark
Code 5	Student performed at or above the Advanced International Benchmark

PIRLS Within-country Scoring Reliability Data File (ASR)

The within-country scoring reliability data files contain data that can be used to investigate the reliability of the PIRLS constructed-response item scoring. The scoring reliability data files contain one record for each booklet that was double scored during the within-country scoring reliability exercise (see Barth, Kennedy, & Trong, 2007). For each constructed-response item in the achievement test, the following three variables are included in the scoring reliability data files:

- Original Score Variable (achievement item response scores obtained from the first scorer)
- Second Score Variable (achievement item response scores obtained from the second scorer)
- Score Agreement Variable (degree of agreement between the two scores).

Scoring Reliability Variable Naming Convention

The variable names for the Original Score, Second Score, and Score Agreement variables in the scoring reliability data files are based on the same naming convention as that for the achievement item variables shown earlier. The second character in the variable name differentiates between the three reliability variables:

- The Original Score variable has the number "0" as the second character, in accordance with the achievement item naming convention (e.g., R021K12C)
- The Second Score variable has the letter "R" as the second character (e.g., RR21K12C)
- The Score Agreement variable has the letter "I" as the second character (e.g., RI21K12C).

Reliability Variable Score Values

The values contained in both the Original Score and Second Score variables are the scores assigned using the PIRLS scoring guides. The Score Agreement variable is given the value "1" if both scores are identical, or the value "0" if they are not.

In general, the data in the Original Score variables are identical to those contained in the Student Achievement data files. In some cases, however, the response scores for specific items were recoded after a review of the international item statistics revealed inconsistencies in the original scoring guides or showed that the original scores were not functioning as desired.[1] The recoded score values were used in computing the achievement scores reflected in the international reports.

The response scores in the student achievement data files reflect the recoded values. In contrast, the Original Score variables in the scoring reliability data files contain the original unrecoded response scores. This was done so that the scoring reliability measure indicated in the Score Agreement variables was based on the original scoring guides used during the constructed-response scoring sessions conducted in each country.

Achievement Item Information File

An item information file is provided to enable users of the database to readily produce summaries of item characteristics. The item information file, available on the DVD, includes the following information for each item in the PIRLS 2006 assessment:

- The item's permanent and unique identifier
- The item's name and label
- The passage to which the item belongs and its location within the passage

1 Item R011F12 is the only PIRLS item affected by such a recode.

- The item's type, either multiple-choice or constructed response
- The item's correct response key (coded as "X" for constructed-response items)
- The item's points value, either 1, 2 or 3 points
- An indicator showing if the item was included in the IRT scaling (SCORE)
- An indicator if the item is released after the 2006 assessment
- The item's classification into the purposes of reading and the processes of comprehension.

The PIRLS item information file is named AITINFR2.XLS and is an Excel spreadsheet with one row for each PIRLS 2006 achievement item.

Achievement Data Almanacs

The achievement data almanacs provide weighted summary statistics for each individual reading achievement item included in the PIRLS assessment. There are separate achievement almanacs for the two purposes of reading—reading for literary experience and reading to acquire and use information. The achievement data almanacs also display for each item its classification in the purposes of reading and the processes of comprehension, the name of the passage it belongs to, a brief description of the item, its variable name, whether it is a multiple-choice or constructed-response item, and the correct response key if it is a multiple-choice item (the key is coded as "X" for constructed-response items). The trend almanacs provide summary statistics for achievement items used in both the 2001 and 2006 assessments. The achievement data almanac files are listed in Exhibit 2.6.

Exhibit 2.6 Achievement Data Almanacs for PIRLS 2006

Data Almanac Files	Contents
P06_ItemAlmanac_LIT	Achievement Data Almanac for Literary Experience Items
P06_ItemAlmanac_INF	Achievement Data Almanac for Acquire and Use Information Items
P06_TrendItemAlmanac_LIT	Trend Achievement Data Almanac for Literary Experience Items
P06_TrendItemAlmanac_INF	Trend Achievement Data Almanac for Acquire and Use Information Items

The data almanacs are provided as PDF files, which can be read with Adobe Acrobat Reader 4.0 or higher, and as Word documents. The files display student-weighted summary statistics for each participating country on each variable. The almanacs also display the international averages for each variable, with each country weighted equally. The Canadian provinces are not included in the calculation of international averages.

There are two types of displays in the almanacs, depending on whether an item is a multiple-choice item or a constructed-response item. The statistics displayed in these almanacs are as follows:

- N: The number of students to whom the item was administered.
- DIFF: Percent of students that responded correctly to a multiple-choice item.
- A, B, C, D: The percent of students choosing each one of the response options for a multiple-choice item.
- 0, 1, 2, 3: The percent of student responses assigned each of the codes in the scoring guide for a constructed-response item.
- OMITTED: The percent of students that omitted to respond to the test item.
- NOT REACHED: The percent of students that did not reach the test item.
- OTHER INCORRECT: The percent of students who either omitted or did not reach the item (i.e., the sum of the previous two categories).
- V1, V2, V3: The percent of students that scored 1 point or better on the item (V1), 2 points or better (V2) or 3 points or better (V3).
- 1.GIRL % RIGHT, 2.BOY % RIGHT: The percent of girls and boys that either got a multiple-choice item right, or obtained the maximum score on a constructed-response item.

Columns “2” and “V2” will only appear for 2- and 3- point constructed-response items. Conversely, columns “3” and “V3” will only appear for 3- point constructed-response items.

Background Questionnaire Data Files

There are five PIRLS background data files—student, home, teacher, school, and curriculum—corresponding to the five background questionnaires administered in PIRLS. These data files contain the responses to the questions asked in the background questionnaires.

Student Background Data File (ASG)

The student background data file contains students' responses to the PIRLS 2006 student questionnaire. These files also contain a number of identification variables, tracking variables, sampling variables, and derived variables that were used for producing exhibits in the international report. In addition, the student background files contain the student reading achievement scores (plausible values).

Home Background Data File (ASH)

The home background data files contain the responses of parents or primary caregivers to the questions in the *Learning to Read Survey*, or home questionnaire. To perform student-level analyses with student reading achievement scores and the home background data, these files must be merged with the student background data files, using the country and student identification variables, to perform analyses on the variables they contain. The PIRLS database DVD contains the IEA's IDB Analyzer software, as well as SPSS and SAS macros, for merging these files. Details of the merging procedure are described in Chapters 3, 4, and 5 of this User Guide.

Teacher Background Data File (ATG)

The teachers of the students who were sampled for PIRLS were administered a teacher background questionnaire with questions pertaining to their background and their teaching practices in the classes of the sampled students. Each teacher was asked to respond to a questionnaire for each class taught that contained sampled students. The teacher background data files contain one record for each teacher who teaches the sampled students. If a teacher taught more than one class, he or she was expected to complete a questionnaire for each class taught. In the teacher background data files, each teacher is assigned a unique identification number (IDTEACH) and a link number (IDLINK) that is specific to the class taught by that teacher and to which the information in the data record

corresponds. The combination of the teacher identification and link numbers uniquely identifies a teacher teaching a specific class.

It is important to note that the teacher background data files do not constitute a representative sample of teachers in a country, but rather consist of the teachers who teach representative samples of students. The teacher data should therefore be analyzed only in conjunction with the student-teacher linkage data file. Chapters 3, 4, and 5 of this User Guide describe student-level analyses using the student-teacher linkage file in the IEA's IDB Analyzer software, as well as SPSS and SAS, respectively.

School Background Data File (ACG)

The school background data files contain school principals' responses to the questions in the school background questionnaire. School-level analyses where the schools are the units of analysis can be performed directly from the school background data files. To perform student-level analyses, the school background data files must be merged with the Student Background data files using the country and school identification variables. Details of the merging procedure using the IEA's IDB Analyzer, SPSS, or SAS, are described in Chapters 3, 4, and 5 of this User Guide, respectively.

Student-Teacher Linkage File (AST)

The student-teacher linkage data files contain information required to link the student and teacher data files. The student-teacher linkage data files contain one entry per student-teacher linkage combination in the data. For instance, if three reading teachers are linked to a student, there are three entries in the file corresponding to that student. The sole purpose of the student-teacher linkage data files is to link teacher-level data with student-level data to perform appropriate student-level analyses.

PIRLS Curriculum Questionnaire Data File (AUG)

The Curriculum Questionnaire data file for PIRLS contains the responses provided by the National Research Coordinators of the participating countries to the PIRLS 2006 Curriculum Questionnaire.

Background Variable Naming Convention

The background variable naming convention is based on a 7- or 8-character string (e.g., ASBGBOOK). The following rules are applied in naming the background variables:

- The first character is always "A."
- The second character indicates the response level. The letter "C" is used identify data from the school principals, the letter "T" is used for teacher data and the letter "S" for student data.
- The third character is used to distinguish responses from the background questionnaires, with the letter "B", and derived variables with the letter "D".
- The fourth character is used to distinguish responses from the *Learning to Read Survey*, where the letter "H" is used, and responses from all other questionnaires, where the letter "G" is used.
- The fifth through eighth characters of all background questionnaire variables in the PIRLS data files are used to assign a unique and concise label to each question.

The variable names in the curriculum questionnaire file follow a different naming convention, based on location code, as described in the following section.

Linking Questions in the Questionnaires to Variables in the International Database

To identify the location of a background variable in its corresponding background questionnaire, each questionnaire was assigned a unique identification code as shown in Exhibit 2.7. This unique code is followed by the sequence number of the question within the questionnaire. For example, if the location of a variable is given as SG1-06A, it refers to Part A of Question 6 in the student background questionnaire. This convention is followed in the data almanacs and in the description of the variables included in Supplements 1 and 2 to this User Guide (provided on the DVD).

Exhibit 2.7 Background Variable Location Convention

Questionnaire	Location
Student Questionnaire	SG***
Home Questionnaire (Learning to Read Survey)	SH***
Teacher Questionnaire	TG***
School Questionnaire	CG***
Curriculum Questionnaire	ACQ, RCQ***

Data Coding Conventions

A series of conventions also were adopted to code the data included in the data files. This section describes these conventions.

Background Question Response Code Values

The values assigned to each of the background question variables depend on the item format and the number of options available. For the multiple-choice questions, one-digit numerical values are used to correspond to the response option. This number corresponds to the sequence of the letter in the alphabet. For example, response option A is represented with a 1, response option B with a 2, etc. Open-ended questions such as "the number of students in a class" are coded with the actual number given as a response.

Codes for Missing Values

A subset of the values for each variable type is reserved for specific codes related to various categories of missing data. The missing categories defined below are assigned values depending on the field width of the variable and the variable type.

Omitted Response Codes (SAS: . ; SPSS: sysmis, 9, 99, 999, ...)

"Omitted" response codes are used for questions that a student, parent, teacher, or school principal should have answered but did not. The length of the omitted response code given to a variable in the SPSS data files depends on the number of characters needed to represent the variable. In all cases, the space necessary to represent the variable is filled with 9's. For questionnaire data, no differentiation has been made between items left blank and items with invalid answers, such as checking two or more response options in a categorical question, or unreadable or uninterpretable responses to open-ended questions. In a few cases, data received from a country in an invalid or inconsistent manner were also coded as "omitted."

Not Administered Response Codes (SAS: .A; SPSS: 8, 98, 998, ...)

Special codes were given to items that were "Not Administered" to distinguish these cases from data that are missing due to non-response. The specific not administered code value given in the SPSS data files depends on the variable field length. In general, the not administered code is used when a question was not administered, such as when a question was left out of the instrument or misprinted. The not administered code is used in the following cases:

- Question left out or misprinted: When a particular question (or a whole page) was misprinted, or not available to the respondent, the corresponding variables are coded as "Not Administered."
- Background questions omitted: Variables corresponding to questions in the student, home, teacher, or school background questionnaires that were considered not applicable in some countries were not included in the national versions of the questionnaires. These questions are coded as "Not Administered."
- Background questions mistranslated or not internationally comparable: In some cases, questions in the international version of the questionnaires were mistranslated or modified to fit the national context. Whenever possible, modified questions were recoded to match as closely as possible the international version. When this was not possible, modified questions were recoded as "Not Administered."

Not Applicable Response Codes (ASCII: 6, 96,... ; SAS: .B ; SPSS: 6, 96,...)

"Not Applicable" response codes are used for the background questionnaire items in which responses are dependent on a filter question. Generally, a "No" response to a filter question leads to any follow-up questions being coded as "Not Applicable" since there are no appropriate responses to these follow-up questions.

Summary Indices and Derived Variables from Questionnaire Data

In the PIRLS questionnaires, there were often several questions asked about various aspects of a single construct. In these cases, responses to the individual items were combined to create a derived variable which provided a more comprehensive picture of the construct of interest than the individual variables could on their own.

Student records were included in the derived variable calculation only if there were data available for two thirds of the variables involved. For example, if a derived variable was based on six component variables, students who were missing responses to more than two of these were counted as missing on the derived variable.

In the PIRLS reports, an index is a special type of derived variable that assigns students to one of three levels—high, medium, and low—on the basis of their responses to the component variables. The high category of an index represents the responses that are expected to characterize aspects of a positive literacy environment, and the low category those responses that are least supportive of literacy.

Supplement 3 of the user guide provides a description of all derived variables included in the international database.

Background Data Almanacs

Background data almanac files contain weighted summary statistics for each participating country on each variable in the student, home, teacher, and school background questionnaires including derived variables based on these background variables. The background data almanac files corresponding to each variable type are listed in Exhibit 2.8 and are described in the following section.

Exhibit 2.8 Background Data Almanacs for PIRLS 2006

Data Almanac Files	Contents
P06_StudentAlmanac	Student Background Data Almanac with Reading Achievement
P06_SchoolAlmanac	School Background Data Almanac with Reading Achievement
P06_TeacherAlmanac	Teacher Background Data Almanac with Reading Achievement
P06_HomeAlmanac	Home Background Data Almanac with Reading Achievement

The data almanacs are provided as PDF files, which can be read with Adobe Acrobat Reader 4.0 or higher, and Word documents. The files display student-weighted summary statistics for each participating country on each variable. The almanacs also display the international averages for each variable, with each country weighted equally. The Canadian provinces are not included in the calculation of international averages.

There are two types of displays in the background data almanacs, depending on whether the data are categorical (i.e., have a small number of discrete values) or continuous. The display for categorical variables includes:

- The sample size (number of students, teachers or schools included in the sample)
- The number of valid cases (number of students, parents, teachers or schools for whom valid data were obtained)
- The weighted percentages of students corresponding to each valid response option (percentages based only on the students with valid data)
- The weighted percentages of students for whom none of the valid response options were selected, coded as "Not Administered" or "Omitted" (percentages based on the sample size)
- The weighted mean achievement values of students corresponding to each valid response option, as well as the "Not Administered" and "Omitted" codes
- In cases where a variable can be coded as "Not Applicable" because of an earlier filter question, the weighted percentage of students for whom the

variable is coded as "Not Applicable" is also displayed, based only on the students with valid data, along with the corresponding weighted mean achievement

The display for continuous variables includes:

- The sample size (number of students, teachers or schools included in the sample)
- The number of valid cases (number of students, parents, teachers or schools for whom valid data were obtained)
- The weighted percentages of students for whom the variable is coded as "Not Administered" or "Omitted" (percentages based on the sample size)
- The weighted mean, mode, minimum, maximum, and the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles across students (based only on the students with valid data)
- In cases where a variable can be coded as "Not Applicable" because of an earlier filter question, the weighted percentage of students for whom the variable is coded as "Not Applicable" is also displayed, based only on the students with valid data

The background data almanacs also display for each variable the question as it was asked, its location in the corresponding questionnaire, and its variable name in the data files.

2.4 Using Sampling Weights in Analyzing the PIRLS 2006 Data

An important characteristic of the PIRLS studies, and one that has crucial implications for data analysis, is that they use data from carefully-drawn random samples of schools, classes, and students to make inferences about the reading achievement at the fourth grade. For analyses based on these sample data to accurately reflect population attributes, it is necessary that they take the design of the sample into account. This is accomplished in part by assigning a sampling weight to each respondent in the sample, and weighting the respondent by its sampling weight in all analyses. 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 (see Joncas, 2007).

Because the students within each country were selected using probability sampling procedures, the probability of each student being selected as part of the

sample is known. The sampling weight is the inverse of this selection probability. In a properly selected and weighted sample, the sum of the weights for the sample approximates the size of the population. In PIRLS, the sum of the sampling weights of all students in a country is an estimate of the size of the fourth grade student population in that country. The student sampling weight, known as TOTWGT in the international database, must be used whenever student population estimates are required. The use of TOTWGT ensures that the various subgroups that constitute the sample are properly and proportionally represented in the computation of population estimates, and that the sample size will be inflated to approximate the size of the population.

Because statistics generated from the international database are estimates of national performance based on samples of students, rather than the value that could be calculated if every student in every country had answered every question, it is important to have a way of quantifying the uncertainty associated with these statistics. In PIRLS, the jackknife procedure is used to provide a robust estimate of the standard error of each statistic presented in the international report. When used with achievement scores, the jackknife standard errors include both an error component due to sampling variation and an error component due to variation among the five plausible values generated for each student. The standard errors may be used to create confidence intervals for statistics computed from the PIRLS data.

The PIRLS 2006 international database includes a set of programs and macros in both SAS and SPSS that enable analysts to apply the jackknife algorithm to a range of analyses of school, teacher, and student variables.

Sampling Weights Included in the Student Data Files

Several sampling and weighting variables are included in the data files.

TOTWGT	Total Student Weight
SENWGT	Student Senate Weight
HOUWGT	Student House Weight
TCHWGT	Overall Teacher Weight
SCHWGT	School-level Weight
JKZONE	The jackknife sampling zone, or stratum, to which the student's school is assigned

JKREP	The jackknife replicate, or primary sampling unit, to which the student's school is assigned
JKCZONE	The jackknife sampling zone, or stratum, to which the school is assigned
JKCREP	The jackknife replicate, or primary sampling unit, to which the school is assigned
WGTFAC1	School Weighting Factor
WGTADJ1	School Weighting Adjustment
WGTFAC2	Class Weighting Factor
WGTADJ2	Class Weighting Adjustment
WGTFAC3	Student Weighting Factor
WGTADJ3	Student Weighting Adjustment

Exhibit 2.9 illustrates the location of the various sampling and weighting variables among the different data file types.

Exhibit 2.9 Location of Sampling and Weighting Variables in the PIRLS Database

Sampling and Weighting Variables	Data File Types					
	ASG	ASA	ASH	AST	ATG	ACG
JKREP	✓	✓		✓		
JKZONE	✓	✓		✓		
JKCREP						✓
JKCZONE						✓
TOTWGT	✓	✓				
SENWGT	✓	✓				
HOUWGT	✓	✓				
TCHWGT				✓		
SCHWGT						✓
WGTADJ1	✓					✓
WGTADJ2	✓					
WGTADJ3	✓					
WGTFAC1	✓					✓
WGTFAC2	✓					
WGTFAC3	✓					

Although TOTWGT has desirable properties, it also has drawbacks for some analyses. Because TOTWGT sums to the student population size in each country, analyses using TOTWGT that combine countries will have more students from larger countries and fewer from smaller countries, which may not be desirable for some purposes. For cross-country analyses in which each country should be treated equally, PIRLS provides SENWGT, a transformation of TOTWGT that results in a sample size of 500 in each country. Additionally, since TOTWGT inflates sample sizes to approximate the population size, software systems that use sample size to compute significance tests will give misleading results for analyses weighted by TOTWGT. HOUWGT, a transformation of TOTWGT, ensures that the weighted sample corresponds to the actual sample size in each country.

The weight variable TCHWGT is specifically designed for using teacher background data in student-level analyses. The weight variable SCHWGT is designed for use in school-level analyses where the schools are the units of

analysis. The weight variables TOTWGT, SENWGT and HOUWGT are designed for use in student-level analyses from all student-level files.

The sampling variables beginning with the letters "JK" are used to compute standard errors based on the jackknife repeated replication methodology. All weighting variables beginning with the letters "WGT" are included to provide insight into the multi-stage sampling and weighting methodology applied to the PIRLS data.

2.5 Structure and Design Variables Included in PIRLS 2006 Data Files

Besides the variables used to store responses to the background questionnaires and achievement booklets, the PIRLS database also contains variables meant to store information that identify and describe the respondents and design information required to properly analyze the data.

Identification Variables

In all data files, several identification variables are included that provide information to identify countries, students, teachers, or schools. These variables also are used to link cases between the different data file types. The identification variables have the prefix "ID" and are described below.

IDCNTRY

IDCNTRY is a five-digit country identification code based on the ISO 3166 classification as shown in Exhibit 1.1. This variable should always be used as the first linking variable whenever files are linked within and across countries.

IDPOP

IDPOP identifies the target population and is always set to "1" for PIRLS.

IDSCHOOL

IDSCHOOL is a four-digit identification code that uniquely identifies the participating schools within each country. The school codes are not unique across countries. Schools across countries can be uniquely identified only with the IDCNTRY and IDSCHOOL combination of linking variables.

IDGRADE

IDGRADE identifies the target grade of the participating students. In PIRLS 2006, the value is "4" for most countries, or "5" for the few countries that tested at the fifth grade.

IDCLASS

IDCLASS is a six-digit identification code that uniquely identifies the sampled classrooms within a country. The variable IDCLASS has a hierarchical structure and is formed by concatenating the IDSCHOOL variable and a two-digit sequential number identifying the sampled classrooms within a school. Classrooms can be uniquely identified in the database by the combination of IDCNTRY and IDCLASS as linking variables.

IDSTUD

IDSTUD is an eight-digit identification code that uniquely identifies each sampled student in a country. The variable IDSTUD also has a hierarchical structure and is formed by concatenating the IDCLASS variable and a two-digit sequential number identifying all students within each classroom. Students can be uniquely identified in the database by the combination of IDCNTRY and IDSTUD as linking variables.

IDBOOK

IDBOOK identifies the specific assessment booklet that was administered to each student. The reader is given the value "0" whereas all other booklets are given their numerical value from 1 through 12.

IDSTRATE & IDSTRATI

IDSTRATE and IDSTRATI are identification variables generated by the school sampling process. IDSTRATE identifies the explicit strata and IDSTRATI the implicit strata from which the participating schools were sampled. The codes assigned to these two variables vary from country to country and are documented in Supplement 4 to the User Guide.

IDTEACH

IDTEACH is a six-digit identification code that uniquely identifies a teacher within a school. It has a hierarchical structure and is formed by the concatenation of IDSCHOOL and a two-digit sequential number within each school.

IDLINK

IDLINK uniquely identifies the class for which a teacher answered a questionnaire. The combination of linking variables IDCNTRY, IDTEACH, and IDLINK uniquely identifies all teacher-class combinations in the database.

Exhibit 2.10 shows in which data files the various identification variables are located. It also highlights the combinations of variables used to uniquely identify the records contained in the data file types. In the student background and achievement data files and the home background data files, the variables IDCNTRY and IDSTUD provide a unique identification number to identify all students in the database. Since teachers may teach more than one class, the combination of the IDCNTRY, IDTEACH and IDLINK variables in the teacher background data files is needed to uniquely identify all teachers and the classes they teach. Teacher background variables are linked to the appropriate students using the student-teacher linkage data file. The variable IDSCHOOL, contained in all files, is a unique identification number for each school within a country. Combined with IDCNTRY, it can be used to link school background data to corresponding students or teachers.

Exhibit2.10 Location of Identification Variables in the PIRLS Database

Identification Variables	Data File Types					
	ASA	ASG	ASH	AST	ATG	ACG
IDCNTRY	✓	✓	✓	✓	✓	✓
IDGRADE	✓	✓	✓	✓	✓	✓
IDPOP	✓	✓	✓	✓	✓	✓
IDSCHOOL	✓	✓	✓	✓	✓	✓
IDCLASS	✓	✓	✓	✓	✓	
IDSTUD	✓	✓	✓	✓		
IDBOOK	✓	✓				
IDSTRATE	✓	✓				
IDSTRATI	✓	✓				
IDTEACH				✓	✓	
IDLINK				✓	✓	

Tracking Variables

Information about students, teachers, and schools provided by the survey tracking forms[2] is incorporated in the tracking variables. These variables have the prefix "IT." All tracking variables are included in the student background data files. ITPART is included in all data file types and ITLANG is also included in the student achievement data files.

ITSEX

Gender of each student as stated in the Student Tracking Forms

ITBIRTHM and ITBIRTHY

Month and year of birth of each student as stated in the Student Tracking Forms

ITDATEM and ITDATEY

Month and year of testing for each student

2 Survey Tracking Forms are lists of students, teachers, or schools used for sampling and administrative purposes.

ITLANG

Language of testing for each student. It is set to "1" for all countries that tested in a single language. For countries that administered the test in more than one language, additional numerical codes are used that correspond to the order of the testing languages as shown in Supplement 2 to the User Guide.

ITPART

Participation status of each respondent and is included in all data files. In the student achievement and background data files and the student-teacher linkage data files, this variable is always set to "3" since all students in the database took part in the PIRLS assessment. In the home, teacher and school background data files, it is set to "1" if the corresponding respondent completed a questionnaire, or is set to "0" otherwise.

2.6 Data Codebook Files

All information related to the structure of the data files, as well as the source, format, descriptive labels, and response option codes for all variables, is contained in codebook files. Each data file is accompanied by a codebook file.

The naming convention for a codebook file is as follows:

- The first three characters of the filename are in every respect identical to the file naming convention presented earlier.
- The next three characters identify the files as PIRLS codebooks and are always "PRL".
- The seventh and eighth characters are always "R2" to indicate the PIRLS 2006 study cycle.
- The three-character file extension is always .SDB, which stands for standard dBase format.

Codebook files can be read using Excel, or any standard database or spreadsheet program. Codebook files also are provided in plain .TXT format and printable .PDF format. They describe the contents and structure of the PIRLS database files. Important codebook fields include FIELD_LABL, which contains extended textual information for all variables, QUEST_LOC, which provides the location of questions and achievement items within their respective survey instruments, and FIELD_CODE, which lists all acceptable responses allowed in the database.

References

Barth, J. Kennedy, A. M. & Trong, K.L. (2007). PIRLS survey operations procedures. In M.O. Martin, I.V.S. Mullis, and A.M. Kennedy (Eds.), *PIRLS 2006 technical report* (pp. 61-72). Chestnut Hill, MA: Boston College.

Foy, P., Galia, J., and Li, I. (2007). Scaling methods and procedures for the PIRLS 2006 reading achievement scales. In M.O. Martin, I.V.S. Mullis, and A.M. Kennedy (Eds.), *PIRLS 2006 technical report* (pp. 149-172). Chestnut Hill, MA: Boston College

Joncas, M. (2007). PIRLS 2006 Sampling weights and participation rates. In M.O. Martin, I.V.S. Mullis, and A.M. Kennedy (Eds.), *PIRLS 2006 technical report* (pp. 105-130). Chestnut Hill, MA: Boston College.

Martin, M.O., Mullis, I.V.S., & Kennedy, A.M. (Eds.). (2007). *PIRLS 2006 technical report.* Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M., & Foy, P. (2007). *PIRLS 2006 international report: IEA's progress in international reading literacy study in primary schools in 40 countries*. Chestnut Hill, MA: Boston College.

Chapter 3

Using the IEA's IDB Analyzer to Analyze the PIRLS 2006 International Database

3.1 Overview

This chapter describes the use of the IEA's International Database (IDB) Analyzer (IEA, 2007) for analyzing the PIRLS 2006 international data files. Example analyses will illustrate the capabilities of the IDB Analyzer to compute a variety of statistics, including percentages of students in specified subgroups, average reading achievement in those subgroups, regression coefficients, and correlations. The examples use student, home, teacher, and school data to replicate some of the PIRLS 2006 results included in the *PIRLS 2006 International Report* (Mullis, Martin, Kennedy, & Foy, 2007) as well as other useful analyses for investigating policy-relevant research questions.

This chapter will focus on the SPSS files in the DATA folder, as well as the SPSS syntax files within the IDB Recode Programs folder (syntax_asbgallr2.sps, syntax_atgmerged.sps, syntax_ashmerged.sps). These are the programs that will be used to recode variables used in the example analyses presented later in this chapter. Other available SPSS program files are not relevant to this chapter and will be discussed in detail in Chapter 4.

3.2 Overview of the IEA's IDB Analyzer

Developed by the IEA Data Processing and Research Center, The IEA's International Database Analyzer (IDB Analyzer) is a plug-in for the Statistical Package for the Social Sciences (SPSS) (SPSS, 2005). The IDB Analyzer enables the user to combine data files from IEA's large-scale assessments and conduct

analyses using SPSS without writing programming code. The IDB Analyzer generates SPSS syntax that takes into account information from the sampling design in the computation of sampling variance. In addition, SPSS code is generated to make use of the plausible values for calculating estimates of reading achievement and their corresponding standard errors, combining both sampling and imputation variance.

The IDB Analyzer consists of two modules, the merge module and the analysis module, which are executed as independent applications. The merge module is used to create analysis datasets by combining data files of different types or from different countries, and selecting subsets of variables for analysis. The analysis module provides procedures for computing means, percentages, standard deviations, correlations, and regression coefficients for variables of interest. These procedures can be applied overall for a country and for specific subgroups within a population. Both modules can be accessed either by the using the START menu in Windows (Start → All Programs → IEA → IDB Analyzer → Merge Data / Analyze Data), or selecting the corresponding menu option from the IEA-IDB Analyzer menu in the SPSS Data window.

3.3 Merging Files with the IDB Analyzer

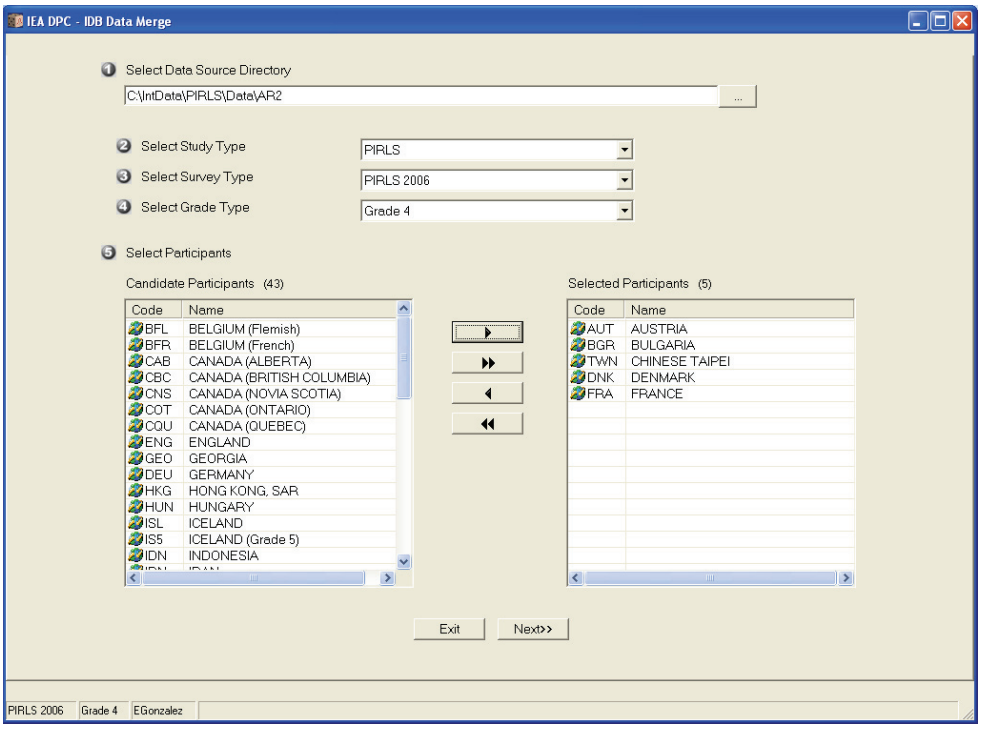
The PIRLS 2006 data files are distributed separately for each country and by file type. In addition to allowing the user to combine like datasets from more than one country for cross-country analyses, the merge module allows for the combination of data from different sources (e.g., student, school, teacher, and home) into one SPSS dataset within a country.

Use the following steps to create an SPSS data file with more than one country's data, or combine more than one file type within a country:

1) Open the merge module of the IDB Analyzer.
2) In the field **Select Data Source Directory**, browse to the path where all country data files are located. For example, in Exhibit 3.1 all country data files are located in the folder "C:\IntData\PIRLS\Data\AR2". The program will automatically recognize and complete the Study Type, Survey Type, and Grade Type field and list all countries available in this path as possible candidates for the merging. If the directory contains data for more than one IEA study, the IDB Analyzer will prompt you to select the desired study files for your analyses.

3) Select the countries of interest from the **Candidate Participants** list. For multiple selections of countries, hold the CTRL key of your keyboard when selecting the countries. In Exhibit 3.1, Austria, Bulgaria, Chinese Taipei, Denmark, and France have been selected.

Exhibit 3.1 IDB Analyzer Merge Module Set-Up: Selecting Countries



4) Press the **Next**>> button to proceed.
The software will take you to the second screen of the merge module, as shown in Exhibit 3.2. Here you will choose the file types for merging and the variables you want included in the merged data file.

5) Select the file types for merging by ticking the corresponding boxes. If you are unsure which file type is which, move the cursor over the file type and text will appear on the screen describing the data files. For example, in Exhibit 3.2 the student background (ASG) data file, is selected.

6) Select the variables you require from the list of background variables and achievement scores. Please note that all identification and sampling variables are selected automatically by the IDB Analyzer.

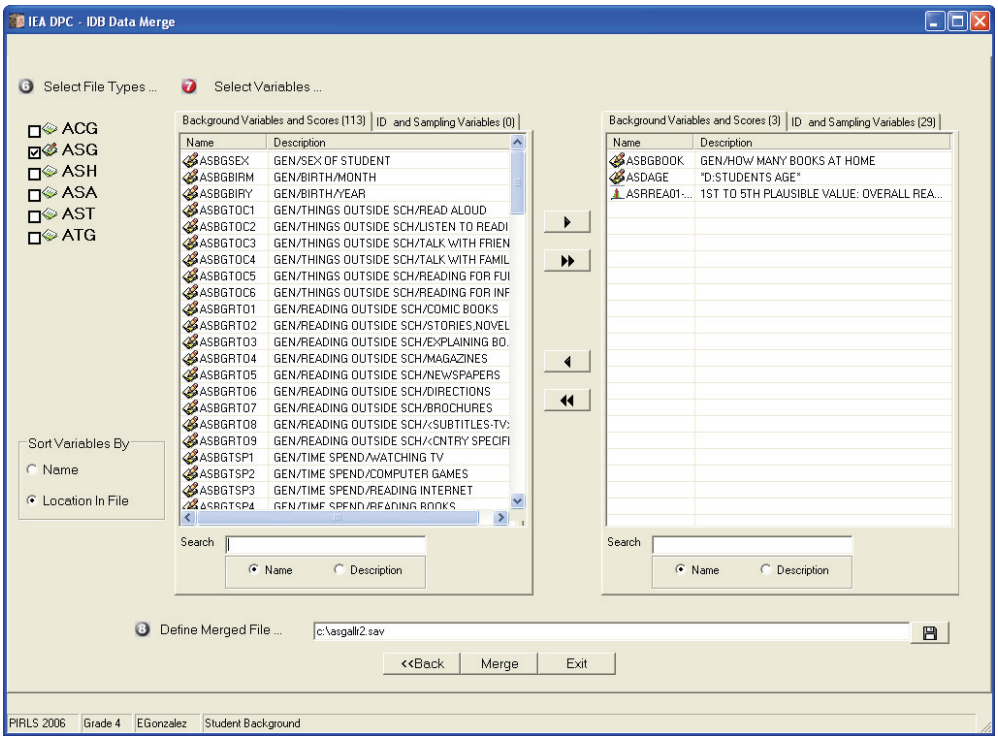
7) Specify the name and the path of your merged data file. To avoid overwriting the original files, save the merged file in a directory other

than where the original files are located and use a different naming convention for the merged file. In the example shown in Exhibit 3.2, the merged file ASGALLR2.SAV will be saved in the root directory and will contain the student background data files for Austria, Bulgaria, Chinese Taipei, Denmark, and France.

8) Click on the **Merge** button to merge the specified data files. The program will alert you as soon as the merge process has been completed. The resulting data file will open automatically in a new SPSS Data Editor window. Simultaneously, the merge procedure generates a syntax file containing the SPSS commands generated by the IDB Analyzer, which can be saved and re-run directly within SPSS, or edited for creating additional merged files.

Please be sure to check the resulting SPSS output file for warnings. If warnings appear, please check them carefully as often they indicate that the merge process was not performed properly and your resulting data file might not be what you expected.

Exhibit 3.2 IDB Analyzer Merge Module Set-Up: Selecting File Types and Variables



Merging Teacher and Student Data Files

When considering analyses with teacher-level variables, it is important to realize that the teachers in the PIRLS database do not constitute representative samples of fourth grade teachers in the participating countries. Rather, they are the teachers of nationally representative samples of fourth grade students. Therefore, analyses with teacher data should be made with students as units of analysis, and reported in terms of how many students are taught by teachers with a particular attribute, and not in terms of how many teachers in a country have a particular attribute.

When analyzing teacher data, it is necessary first to link the students to their teachers. The student-teacher linkage (AST) data files were created to facilitate this linkage. Each student record in a student background data file links to at least one reading teacher in its corresponding teacher background data file. There is usually only one student-teacher link, but occasionally two, and in rare circumstances, three.

The student-teacher linkage (AST) data files contain one record for each student-teacher combination. For example, if a student is taught by two reading teachers, the AST data file has two records for that student, one for each of his or her reading teachers. Each record in the data file holds the number of reading teachers linked to that student.

Student achievement scores (plausible values), jackknife replication information, and a weighting variable, TCHWGT, appropriate for conducting student-level analyses of teacher data, have been included in the AST data files in order to simplify the merging process for analyses that link teacher variables to student achievement. For such analyses, it is only necessary to merge the teacher background (ATG) data files with the student-teacher linkage (AST) data files. For analyses linking teacher variables to student background variables, it is also necessary to merge the student background (ASG) data files with the teacher background (ATG) data files after having been combined with the student-teacher linkage data files.

When merging file types using the IDB Analyzer, the student-teacher linkage (AST) file is automatically selected when you select the teacher background (ATG) data file. Therefore, to merge the teacher and student background data files simply select both the ATG and ASG file types.

You must choose variables of interest to be included in the merged data file separately by file type:

1) Click on the file type name ASG (student background file) so that it appears highlighted. The Background Variables and Scores listed on the left-hand side of the display will include all available variables for the specified file type.
2) Select your variables of interest and press the right arrow button to move these variables into the list on the right-hand side of the display.
3) Next, click on the file type name ATG (teacher background file) as the other file type to be merged, selecting variables from the Background Variables and Scores listed in the same manner.

Merging School and Student Data Files

The PIRLS 2006 school samples were designed to optimize the student samples and the student-level estimates, and for this reason it is preferable to analyze school-level variables as attributes of students, rather than as elements in their own right. However, the school samples are representative probability samples of schools and so it is possible to compute weighted numbers of schools with particular characteristics for providing reasonable estimates of percentages and means across primary schools in each country. To merge the school and student background data files select both the ACG (school background) and ASG (student background) file types. You must choose variables of interest to be included in the merged data file separately by file type. This can be done by following the steps in the previous section.

Merging Home and Student Data Files

The students' parents' or primary caregivers' responses to the *Learning to Read Survey* are included in the home background (ASH) data files. Although home background variables are located in separate files, they are in essence attributes of students and must be analyzed in the same manner as student background variables. This will require us to merge the home background data files with the student background data files by selecting both the ASH (home background) and ASG (student background) file types. Variables of interest to be included in the merged data file must be chosen separately by file type according to the steps in the section on merging teacher and student data files.

Merging Data Files for Chapter Examples

To create the data files used in the examples that follow in this chapter, you will need to create the following merged files:

ASGALLR2.SAV	Merge the student background (ASG) data files for all countries
ATGMERGED.SAV	Merge the teacher (ATG) and student (ASG) background data files for all countries
ACGMERGED.SAV	Merge the school (ACG) and student (ASG) background data files for all countries
ASHMERGED.SAV	Merge the home (ASH) and student (ASG) background data files for all countries

3.4 Conducting Analyses with the IDB Analyzer

The analysis module of the IDB Analyzer is used to analyze the country data files, as well as any files that have been created using the merge module. The analysis module can perform the following procedures:

Percentages and Means

Computes percentages, means, and standard deviations for selected variables, by subgroups defined by grouping variable(s)

Percentages Only

Computes percentages, by subgroups defined by grouping variable(s)

Regression

Computes regression coefficients for selected variables predicting a dependent variable, by subgroups defined by grouping variable(s)

Benchmarks

Computes percentages of students meeting a set of user-specified achievement benchmarks, by subgroups defined by grouping variable(s)

Correlations

Computes means, standard deviations, and correlation coefficients for selected variables, by subgroups defined by grouping variable(s)

All procedures offered within the analysis module of the IDB Analyzer make use of appropriate sampling weights and standard errors are computed using the jackknife repeated replicate method (JRR) (see Joncas, 2007). Percentages and means, and the regression and correlation analyses may be specified with or without achievement scores. To conduct an analysis using the achievement scores, select the option **With Achievement Scores** under the section **Analysis Type**. When achievement scores are used, the analyses are conducted using all five plausible values and the standard errors are calculated accordingly.

The IDB Analyzer requires the selection of several variables:

Grouping Variables

A list of variables that define the subgroups. The list must consist of at least one grouping variable. By default, the IDB Analyzer includes IDCNTRY. Additional variables may be selected from the available list. If the option **Exclude Missing from Analysis** is checked, only cases that have non-missing values in the grouping variables will be used in the analysis.

Analysis Variables

This is the list of variables for which means or percentages are to be computed or the independent variables for a regression analysis. You may select more than one analysis variable.

To compute means for the achievement scores, you must check the option **With Achievement Scores** under **Analysis Type** and select the **Achievement Scores** variable.

Achievement Scores

This section is used to identify the set of plausible values to be used when an achievement score is the analysis variable for computing percentages and means or when an achievement score is the dependent variable in a regression analysis.

Dependent Variable

This is the variable to be used as the dependent variable when a regression analysis is specified. Only one dependent variable can be listed. To use achievement scores as the dependent variable, you must check the option **With Achievement Scores** under **Analysis Type** and select the **Achievement Scores** variable.

Benchmarks

These are the values that will be used as cut points of the achievement distribution for computing the percentages of students meeting the specified benchmarks.

Weight Variable

The sampling weight that will be used in the analysis. The IDB Analyzer automatically selects the appropriate weight variable for the analysis. Generally, this will be TOTWGT, although TCHWGT will be used when analyzing teacher data files.

Jackknifing Variables

Variables that capture the assignment of cases to sampling zones (JKZONE) and whether the case is to be dropped or have its weight doubled when computing the set of replicate weights (JKREP). The IDB Analyzer automatically uses these variables to compute the 75 replicate weights that are used in all analysis types and cannot be changed.

3.5 Performing Analyses with Student-level Variables

Many analyses of the PIRLS data can be undertaken using student-level data only. This section presents examples of actual analyses used to produce exhibits from the *PIRLS 2006 International Report*. Examples of regression analyses and computing percentages of students reaching benchmarks also are included in this section.

Student-Level Analysis Without Achievement Scores

In our first example, we will replicate an analysis of students' reports on the number of hours spent reading stories or articles in books or magazines outside

of school. The results, presented in Exhibit 4.5 of the *PIRLS 2006 International Report*, are reproduced here in Exhibit 3.3. This example will focus on the results presented in the fourth data column—the average number of hours spent reading stories or articles in books or magazines outside of school, overall. Since we want to report the average number of hours (with corresponding standard errors), we will be computing means without achievement scores.

We need to undertake a number of steps to replicate the results in this exhibit. After reviewing the codebooks and questionnaire information, we identify the student background variable ASBGTSP4 as a categorical variable measuring the number of hours spent reading stories or articles in books or magazines outside of school (see Supplement 1 for a copy of the *Student Questionnaire*). Our next step is to review the documentation of national adaptations to the questionnaire to ensure that no countries made changes to this variable (see Supplement 2). If no national adaptations were made, we can continue with our analysis without any modifications.

Before proceeding with the analysis module of the IDB Analyzer, we must first recode the variable ASBGTSP4 so that the values represent the mid-points of the intervals they cover. To this effect, the value 1 is recoded to 5 hours, the value 2 to 4 hours, the value 3 to 2 hours, the value 4 to 0.5 hours and the value 5 to zero hours. By doing these recodes we make it easier to interpret the results from our analysis.

The SPSS syntax file for recoding ASBGTSP4 for this example is in the IDB Recode Programs folder on the DVD. The syntax is also provided in Exhibit 3.4, where the new values are assigned to the variable NEWTSP4.

Exhibit 3.3 Example Exhibit of Student-Level Analysis Without Achievement Scores Taken From the PIRLS 2006 International Report (Exhibit 4.5)

Exhibit 4.5 **Students Read Stories or Articles Outside of School** — **PIRLS 2006** 4th Grade

Countries	Average Number of Hours on a Typical Day Spent Reading								
	Stories or Articles on the Internet				Stories or Articles in Books or Magazines				
	Overall	Girls	Boys		Overall	Girls	Boys		
Austria	0.8 (0.02)	0.8 (0.03)	0.9 (0.03)	▲	1.6 (0.03)	1.8 (0.04)	▲	1.3 (0.04)	
Belgium (Flemish)	0.6 (0.03)	0.5 (0.03)	0.6 (0.04)	▲	0.6 (0.02)	0.7 (0.03)	▲	0.5 (0.02)	
Belgium (French)	1.2 (0.04)	1.2 (0.05)	1.3 (0.05)		1.2 (0.03)	1.3 (0.04)	▲	1.1 (0.04)	
Bulgaria	1.0 (0.05)	0.9 (0.05)	1.1 (0.06)	▲	1.5 (0.05)	1.6 (0.07)	▲	1.3 (0.05)	
Canada, Alberta	0.8 (0.03)	0.8 (0.04)	0.9 (0.04)		1.3 (0.03)	1.4 (0.04)	▲	1.2 (0.04)	
Canada, British Columbia	0.8 (0.03)	0.7 (0.04)	0.8 (0.04)	▲	1.3 (0.04)	1.4 (0.05)	▲	1.2 (0.05)	
Canada, Nova Scotia	0.9 (0.03)	0.9 (0.04)	1.0 (0.04)	▲	1.3 (0.03)	1.5 (0.04)	▲	1.2 (0.04)	
Canada, Ontario	0.9 (0.04)	0.8 (0.05)	0.9 (0.04)		1.4 (0.05)	1.5 (0.06)	▲	1.3 (0.06)	
Canada, Quebec	1.0 (0.04)	1.0 (0.04)	1.1 (0.06)		1.3 (0.04)	1.5 (0.06)	▲	1.2 (0.04)	
Chinese Taipei	1.0 (0.03)	1.1 (0.04)	▲ 0.9 (0.03)		1.2 (0.03)	1.4 (0.04)	▲	1.1 (0.04)	
Denmark	0.6 (0.03)	0.5 (0.03)	0.7 (0.04)	▲	1.0 (0.03)	1.1 (0.04)	▲	0.9 (0.04)	
England	0.9 (0.03)	0.8 (0.04)	0.9 (0.04)		1.2 (0.03)	1.4 (0.05)	▲	1.1 (0.04)	
France	0.9 (0.03)	0.9 (0.04)	0.9 (0.05)		1.2 (0.03)	1.2 (0.04)	▲	1.1 (0.04)	
Georgia	0.9 (0.05)	0.8 (0.06)	0.9 (0.06)		1.5 (0.05)	1.6 (0.07)	▲	1.4 (0.06)	
Germany	0.6 (0.02)	0.6 (0.03)	0.6 (0.03)	▲	1.5 (0.03)	1.7 (0.04)	▲	1.4 (0.05)	
Hong Kong SAR	1.1 (0.03)	1.1 (0.04)	1.0 (0.04)		1.0 (0.03)	1.1 (0.04)	▲	1.0 (0.04)	
Hungary	0.7 (0.03)	0.7 (0.04)	0.7 (0.03)		1.3 (0.04)	1.4 (0.06)	▲	1.1 (0.04)	
Iceland	0.6 (0.02)	0.6 (0.02)	0.7 (0.03)	▲	0.8 (0.02)	0.9 (0.03)	▲	0.7 (0.03)	
Indonesia	1.3 (0.05)	1.2 (0.05)	1.3 (0.06)		1.6 (0.04)	1.7 (0.06)		1.6 (0.05)	
Iran, Islamic Rep. of	0.3 (0.03)	0.3 (0.05)	0.4 (0.05)		1.5 (0.05)	1.5 (0.06)		1.4 (0.08)	
Israel	1.5 (0.04)	1.4 (0.05)	1.5 (0.05)	▲	1.4 (0.04)	1.6 (0.05)	▲	1.3 (0.04)	
Italy	0.7 (0.03)	0.6 (0.04)	0.8 (0.04)	▲	1.3 (0.04)	1.4 (0.05)	▲	1.1 (0.05)	
Kuwait	2.1 (0.06)	2.0 (0.07)	2.1 (0.08)		2.1 (0.05)	2.2 (0.07)		2.1 (0.07)	
Latvia	1.0 (0.04)	0.9 (0.04)	1.1 (0.05)	▲	1.2 (0.03)	1.4 (0.05)	▲	1.0 (0.04)	
Lithuania	0.9 (0.03)	0.9 (0.04)	1.0 (0.03)		1.4 (0.03)	1.7 (0.04)	▲	1.2 (0.04)	
Luxembourg	0.5 (0.01)	0.5 (0.02)	0.6 (0.02)	▲	0.9 (0.02)	1.0 (0.02)	▲	0.8 (0.03)	
Macedonia, Rep. of	1.8 (0.08)	1.7 (0.09)	1.8 (0.08)	▲	2.6 (0.07)	2.7 (0.08)	▲	2.4 (0.07)	
Moldova, Rep. of	1.0 (0.06)	0.9 (0.08)	1.0 (0.06)		1.8 (0.05)	1.9 (0.07)	▲	1.7 (0.06)	
Morocco	1.3 (0.08)	1.3 (0.09)	1.4 (0.08)		1.3 (0.07)	1.4 (0.09)		1.3 (0.08)	
Netherlands	0.5 (0.02)	0.5 (0.02)	0.5 (0.03)		0.8 (0.02)	0.9 (0.04)	▲	0.6 (0.03)	
New Zealand	0.9 (0.03)	0.9 (0.04)	1.0 (0.04)	▲	1.4 (0.04)	1.6 (0.05)	▲	1.3 (0.04)	
Norway	0.6 (0.03)	0.5 (0.04)	0.6 (0.03)		0.9 (0.04)	0.9 (0.05)	▲	0.8 (0.05)	
Poland	0.9 (0.03)	0.8 (0.03)	1.0 (0.04)	▲	1.5 (0.03)	1.7 (0.04)	▲	1.3 (0.05)	
Qatar	2.3 (0.03)	2.2 (0.04)	2.4 (0.04)	▲	2.2 (0.03)	2.3 (0.04)	▲	2.1 (0.04)	
Romania	0.9 (0.06)	0.8 (0.06)	1.0 (0.07)	▲	1.6 (0.05)	1.8 (0.07)	▲	1.5 (0.06)	
Russian Federation	0.5 (0.03)	0.4 (0.02)	0.6 (0.04)	▲	1.5 (0.04)	1.6 (0.05)	▲	1.3 (0.04)	
Scotland	0.9 (0.04)	0.9 (0.05)	0.9 (0.05)		1.2 (0.03)	1.4 (0.05)	▲	1.1 (0.05)	
Singapore	1.1 (0.03)	1.1 (0.03)	1.0 (0.04)		1.4 (0.02)	1.6 (0.04)	▲	1.2 (0.03)	
Slovak Republic	0.7 (0.03)	0.7 (0.04)	0.8 (0.04)	▲	1.5 (0.04)	1.7 (0.05)	▲	1.3 (0.05)	
Slovenia	0.7 (0.03)	0.6 (0.03)	0.8 (0.04)	▲	1.0 (0.02)	1.1 (0.03)	▲	0.9 (0.03)	
South Africa	2.1 (0.07)	2.1 (0.07)	2.1 (0.07)		2.7 (0.06)	2.7 (0.06)	▲	2.6 (0.07)	
Spain	0.9 (0.03)	0.8 (0.04)	1.0 (0.05)	▲	1.2 (0.03)	1.2 (0.05)		1.2 (0.05)	
Sweden	0.5 (0.02)	0.4 (0.02)	0.6 (0.04)	▲	0.7 (0.02)	0.7 (0.03)	▲	0.6 (0.03)	
Trinidad and Tobago	1.5 (0.07)	1.5 (0.09)	1.5 (0.07)		1.7 (0.06)	1.9 (0.08)	▲	1.6 (0.07)	
United States	1.0 (0.05)	1.1 (0.06)	1.0 (0.05)		1.4 (0.04)	1.6 (0.04)	▲	1.2 (0.06)	
International Avg.	1.0 (0.01)	0.9 (0.01)	1.0 (0.01)	▲	1.4 (0.01)	1.5 (0.01)	▲	1.3 (0.01)	

▲ Average significantly higher than other gender

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

Exhibit 3.4 Example SPSS Program to Recode Variables for Analysis

```
get file = "<datpath>asgallr2.sav".

* Compute new variable newTSP4 from ASBGTSP4 .
Recode asbgtsp4 (1=5.5) (2=4) (3=2) (4=0.5) (5=0) into newtsp4.

variable labels
    newtsp4 "Recoded time spent reading books".

save outfile = "<datpath>asgallr2.sav".
```

After creating the data file for the analysis, and performing the necessary recodes, we then use the IDB Analyzer analysis module to conduct our analysis in the following steps:

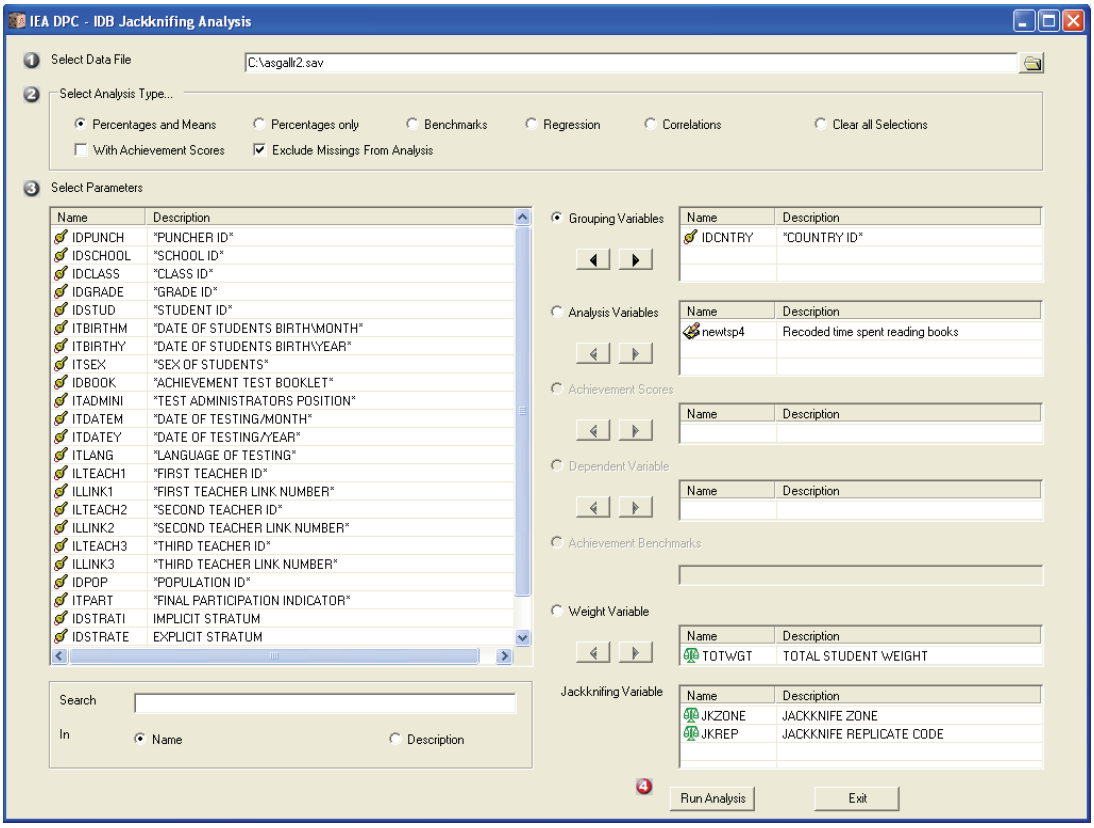
1) Open the analysis module of the IDB Analyzer.
2) Select the data file (the merged file ASGALLR2).
3) As the **Analysis Type** select **Percentages and Means**. Note that there are two options to check: **With Achievement Scores** and **Exclude Missing from Analysis**. Since no achievement scores are used in this analysis, only **Exclude Missing from Analysis** should be checked. This option, indicating whether to exclude cases that have missing values in the grouping variable(s), is checked by default.
4) Define the variables needed for the analysis:
    a. Within **Grouping Variables**, you will need to add ITSEX for this example. To do this, select the variable from the variable list on the left hand side of the display and press the right arrow button belonging to the section of the grouping variables to move the variable.
    b. Next, define the analysis variables. To activate this section, press the radio button **Analysis Variables**. For the example, you will need to select NEWTSP4 from the list of variables and move it to the analysis variables field by pressing the right arrow button in this section.

5) The **Weight Variable** is automatically defined by the software. As the example analysis is at the student level, TOTWGT is selected by default. Additionally, the **Jackknifing Variables** JKZONE and JKREP are selected.

6) Press the button **Run Analysis** in order to complete the procedure. The program will alert you as soon as the analysis has been completed and the results will be displayed in the output window.

Exhibit 3.5 presents the completed Analysis Module for the example. The results are displayed in Exhibit 3.6, though a limited number of countries are displayed to conserve space. This is done for all output exhibits.

Exhibit 3.5 Example IDB Analyzer Setup for Student-Level Analysis Without Plausible Values



In this example, each country's mean value for NEWTSP4 variable is reported for all sampled students. The countries are identified in the first column and the second column reports the number of valid cases. The third column reports the sum of weights of the sampled students, followed by the percent, mean, and

standard deviation, each accompanied by its jackknife standard error. From the first line shown in Exhibit 3.6, we see that in Austria valid data were available for 4,967 students and these sampled students represent a population of 81,526 students. Austrian students spend, on average, 1.57 hours reading stories or articles in books or magazines outside of school (the mean of NEWTSP4), with a standard error of 0.03. For this example, the percents are of little use as they are simply the proportion each country represents among all participating countries.

Exhibit 3.6 Output for Example Student-Level Analysis Without Achievement Scores

Average for NEWTSP4 by (IDCNTRY) PAGE 1

COUNTRY ID	N of Cases	Sum of TOTWGT	Percent	Percent (s.e.)	NEWTSP4 (Mean)	NEWTSP4 (s.e.)	Std.Dev.	Std.Dev. (s.e.)
Austria	4967	81526	.50	.02	1.57	.03	1.66	.02
Bulgaria	3739	61273	.38	.01	1.46	.05	1.69	.03
Chinese Taipei	4462	295892	1.82	.03	1.24	.03	1.51	.02
				...				
South Africa	13092	744365	4.59	.07	2.67	.06	2.18	.01
Spain	4026	384242	2.37	.05	1.19	.03	1.65	.03
Sweden	4306	100097	.62	.02	.68	.02	1.13	.04
Trinidad & Tobago	3748	16249	.10	.00	1.74	.06	1.98	.03
Macedonia	3647	20817	.13	.00	2.58	.07	2.02	.02
United States	4922	3187176	19.63	.45	1.40	.04	1.80	.03

Student-level Analysis with Achievement Scores

In our second example, we want to replicate another set of results presented in the *PIRLS 2006 International Report*. We are interested in investigating the relationship between the frequency of students reading for fun outside of school and reading achievement. These results, presented in Exhibit 4.6 of the *PIRLS 2006 International Report*, are repeated here in Exhibit 3.7. Since the results in this exhibit are based on plausible values, we need to make sure we include these when we create the file using the merge module, and also to indicate that our analysis will make use of achievement scores when we specify the analysis type.

After reviewing the codebooks and the questionnaire information, we observe that the variable ASBGTOC5 contains categorical information on the frequency students read for fun outside of school, and this variable is found in the student background data files. Our next step is to review the documentation of national adaptations to the questionnaires to ensure that there were no deviations listed for this variable (see Supplement 2). If no national adaptations were made, we can continue with our analysis without any modifications.

Exhibit 3.7 Example Exhibit of Student-Level Analysis with Achievement Scores Taken From the PIRLS 2006 International Report (Exhibit 4.6)

Exhibit 4.6 **Students Reading for Fun Outside of School with Trends** — **PIRLS 2006** 4th Grade

Countries	Every Day or Almost Every Day			Once or Twice a Week			Twice a Month or Less		
	2006 Percent of Students	Average Achievement	Difference in Percent from 2001	2006 Percent of Students	Average Achievement	Difference in Percent from 2001	2006 Percent of Students	Average Achievement	Difference in Percent from 2001
Russian Federation	58 (1.1)	570 (3.8)	0 (1.8)	28 (0.8)	559 (3.9)	–1 (1.3)	14 (0.8)	556 (3.9)	2 (1.3)
Canada, Alberta	53 (0.9)	575 (2.5)	◊ ◊	23 (0.8)	555 (3.5)	◊ ◊	23 (0.9)	537 (2.7)	◊ ◊
Canada, British Columbia	53 (1.0)	573 (2.9)	◊ ◊	26 (0.8)	554 (2.5)	◊ ◊	21 (0.9)	531 (4.1)	◊ ◊
Germany	53 (0.9)	563 (2.7)	5 (1.2) ▲	24 (0.6)	545 (3.0)	0 (0.9)	24 (0.8)	525 (2.5)	–5 (1.2) ▼
Lithuania	52 (1.2)	545 (2.1)	–1 (1.9)	30 (1.0)	533 (2.2)	–1 (1.6)	17 (0.8)	520 (2.8)	2 (1.2)
Moldova, Rep. of	52 (1.4)	507 (3.2)	2 (2.6)	34 (1.1)	498 (4.2)	0 (1.9)	14 (1.0)	484 (5.1)	–2 (1.6)
France	51 (1.0)	540 (2.5)	2 (1.6)	24 (0.8)	517 (2.3)	–2 (1.2)	25 (0.9)	491 (2.7)	0 (1.3)
Canada, Ontario	49 (1.4)	567 (3.2)	14 (2.0) ▲	25 (1.1)	552 (3.6)	2 (1.4)	26 (1.1)	534 (4.1)	–16 (1.9) ▼
Belgium (French)	49 (1.1)	517 (3.0)	◊ ◊	26 (0.7)	495 (2.9)	◊ ◊	25 (0.9)	473 (3.1)	◊ ◊
Iceland	49 (0.9)	527 (1.9)	–3 (1.2) ▼	23 (0.7)	511 (2.6)	2 (1.0)	28 (0.7)	485 (2.3)	1 (1.0)
Denmark	49 (1.1)	559 (2.9)	◊ ◊	30 (0.8)	540 (2.7)	◊ ◊	21 (0.9)	528 (3.2)	◊ ◊
Canada, Nova Scotia	48 (0.9)	560 (2.6)	◊ ◊	25 (0.8)	541 (2.7)	◊ ◊	27 (0.8)	515 (3.2)	◊ ◊
Bulgaria	47 (1.6)	561 (4.4)	–4 (2.3)	27 (1.0)	555 (5.0)	–2 (1.5)	26 (1.6)	520 (6.5)	5 (2.2) ▲
Canada, Quebec	47 (1.3)	549 (3.0)	1 (1.9)	26 (1.0)	530 (3.8)	1 (1.4)	27 (1.2)	509 (3.2)	–3 (1.7)
South Africa	45 (0.9)	303 (6.4)	◊ ◊	26 (0.5)	314 (6.4)	◊ ◊	28 (0.8)	307 (6.4)	◊ ◊
Austria	45 (1.1)	555 (3.0)	◊ ◊	25 (0.8)	535 (2.4)	◊ ◊	29 (1.0)	516 (2.9)	◊ ◊
Spain	45 (1.1)	525 (2.9)	◊ ◊	27 (0.7)	515 (2.5)	◊ ◊	28 (1.0)	494 (3.7)	◊ ◊
Hungary	44 (1.2)	565 (3.7)	4 (1.7) ▲	30 (0.9)	547 (3.4)	–2 (1.4)	26 (1.1)	532 (4.2)	–2 (1.5)
Indonesia	44 (1.4)	405 (4.7)	◊ ◊	31 (1.1)	414 (4.6)	◊ ◊	25 (1.2)	403 (4.8)	◊ ◊
Macedonia, Rep. of	43 (1.2)	453 (5.7)	–3 (1.9)	31 (1.0)	451 (4.6)	3 (1.4) ▲	25 (1.1)	435 (5.1)	0 (1.8)
Poland	43 (1.3)	538 (2.5)	◊ ◊	29 (1.0)	518 (3.2)	◊ ◊	27 (1.0)	495 (3.0)	◊ ◊
New Zealand	42 (1.1)	562 (2.4)	–1 (1.8)	24 (0.7)	531 (2.5)	0 (1.1)	34 (1.0)	500 (3.0)	1 (1.6)
Belgium (Flemish)	40 (1.1)	563 (2.1)	◊ ◊	29 (0.8)	545 (2.9)	◊ ◊	31 (1.2)	529 (2.3)	◊ ◊
Slovak Republic	39 (1.0)	545 (2.9)	0 (1.5)	33 (0.9)	535 (3.2)	0 (1.3)	27 (1.1)	507 (5.4)	0 (1.6)
Italy	38 (1.3)	573 (3.3)	7 (1.7) ▲	25 (0.7)	554 (3.2)	1 (1.0)	37 (1.3)	529 (3.8)	–7 (1.7) ▼
Luxembourg	38 (0.6)	581 (1.8)	◊ ◊	27 (0.7)	551 (2.0)	◊ ◊	35 (0.5)	537 (1.5)	◊ ◊
Israel	38 (1.2)	538 (4.2)	–6 (1.6) ▼	28 (0.9)	518 (4.4)	1 (1.2)	35 (1.1)	497 (4.1)	5 (1.5) ▲
Slovenia	37 (0.9)	543 (2.5)	–8 (1.6) ▼	33 (0.7)	519 (3.0)	4 (1.1) ▲	30 (0.9)	500 (2.6)	4 (1.5) ▲
Netherlands	36 (1.1)	566 (2.1)	0 (1.6)	22 (0.7)	550 (1.8)	2 (1.1) ▲	42 (1.1)	530 (1.8)	–2 (1.6)
Sweden	36 (1.0)	569 (2.8)	–8 (1.3) ▼	31 (0.9)	549 (3.2)	–1 (1.2)	33 (1.0)	530 (2.6)	9 (1.3) ▲
United States	35 (1.3)	561 (4.3)	1 (1.8)	22 (0.7)	550 (3.3)	0 (1.3)	43 (1.4)	521 (3.3)	–1 (1.8)
Latvia	35 (1.2)	556 (3.0)	–8 (1.7) ▼	31 (0.8)	543 (2.8)	–2 (1.2)	34 (1.2)	524 (2.6)	10 (1.7) ▲
Hong Kong SAR	35 (1.0)	575 (2.6)	14 (1.3) ▲	33 (0.9)	567 (2.7)	–5 (1.2) ▼	32 (1.0)	549 (2.8)	–8 (1.4) ▼
Iran, Islamic Rep. of	33 (1.2)	428 (4.2)	–1 (1.7)	41 (1.2)	429 (3.9)	2 (1.9)	26 (1.0)	406 (5.2)	0 (1.5)
Norway	33 (1.1)	514 (3.4)	–5 (1.5) ▼	30 (1.0)	505 (3.2)	2 (1.3)	37 (1.2)	481 (3.1)	4 (1.7) ▲
Qatar	33 (0.6)	357 (2.2)	◊ ◊	28 (0.5)	367 (2.5)	◊ ◊	39 (0.6)	352 (2.2)	◊ ◊
Scotland	33 (1.1)	555 (4.4)	2 (1.6)	24 (1.0)	533 (3.1)	0 (1.3)	44 (1.5)	505 (2.7)	–2 (2.2)
England	33 (1.2)	575 (4.0)	0 (1.8)	25 (0.8)	537 (3.5)	–1 (1.2)	42 (1.3)	517 (2.9)	1 (2.0)
Kuwait	32 (1.1)	338 (5.5)	◊ ◊	32 (1.0)	342 (5.5)	◊ ◊	36 (1.2)	332 (5.4)	◊ ◊
Trinidad and Tobago	32 (1.2)	450 (6.7)	◊ ◊	25 (1.0)	442 (5.9)	◊ ◊	43 (1.4)	427 (5.6)	◊ ◊
Georgia	29 (1.4)	479 (4.5)	◊ ◊	29 (1.2)	484 (4.0)	◊ ◊	41 (1.4)	461 (4.3)	◊ ◊
Morocco	29 (1.3)	317 (8.2)	–3 (2.3)	34 (1.4)	326 (6.9)	–3 (2.0)	37 (1.9)	331 (7.1)	6 (3.1)
Singapore	27 (0.9)	587 (3.9)	–3 (1.5) ▼	26 (0.6)	564 (3.1)	3 (0.9) ▲	47 (1.0)	540 (2.7)	0 (1.5)
Romania	25 (1.3)	510 (5.5)	–3 (2.2)	26 (1.1)	502 (5.5)	–4 (1.6) ▼	50 (1.6)	478 (5.5)	7 (2.3) ▲
Chinese Taipei	24 (0.7)	553 (2.6)	◊ ◊	31 (0.8)	539 (2.6)	◊ ◊	45 (1.0)	525 (2.2)	◊ ◊
International Avg.	40 (0.2)	516 (0.6)		28 (0.1)	503 (0.6)		32 (0.2)	484 (0.6)	

▲ Percent in 2006 significantly higher

▼ Percent in 2006 significantly lower

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

Before proceeding with the analysis module of the IDB Analyzer, we must first recode the variable ASBGTOC5 in the merged student data file (ASGALLR2.SAV) into a new variable called NEWTOC5. The recode will combine categories 3 and 4 from the original variable into the value 3 to create the category "Twice a Month or Less". Exhibit 3.8 shows an example of the SPSS code used to perform these recodes.

Exhibit 3.8 Example SPSS Program to Recode Variables for Analysis

```
get file = "<datpath>asgallr2.sav".

* Compute new variable newTOC5 from ASBGTOC5 .
Recode asbgtoc5 (1=1) (2=2) (3,4=3) (else=sysmis) into newTOC5.

variable labels
    newTOC5 "Recoded ASBGTOC5 Read for fun outside school".

value labels newTOC5
     1 'EVERY DAY OR ALMOST EVERY DAY'
     2 'ONCE OR TWICE A WEEK'
     3 'TWICE A MONTH OR LESS' .

save outfile = "<datpath>asgallr2.sav".
```

Once the recode is complete, we can begin our analysis. The analysis type **Percentages and Means** with activation of the checkbox **With Achievement Scores** computes percentages and mean achievement scores based on plausible values with their corresponding standard errors. Once you have opened the analysis module and selected the ASGALLR2 data file, the steps in the IDB Analyzer are as follows:

1) Select **Percentages and Means** under **Analysis Type**. By default, the program will exclude those with missing grouping variables from the analysis.
2) Check the box next to **With Achievement Scores**.
3) Add the variable NEWTOC5 as a second **Grouping Variable**.

4) Next, define the achievement scores to be used for the analysis. To activate this section, you will need to press the radio button **Achievement Scores**. This time you will need to select variable ASRREA01-05 from the list of variables and move it to the analysis variables field by pressing the right arrow button in this section.

5) The **Weight Variable** is automatically defined by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. Additionally the **Jackknifing Variables** JKZONE and JKREP are selected.

6) Press the button **Run Analysis** in order to perform your analysis. The program will alert you as soon as the analysis has been completed and the results will be displayed in the output window.

Exhibit 3.9 below displays the analysis module with the proper selections. The output of this set-up is shown in Exhibit 3.10.

Exhibit 3.9 Example IDB Analyzer Setup for Student-Level Analysis with Achievement Scores

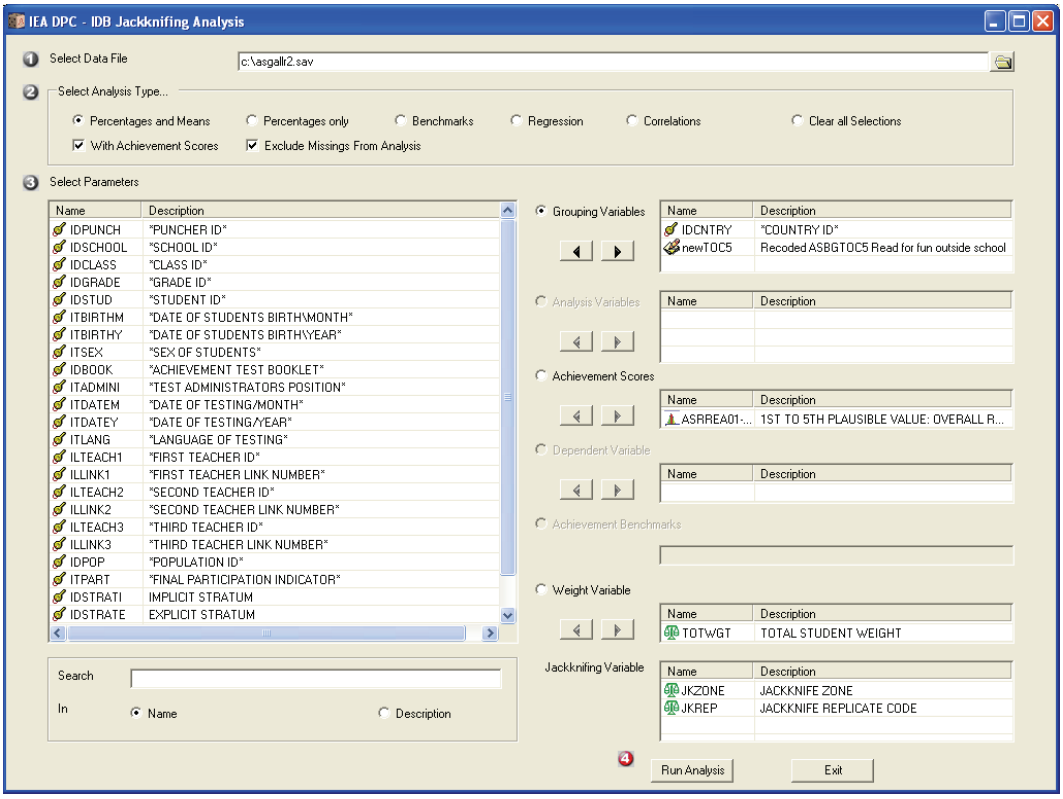


Exhibit 3.10 Output for Example Student-Level Analysis with Achievement Scores

COUNTRY ID	Recoded ASBGTOC5 Read for fun outside school	N of Cases	Sum of TOTWGT	Percent	Percent (s.e.)	ASRREA0 (Mean)	ASRREA0 (s.e.)	Std.Dev.	Std.Dev. (s.e.)
Austria	EVERY DAY OR ALMOST EVERY DAY	2281	36917	45.16	1.10	555.44	3.00	64.27	1.87
	ONCE OR TWICE A WEEK	1255	20821	25.47	.80	535.13	2.43	58.91	1.70
	TWICE A MONTH OR LESS	1445	24016	29.38	.98	515.74	2.88	58.79	1.84
Bulgaria	EVERY DAY OR ALMOST EVERY DAY	1796	29070	47.24	1.59	560.61	4.37	81.02	3.36
	ONCE OR TWICE A WEEK	1033	16313	26.51	1.04	555.34	5.00	74.98	3.19
	TWICE A MONTH OR LESS	928	16154	26.25	1.56	520.18	6.50	82.53	3.00
Chinese Taipei	EVERY DAY OR ALMOST EVERY DAY	1088	71759	24.10	.73	553.02	2.58	63.48	1.95
	ONCE OR TWICE A WEEK	1401	93060	31.25	.78	539.16	2.57	60.87	1.36
	TWICE A MONTH OR LESS	2000	132983	44.65	1.00	525.47	2.23	63.21	1.22
				...					
South Africa	EVERY DAY OR ALMOST EVERY DAY	6161	345653	45.50	.91	303.08	6.41	138.87	4.52
	ONCE OR TWICE A WEEK	3443	199247	26.23	.53	314.45	6.35	138.14	4.02
	TWICE A MONTH OR LESS	3783	214845	28.28	.75	306.94	6.44	131.88	3.64
Spain	EVERY DAY OR ALMOST EVERY DAY	1801	170563	44.77	1.15	525.03	2.88	71.25	2.33
	ONCE OR TWICE A WEEK	1078	102245	26.84	.70	514.71	2.45	65.40	1.70
	TWICE A MONTH OR LESS	1117	108198	28.40	1.01	493.87	3.69	70.25	2.54
Sweden	EVERY DAY OR ALMOST EVERY DAY	1572	36017	35.97	1.02	568.61	2.81	62.98	1.78
	ONCE OR TWICE A WEEK	1333	30843	30.80	.86	549.35	3.22	60.90	1.96
	TWICE A MONTH OR LESS	1401	33281	33.23	.96	530.13	2.56	60.15	1.87
Trinidad & Tobago	EVERY DAY OR ALMOST EVERY DAY	1215	5218	31.70	1.17	449.84	6.74	109.04	3.94
	ONCE OR TWICE A WEEK	956	4170	25.33	.95	441.71	5.89	99.18	3.90
	TWICE A MONTH OR LESS	1619	7072	42.97	1.43	426.74	5.58	98.55	2.93
Macedonia	EVERY DAY OR ALMOST EVERY DAY	1576	9077	43.32	1.17	452.96	5.69	103.87	2.73
	ONCE OR TWICE A WEEK	1156	6552	31.27	1.00	451.30	4.56	99.20	2.90
	TWICE A MONTH OR LESS	933	5324	25.41	1.07	434.69	5.11	95.04	2.57
United States	EVERY DAY OR ALMOST EVERY DAY	1717	1131176	35.28	1.25	560.82	4.26	76.84	2.38
	ONCE OR TWICE A WEEK	1095	705135	21.99	.74	549.57	3.35	68.48	1.91
	TWICE A MONTH OR LESS	2139	1369921	42.73	1.43	521.22	3.27	68.61	1.49

In this example, each country's results are presented on three lines, one for each value of the NEWTOC5 variable. The countries are identified in the first column and the second column describes the category of NEWTOC5 being reported. The third column reports the number of valid cases and the fourth the sum of weights of the sampled students. The next two columns report the percentage of students in each category and its standard error, followed by the estimated mean reading achievement and its standard error. The standard deviation of the achievement scores and corresponding standard error are in the last two columns.

From the first three lines in Exhibit 3.10, we see that, in Austria, 45.16 percent of students reported reading for fun every day or almost every day, 25.46 percent reported reading for fun once or twice a week, and 29.38 percent twice a month or less. We also see that the mean reading achievement was 555.44 (standard error of 3.00) for those reporting reading every day or almost every day, 535.13 (standard error of 2.43) for those reading once or twice a week, and 515.74 (standard error of 2.88) for those reading twice a month or less.

Student-level Regression Analysis

This section demonstrates a regression analysis, using variables selected in the example merged data file ASGALLR2. The example used in this section does not have a corresponding exhibit in the *PIRLS 2006 International Report*.

In this example we will examine gender as a predictor of the number of books in the home. The linear regression analysis will use the variable REGSEX as the predictor and the number of books in the home (BOOK), using the weighting variable TOTWGT. The data will be read from the merged data file ASGALLR2.SAV and the standard errors will be computed based on 75 replicate weights.

Please note for this example the values of the variable ITSEX are recoded into variable REGSEX and the values of the variable ASBGBOOK are recoded into variable BOOK. By using these recoded variables, the intercept or constant will be the estimated average number of books in the homes of girls, whereas the regression coefficient (REGSEX (estimate)) will be the estimated number of additional books in the homes of boys. This will also allow us to perform a t-test to determine if the number of books in the home is significantly different between girls and boys. The syntax for these recodes is provided in Exhibit 3.11.

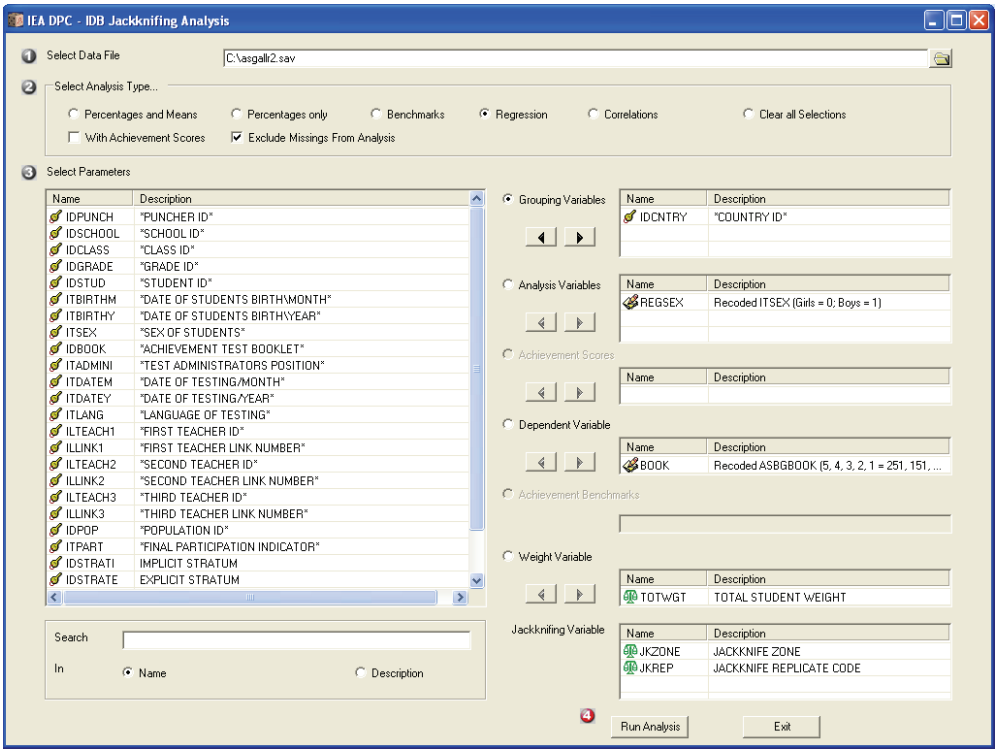
Exhibit 3.11 Example SPSS Program to Recode Variables for Regression Analysis

```
get file = "<datpath>asgallr2.sav".

* Compute new variable REGSEX from ITSEX .
compute REGSEX = ITSEX-1.
value labels REGSEX 0 'Girl'
                    1 'Boy' .

* Compute new variable BOOK from ASBGBOOK .
recode ASBGBOOK (5 = 251) (4 = 151) (3 = 63) (2 = 18) (1 = 5) (else
=sysmis) into BOOK .
value labels BOOK   5 '5 Books at home'
                   18 '18 Books at home'
                   63 '63 Books at home'
                  151 '151 Books at home'
                  251 '251 Books at home' .

variable labels
    regsex "Recoded ITSEX (Girls = 0; Boys = 1)"
    book   "Recoded ASBGBOOK (5, 4, 3, 2, 1 = 251, 151, 63, 18, 5)".

save outfile = "<datpath>asgallr2.sav".
```

Within the analysis module, the steps are as follows, with the completed analysis window shown in Exhibit 3.12:

1) Select the data file called ASGALLR2 that contains the recoded variables BOOK and REGSEX.
2) Select **Regression** as the Analysis Type.
3) By default, IDCNTRY is selected as the **Grouping Variable**. No other variable needs to be added for this example.
4) Next, press the radio button **Analysis Variables** to activate the section and define the analysis variable. For this analysis, you will need to select variable REGSEX from the list of variables and move it to the analysis variables field by pressing the right arrow button in this section.
5) Select the radio button **Dependent Variable**. Select variable BOOK from the variable list and move it to the dependent variable field by pressing the right arrow button in this section.
6) The **Weight Variable** is automatically defined by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. Additionally, the **Jackknifing Variables** JKZONE and JKREP are selected.
7) Press the button **Run Analysis** to perform the analysis. The program will alert you as soon as the analysis has been completed. The resulting data file, called REG, will open in SPSS. To use this file in further analyses you will need to save the resulting data file.

Exhibit 3.12 IDB-Analyzer Analysis Module Set-Up: Computing Regression Equations



The SPSS output of the linear regression results is presented in Exhibit 3.14. From the first line of the results, we see that in Austria the estimated average number of books in the homes of fourth grade girls, labeled "Constant (estimate)", is 82.81, with a standard error of 2.46. The Austrian fourth grade boys have an estimated 4.24 fewer books in the home, labeled "REGSEX (estimate)", than girls. With an estimated standard error of 3.04, this difference is not statistically significant at a 95% confidence level (REGSEX (t-test) is smaller than the absolute value of 1.96).

Exhibit 3.13 SPSS Output for Regression Analysis Example

```
Predictors: REGSEX / Predicted: BOOK                                                    PAGE    1

                        N of                  Constant     Constant      REGSEX       REGSEX       REGSEX
*COUNTRY ID*            Cases     Mult_RSQ    (estimate)     (s.e.)     (estimate)     (s.e.)     (t-test)
_____________ _____     _____     _____       _____        _____        _____        _____        _____

Austria                  4923         .00        82.81         2.46        -4.24         3.04        -1.40
Bulgaria                 3723         .00        85.55         3.97         -.24         3.99         -.06
Chinese Taipei           4465         .00        83.65         1.96        -2.48         2.69         -.92
                                                     …
South Africa            10831         .00        49.32         2.33        -2.22         1.88        -1.18
Spain                    3895         .00        89.16         2.78          .49         3.13          .16
Sweden                   4250         .00       117.38         2.51        -2.43         2.75         -.88
Trinidad And Tobago      3691         .00        82.58         3.14        -7.18         4.29        -1.67
Macedonia                3433         .00        65.16         2.93         -.17         3.06         -.06
United States            4941         .00        94.24         3.03        -3.36         2.62        -1.28
```

Student-level Regression Analysis with Achievement Scores

The next example of a student-level regression analysis will examine gender as a predictor of reading achievement based on the five plausible values (ASRREA01 through ASRREA05), using the weighting variable TOTWGT. The data will be read from the recoded data file ASGALLR2.SAV (see syntax in Exhibit 3.12). Standard errors will be computed based on 75 replicate weights. The example used in this section does not have a corresponding exhibit in the *PIRLS 2006 International Report*.

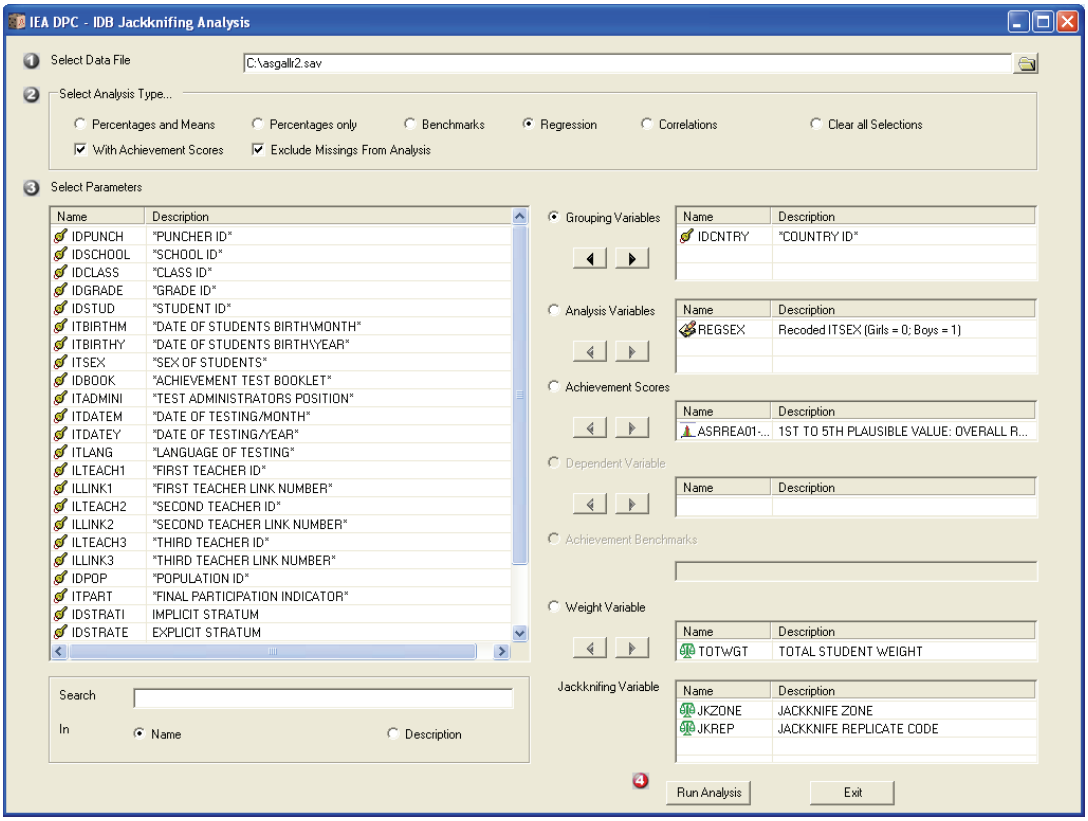
Within the analysis module, the steps are as follows, with the completed analysis window shown in Exhibit 3.14:

1) Select the data file called ASGALLR2.SAV.
2) Select **Regression** as the Analysis Type.
3) Make sure to check the box **With Achievement Scores**.
4) IDCNTRY is the only **Grouping Variable** required for this example.
5) Press the radio button **Analysis Variables** to activate the session and select variable REGSEX from the list of variables and move it to the analysis variables field by pressing the right arrow button in this section.
6) Activate the **Achievement Scores** and select variable ASRREA01-05 from the list of variables and move it to the field by pressing the right arrow button in this section.

7) The **Weight Variable** is automatically defined by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. Additionally the **Jackknifing Variables** JKZONE and JKREP are selected.

8) Press the button **Run Analysis** in order to perform your analysis. The program will alert you as soon as the analysis has been completed.

Exhibit 3.14 IDB Analyzer Analysis Module Set-Up: Regression with Achievement Scores



Results of this analysis are presented in Exhibit 3.15. From the first line of the results, we see that in Austria the estimated mean reading achievement of fourth grade girls, labeled "Constant (estimate)", is 543.26, with a standard error of 2.32. The Austrian fourth-grade boys have an estimated mean reading achievement 9.82 points (REGSEX (estimate)) lower than girls. With an estimated standard error of 2.35, this difference is indeed statistically significant at a 95% confidence level (REGSEX (t-test) value is larger than the absolute value of 1.96).

Exhibit 3.15 SPSS Output for Regression with Achievement Scores

COUNTRY ID	N of Cases	Mult_RSQ	Constant (estimate)	Constant (s.e.)	REGSEX (estimate)	REGSEX (s.e.)	REGSEX (t-test)
Austria	5067	.01	543.26	2.32	-9.82	2.35	-4.18
Bulgaria	3863	.02	557.52	4.41	-20.74	3.83	-5.41
Chinese Taipei	4589	.01	542.08	2.22	-12.83	1.86	-6.89
			…				
South Africa	14563	.02	319.43	6.32	-35.98	4.55	-7.90
Spain	4087	.00	514.61	2.60	-4.08	2.78	-1.47
Sweden	4394	.02	558.55	2.60	-17.76	2.45	-7.24
Trinidad And Tobago	3951	.02	451.22	4.95	-30.86	5.56	-5.55
Macedonia	4002	.01	453.39	4.44	-21.38	3.49	-6.13
United States	5187	.00	544.89	3.33	-10.07	3.17	-3.17

Calculating Percentages of Students Reaching Benchmarks

This section describes the IDB Analyzer's Benchmarks analysis, which computes percentages of students reaching benchmarks or selected cut points on the achievement distribution within specified subgroups, along with appropriate standard errors for those percentages. As an example, we will compute the percentages of students reaching the PIRLS 2006 High International Benchmark using the merged ASGALLR2 data file.

The steps in the analysis module of the IDB Analyzer are as follows:

1) Select **Benchmarks** as the analysis type.
2) No additional Grouping Variable is required for this analysis.
3) Press the radio button **Achievement Scores** to activate the session and select variable ASRREA01-05 from the list of variables and move it to the field by pressing the right arrow button in this section.
4) Press the radio button **Benchmarks** to activate the session and specify the cut points in the distribution. For our example we will use the High International Benchmark of achievement (550). Enter this value in the input field.
5) The **Weight Variable** is automatically defined by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. Additionally, the **Jackknifing Variables** JKZONE and JKREP are selected.
6) Press the button **Run Analysis** in order to perform your analysis. The program will alert you as soon as the analysis has been completed.

The completed analysis window is shown in Exhibit 3.16, and results are provided in Exhibit 3.17. From the first few lines of the results shown in Exhibit 3.17, we see

that in Austria there are 54.93 percent of students below the High International Benchmark of 550, and 45.07 percent of students reaching the High International Benchmark, with a standard error of 1.53.

Exhibit 3.16 IDB Analyzer Analysis Module Set-Up: Computing Percents Reaching Benchmarks

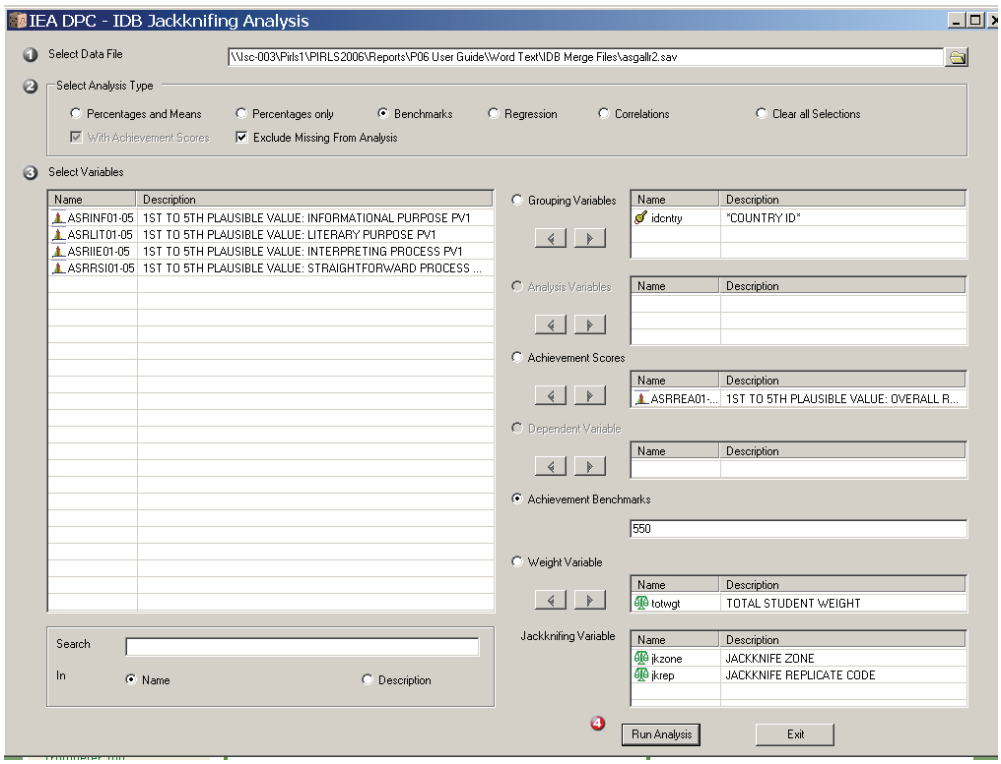


Exhibit 3.17 SPSS Output for Benchmarks

COUNTRY ID	Performance Group	N of Cases	Sum of TOTWGT	Percent	Percent (s.e.)
Austria	Below 550	2792	45687	54.93	1.53
	Above 550	2275	37482	45.07	1.53
Bulgaria	Below 550	1732	30660	48.38	2.26
	Above 550	2131	32712	51.62	2.26
Chinese Taipei	Below 550	2593	173465	56.97	1.34
	Above 550	1996	131024	43.03	1.34
South Africa	Below 550	14122	775839	94.18	.91
	Above 550	535	47981	5.82	.91
Spain	Below 550	2760	269729	68.97	1.30
	Above 550	1334	121355	31.03	1.30
Sweden	Below 550	2083	48337	47.48	1.51
	Above 550	2311	53472	52.52	1.51
Trinidad And Tobago	Below 550	3335	14945	86.94	1.22
	Above 550	616	2245	13.06	1.22
Macedonia	Below 550	3409	19492	85.01	1.12
	Above 550	593	3436	14.99	1.12
United States	Below 550	2857	1778834	53.07	2.03
	Above 550	2333	1573125	46.93	2.03

Computing Correlations with Background and Achievement Variables

In addition to the analyses described above, the IDB Analyzer is also able to compute correlations between background variables, and between background variables and achievement variables. While these types of analyses will not be described here in full, the principles for conducting them are the same as those described previously: you need to select the grouping variables, the analysis variables, the achievement score (if necessary), and confirm the weight and sampling information variables. The output will display, for each group defined by the grouping variables, the correlation coefficients for each possible pair of variables. When using only background variables, the diagonal and elements above the diagonal of the correlation matrix are displayed with their corresponding standard errors. When using achievement scores, a single column is displayed containing the correlations between each of the background variables specified in the model and the achievement score selected. Due to limitations in the current implementation of the correlation analysis, the grouping variables are not displayed with the corresponding labels. The user will need to keep a record of the labels for each of the values of the grouping variables in order to interpret the results.

3.6 Performing Analyses with Teacher Data

As our example of an analysis using teacher data, we will investigate the percentage of students according to the age of their PIRLS reading teachers within each country. The results of such an analysis are presented in Exhibit 6.3 of the *PIRLS 2006 International Report* and are reproduced here in Exhibit 3.18. We will use the **Percent Only** analysis type to estimate the percentages of students within the reporting categories of teachers' age.

As with the previous examples, we first proceed to identify the variables relevant to the analysis in the appropriate files, and review the documentation for any specific national adaptations to the questions of interest (Supplements 1 and 2). Since we are using a teacher-level variable, we need to look in the teacher background data files for the variable that contains the information on the reading teachers' age (ATBGAGE).

Exhibit 3.18 Example Exhibit of Teacher Variable Analysis from the PIRLS 2006 International Report (Exhibit 6.3)

Exhibit 6.3 **Teachers' Gender, Age, and Number of Years Teaching**

PIRLS 2006
4th Grade

Countries	Percentage of Students by Teacher Characteristics						Trends in Number of Years Teaching All Grades		
	Gender		Age						
	Female	Male	29 Years or Under	30-39 Years	40-49 Years	50 Years or Older	2006	Difference from 2001	
Austria	88 (2.1)	12 (2.1)	6 (1.7)	17 (2.5)	39 (3.0)	38 (3.2)	22 (0.7)	◊ ◊	
Belgium (Flemish)	75 (3.0)	25 (3.0)	28 (3.2)	29 (3.3)	29 (2.9)	14 (2.2)	16 (0.6)	◊ ◊	
Belgium (French)	80 (3.0)	20 (3.0)	16 (2.2)	33 (3.2)	37 (3.5)	14 (2.6)	17 (0.6)	◊ ◊	
Bulgaria	94 (1.8)	6 (1.8)	3 (1.4)	26 (2.8)	40 (4.0)	31 (3.9)	21 (0.6)	4 (0.9)	▲
Canada, Alberta	80 (3.2)	20 (3.2)	22 (3.7)	20 (3.3)	26 (3.6)	32 (3.3)	15 (0.8)	◊ ◊	
Canada, British Columbia	r 72 (3.6)	28 (3.6)	r 9 (2.2)	21 (3.0)	32 (4.1)	38 (4.2)	17 (0.8)	◊ ◊	
Canada, Nova Scotia	84 (2.9)	16 (2.9)	12 (2.4)	22 (3.4)	25 (3.0)	40 (3.4)	18 (0.7)	◊ ◊	
Canada, Ontario	75 (4.8)	25 (4.8)	18 (3.9)	42 (5.2)	21 (4.3)	20 (4.3)	12 (0.9)	-4 (1.3)	▼
Canada, Quebec	86 (3.2)	14 (3.2)	12 (2.3)	37 (4.0)	19 (3.5)	32 (3.7)	17 (0.8)	-1 (1.2)	
Chinese Taipei	83 (3.2)	17 (3.2)	24 (3.7)	44 (4.1)	27 (3.5)	5 (1.6)	12 (0.6)	◊ ◊	
Denmark	90 (2.1)	10 (2.1)	16 (2.9)	24 (3.4)	19 (2.9)	41 (4.3)	16 (1.1)	◊ ◊	
England	75 (3.5)	25 (3.5)	30 (3.8)	33 (4.1)	14 (2.9)	23 (3.7)	12 (0.9)	-2 (1.3)	
France	71 (3.3)	29 (3.3)	17 (2.9)	31 (3.0)	34 (3.4)	18 (2.6)	15 (0.7)	-3 (1.1)	▼
Georgia	100 (0.3)	0 (0.0)	9 (1.6)	28 (3.4)	24 (3.4)	39 (3.6)	20 (0.9)	◊ ◊	
Germany	89 (2.4)	11 (2.4)	5 (1.7)	23 (2.9)	22 (3.0)	49 (3.8)	20 (0.9)	-3 (1.1)	▼
Hong Kong SAR	78 (3.7)	22 (3.7)	29 (4.2)	33 (4.1)	19 (3.5)	19 (2.6)	13 (0.7)	0 (1.2)	
Hungary	97 (1.6)	3 (1.6)	6 (2.1)	21 (3.1)	48 (3.7)	24 (3.5)	21 (0.7)	3 (1.0)	▲
Iceland	93 (0.2)	7 (0.2)	13 (0.3)	31 (0.4)	35 (0.4)	22 (0.3)	12 (0.1)	-1 (0.1)	▼
Indonesia	56 (3.9)	44 (3.9)	13 (2.2)	31 (4.1)	39 (3.9)	17 (3.0)	16 (0.8)	◊ ◊	
Iran, Islamic Rep. of	50 (2.3)	50 (2.3)	8 (1.6)	44 (3.6)	41 (3.5)	7 (1.9)	17 (0.5)	3 (0.8)	▲
Israel	92 (1.9)	8 (1.9)	8 (1.8)	41 (4.0)	33 (4.1)	18 (3.4)	16 (0.8)	2 (1.1)	
Italy	98 (1.1)	2 (1.1)	2 (1.3)	14 (2.5)	37 (3.5)	47 (3.6)	22 (0.7)	1 (1.0)	
Kuwait	86 (2.3)	14 (2.3)	36 (3.8)	48 (3.8)	15 (3.1)	1 (0.0)	8 (0.6)	◊ ◊	
Latvia	99 (0.6)	1 (0.0)	7 (2.0)	30 (3.6)	34 (3.6)	29 (3.4)	21 (0.8)	2 (1.3)	
Lithuania	99 (0.5)	1 (0.5)	1 (0.7)	35 (3.4)	40 (3.2)	23 (3.2)	21 (0.6)	1 (1.1)	
Luxembourg	55 (0.2)	45 (0.2)	32 (0.2)	24 (0.1)	16 (0.1)	27 (0.1)	15 (0.0)	◊ ◊	
Macedonia, Rep. of	70 (3.5)	30 (3.5)	5 (1.7)	24 (3.4)	40 (4.1)	31 (4.2)	r 20 (1.0)	0 (1.5)	
Moldova, Rep. of	90 (2.8)	10 (2.8)	8 (2.3)	23 (3.7)	36 (4.3)	33 (4.2)	24 (0.9)	5 (1.2)	▲
Morocco	56 (4.0)	44 (4.0)	18 (3.0)	25 (3.5)	40 (3.4)	17 (2.9)	17 (0.6)	1 (1.0)	
Netherlands	68 (3.4)	32 (3.4)	26 (3.5)	19 (3.0)	19 (3.5)	36 (4.0)	17 (1.0)	0 (1.4)	
New Zealand	77 (2.7)	23 (2.7)	22 (2.1)	28 (2.6)	26 (2.4)	25 (2.4)	12 (0.6)	-1 (1.1)	
Norway	91 (1.9)	9 (1.9)	9 (2.7)	28 (3.3)	29 (3.9)	35 (3.9)	16 (1.0)	-1 (1.3)	
Poland	100 (0.3)	0 (0.3)	2 (1.1)	31 (3.4)	58 (4.0)	8 (2.3)	20 (0.5)	◊ ◊	
Qatar	r 90 (0.1)	10 (0.1)	r 27 (0.2)	54 (0.3)	16 (0.2)	3 (0.1)	11 (0.0)	◊ ◊	
Romania	89 (2.4)	11 (2.4)	18 (3.1)	25 (3.3)	27 (3.5)	30 (3.2)	22 (0.9)	2 (1.2)	
Russian Federation	98 (1.1)	2 (1.1)	6 (1.6)	38 (3.7)	35 (3.5)	21 (2.8)	22 (0.6)	2 (1.0)	▲
Scotland	96 (1.6)	4 (1.6)	26 (3.8)	19 (2.8)	18 (3.5)	37 (4.2)	16 (1.1)	-2 (1.5)	
Singapore	75 (2.2)	25 (2.2)	37 (2.6)	42 (3.0)	16 (2.4)	6 (1.0)	9 (0.4)	-2 (1.0)	▼
Slovak Republic	93 (1.7)	7 (1.7)	13 (2.4)	37 (3.5)	25 (2.7)	25 (3.2)	17 (0.8)	0 (1.2)	
Slovenia	98 (1.1)	2 (1.1)	12 (2.2)	24 (2.9)	44 (3.0)	19 (2.4)	19 (0.7)	0 (1.0)	
South Africa	71 (2.5)	29 (2.5)	4 (1.2)	44 (2.8)	32 (2.9)	20 (2.3)	15 (0.4)	◊ ◊	
Spain	78 (3.6)	22 (3.6)	11 (2.4)	16 (2.6)	25 (3.4)	49 (3.8)	22 (0.9)	◊ ◊	
Sweden	84 (2.8)	16 (2.8)	9 (1.7)	29 (3.3)	24 (2.9)	38 (3.6)	17 (1.0)	1 (1.2)	
Trinidad and Tobago	82 (3.0)	18 (3.0)	11 (2.5)	37 (3.6)	28 (3.3)	24 (3.0)	19 (0.7)	◊ ◊	
United States	85 (2.7)	15 (2.7)	21 (2.8)	27 (2.8)	28 (3.7)	25 (3.4)	12 (0.7)	-3 (1.1)	▼
International Avg.	83 (0.4)	17 (0.4)	15 (0.4)	30 (0.5)	30 (0.5)	25 (0.5)	17 (0.1)	0 (0.0)	

▲ Number of years in 2006 significantly higher ▼ Number of years in 2006 significantly lower

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

The merged file ATGMERGED.SAV will be used for this example. The variable that identifies the country (IDCNTRY), and the two identification variables (IDTEACH and IDLINK) that will allow us to link the teacher data to the student data are selected automatically by the IDB Analyzer, and so are the variables that contain the sampling information and that will be used to generate the replicate weights for the analysis.

Note that one of the steps in reproducing this analysis is to combine categories 1 and 2 and categories 5 and 6 of the variable ATBGAGE in order to match the results presented in the international report, where teachers are categorized into four groups: 29 years or under, 30-39 years old, 40-49 years old, and 50 years or older. SPSS syntax, presented in Exhibit 3.19, is used to recode ATBGAGE into a new variable, NEWAGE.

Exhibit 3.19 Example SPSS Program to Recode Variables for Teacher Variable Analysis

```
get file = "<datpath>atgmerged.sav".

* Compute new variable newAGE from ATBGAGE .
recode atbgage (1,2=1) (3=2) (4=3) (5,6=4) (else=sysmis) into newAGE.

variable labels
    newAGE "Recoded ATBGAGE Teacher Age".

value labels newAGE
     1 '29 yrs or under'
     2 '30-39 years'
     3 '40-49 years'
     4 '50 years or older' .

save outfile = "<datpath>atgmerged.sav".
```

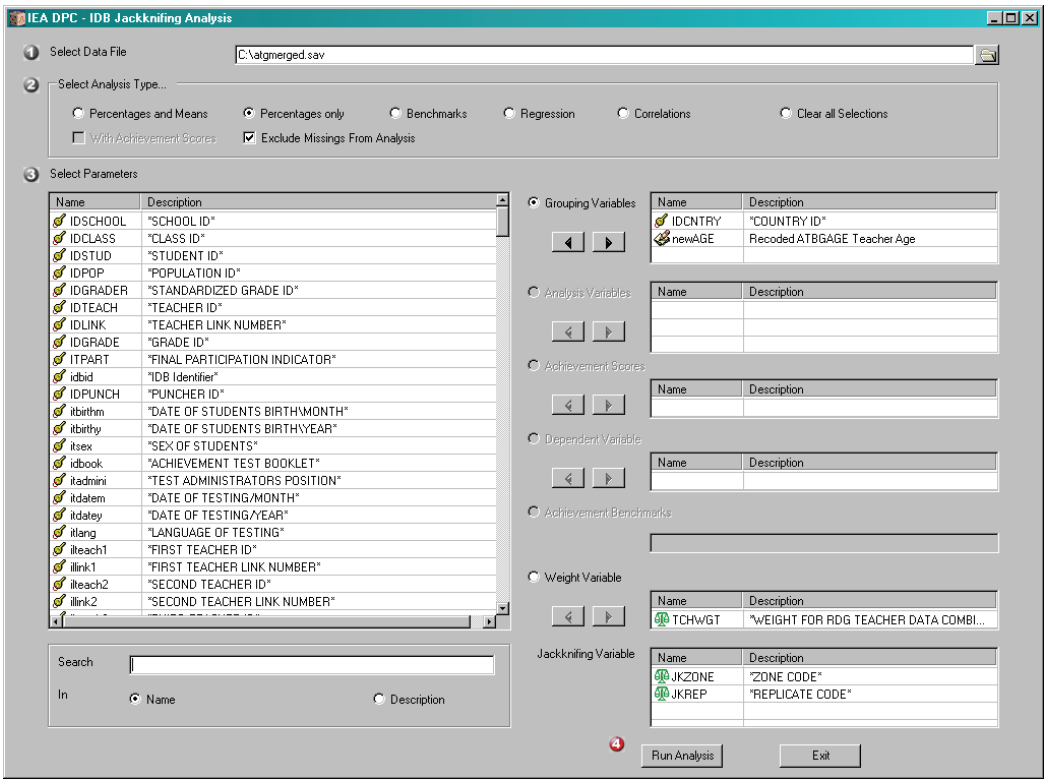
Once the recode is completed, we then proceed to open the analysis module of the IDB Analyzer, select the recoded ATGMERGED.SAV data file for the analysis, and follow the steps below:

1) As type of the analysis select **Percentages only**.
2) Add ITSEX as an additional **Grouping Variable**.
3) The **Weight Variable** is automatically defined by the software. As this is an example for analysis on teacher-level, the weight TCHWGT is selected by default. Additionally, the **Jackknifing Variables** JKZONE and JKREP are selected.

4) Press the button **Run Analysis**. The program will alert you as soon as the analysis has been completed. The resulting data file will be open in SPSS called FINAL. If you want to use it for further analysis it is recommended to save the resulting data file.

Exhibit 3.21 below shows the analysis module display with variable selection completed. Notice that for analysis type we have selected **Percents Only**, with the option With Achievement Scores not checked. The results of this analysis are presented in Exhibit 3.21.

Exhibit 3.20 Example IDB Analyzer Setup for Teacher Variable Analysis



Each country's results are presented on four lines, one for each value of the NEWAGE variable. The results are presented in the same manner as in the previous example, with countries identified in the first column and the second column describing the categories of NEWAGE. From Exhibit 5.16, we see that 6.41 percent of students in Austria were taught by teachers 29 years or younger, 16.52 percent by teachers 30-39 years old, 39.30 percent by teachers 40-49 years old, and 37.77 percent by teachers 50 years or older.

Exhibit 3.21 Output for Example Teacher Variable Analysis

COUNTRY ID	Recoded ATBGAGE Teacher Age	N of Cases	Sum of TCHWGT	Percent	Percent (s.e.)
AUSTRIA	29 yrs or under	320	5289	6.41	1.65
	30-39 years	822	13629	16.52	2.48
	40-49 years	2115	32431	39.30	3.01
	50 years or older	1812	31169	37.77	3.17
BELGIUM (Flemish)	29 yrs or under	1202	18077	27.71	3.22
	30-39 years	1321	18825	28.86	3.34
	40-49 years	1336	19196	29.43	2.85
	50 years or older	551	9139	14.01	2.20
BELGIUM (French)	29 yrs or under	718	7158	15.75	2.18
	30-39 years	1443	15013	33.04	3.17
	40-49 years	1569	16708	36.77	3.55
	50 years or older	614	6557	14.43	2.58
		.....			
SWEDEN	29 yrs or under	428	8179	8.59	1.65
	30-39 years	1307	27345	28.71	3.33
	40-49 years	1083	23098	24.25	2.88
	50 years or older	1697	36636	38.46	3.61
TRINIDAD AND TOBAGO	29 yrs or under	399	1879	11.35	2.54
	30-39 years	1411	6125	36.99	3.62
	40-49 years	1012	4592	27.73	3.34
	50 years or older	992	3963	23.93	3.00
UNITED STATES	29 yrs or under	1011	664097	20.72	2.75
	30-39 years	1376	849591	26.51	2.84
	40-49 years	1252	885575	27.63	3.72
	50 years or older	1316	805499	25.13	3.44

3.7 Performing Analyses with School Variables

When performing analyses with the merged school-level data files, such as ACGMERGED.SAV that was created earlier in this chapter, the data are analyzed to make statements about the number or percentages of students attending schools with one characteristic or another, rather than about the number or percentages of schools with one characteristic or another. Our example of a school-level analysis will investigate the percentage of students who attend schools in areas of different population sizes, and the average reading achievement within each category of school location. The results of this analysis are presented in Exhibit 7.1 of the *PIRLS 2006 International Report* and are reproduced here in Exhibit 3.22. To conduct this analysis, we will use the **Means and Percentages** analysis type in the IDB Analyzer with the option **With Achievement Scores** checked.

The first step in our analysis is to identify the variables of interest in the appropriate files and review the documentation on specific national adaptations to the questions of interest (Supplements 1 and 2). We observe that the variable ACBGCOMM in the school background data file contains information on the size of the community where the schools are located. Our next step is to review the

documentation of national adaptations to the questionnaires to ensure that there were no deviations listed for this variable (see Supplement 2). If no adaptations were made, we can continue with our analysis without any modifications.

Exhibit 3.22 Example Exhibit of School Variable Analysis from the PIRLS 2006 International Report (Exhibit 7.1)

Exhibit 7.1 **Principals' Reports on Their Schools' Locations with Trends** — **PIRLS 2006** 4th Grade

Countries	Urban			Suburban			Rural		
	2006 Percent of Students	Average Achievement	Difference in Percent from 2001	2006 Percent of Students	Average Achievement	Difference in Percent from 2001	2006 Percent of Students	Average Achievement	Difference in Percent from 2001
Austria	31 (3.4)	529 (4.4)	◊ ◊	20 (3.2)	542 (5.5)	◊ ◊	48 (3.7)	543 (2.5)	◊ ◊
Belgium (Flemish)	21 (3.6)	541 (5.4)	◊ ◊	37 (4.7)	546 (3.7)	◊ ◊	42 (4.8)	551 (2.4)	◊ ◊
Belgium (French) r	47 (4.0)	494 (5.5)	◊ ◊	21 (3.9)	496 (5.7)	◊ ◊	33 (3.8)	512 (4.2)	◊ ◊
Bulgaria	70 (3.0)	557 (5.2)	6 (4.1)	5 (1.7)	550 (10.8)	−6 (3.1) ▼	24 (2.5)	516 (10.8)	0 (3.4)
Canada, Alberta	46 (4.5)	559 (3.8)	◊ ◊	26 (3.5)	572 (4.4)	◊ ◊	28 (3.5)	550 (4.1)	◊ ◊
Canada, British Columbia	38 (4.3)	555 (4.3)	◊ ◊	46 (4.7)	565 (3.5)	◊ ◊	17 (3.2)	545 (6.2)	◊ ◊
Canada, Nova Scotia	25 (3.2)	542 (6.5)	◊ ◊	26 (3.2)	551 (3.9)	◊ ◊	50 (3.5)	537 (2.6)	◊ ◊
Canada, Ontario	51 (4.9)	549 (4.1)	8 (6.9)	36 (5.0)	563 (4.3)	1 (7.0)	14 (3.2)	552 (4.9)	−9 (5.1)
Canada, Quebec	51 (4.8)	533 (3.7)	15 (6.7) ▲	28 (4.0)	538 (6.5)	−19 (6.3) ▼	20 (3.6)	528 (5.6)	3 (5.0)
Chinese Taipei	– –	– –	– –	– –	– –	– –	– –	– –	– –
Denmark	33 (4.1)	545 (3.4)	◊ ◊	30 (3.4)	555 (3.9)	◊ ◊	37 (4.0)	542 (4.4)	◊ ◊
England r	45 (4.2)	523 (5.3)	−2 (6.5)	35 (3.9)	553 (5.1)	4 (6.0)	19 (3.7)	564 (5.2)	−1 (5.2)
France	34 (4.0)	522 (4.3)	−2 (5.7)	25 (3.9)	518 (6.3)	−5 (5.7)	41 (3.9)	524 (2.5)	6 (5.3)
Georgia	42 (3.6)	486 (4.5)	◊ ◊	15 (2.7)	465 (8.1)	◊ ◊	43 (2.6)	459 (5.1)	◊ ◊
Germany	37 (3.3)	535 (4.6)	4 (4.4)	19 (3.1)	557 (3.9)	−3 (4.3)	44 (4.0)	555 (2.3)	−1 (5.5)
Hong Kong SAR	58 (4.4)	573 (3.1)	6 (5.3)	37 (4.1)	555 (4.5)	−9 (4.9)	5 (1.9)	540 (11.0)	3 (2.2)
Hungary	28 (2.2)	565 (6.6)	0 (3.3)	40 (2.5)	557 (5.0)	5 (3.4)	31 (1.8)	528 (4.7)	−5 (2.5) ▼
Iceland r	33 (0.3)	518 (2.0)	−3 (0.5) ▼	37 (0.3)	509 (2.3)	−6 (0.5) ▼	30 (0.4)	506 (2.3)	9 (0.5) ▲
Indonesia	12 (2.2)	451 (9.7)	◊ ◊	14 (2.7)	425 (9.6)	◊ ◊	74 (2.9)	393 (4.8)	◊ ◊
Iran, Islamic Rep. of	50 (2.9)	454 (4.1)	1 (4.7)	15 (2.2)	415 (10.3)	6 (3.3) ▲	35 (2.8)	376 (5.7)	−8 (4.2)
Israel	49 (3.9)	534 (6.0)	−3 (5.5)	18 (2.8)	529 (13.1)	−4 (4.3)	33 (3.8)	472 (10.2)	7 (5.0)
Italy	70 (3.6)	554 (2.9)	−6 (4.8)	15 (2.8)	555 (9.2)	1 (3.8)	15 (3.1)	533 (9.9)	5 (3.8)
Kuwait	26 (3.6)	355 (7.6)	◊ ◊	61 (4.0)	321 (5.8)	◊ ◊	14 (3.0)	311 (12.1)	◊ ◊
Latvia	70 (0.7)	548 (2.4)	26 (3.9) ▲	3 (1.5)	528 (6.6)	−15 (4.2) ▼	27 (1.7)	525 (5.9)	−10 (3.4) ▼
Lithuania	72 (2.3)	544 (1.9)	1 (3.6)	3 (1.4)	549 (10.7)	−3 (2.5)	26 (2.1)	516 (3.4)	3 (3.2)
[1] Luxembourg	– –	– –	– –	– –	– –	– –	– –	– –	– –
Macedonia, Rep. of r	51 (3.6)	477 (6.9)	−6 (4.9)	18 (3.5)	443 (13.8)	3 (4.5)	31 (2.6)	401 (9.5)	3 (4.1)
Moldova, Rep. of	29 (2.4)	517 (4.5)	2 (4.1)	6 (2.4)	498 (17.6)	−8 (4.5)	65 (2.5)	492 (3.9)	6 (4.5)
Morocco r	37 (3.3)	363 (7.2)	−3 (5.5)	18 (3.6)	334 (15.8)	−3 (5.7)	45 (3.7)	296 (13.5)	6 (5.5)
Netherlands	26 (4.0)	538 (4.1)	−5 (5.6)	33 (4.7)	553 (3.2)	10 (6.0)	41 (3.5)	547 (2.5)	−5 (5.2)
New Zealand	41 (3.2)	536 (3.3)	3 (5.1)	39 (3.0)	527 (3.6)	−1 (4.7)	21 (2.3)	535 (5.3)	−2 (3.7)
Norway	20 (3.6)	502 (3.8)	1 (5.0)	30 (3.9)	504 (3.9)	4 (5.5)	50 (4.2)	492 (4.2)	−6 (5.5)
Poland	52 (2.1)	528 (2.9)	◊ ◊	5 (1.7)	529 (11.0)	◊ ◊	43 (1.9)	508 (3.8)	◊ ◊
Qatar	65 (0.2)	362 (1.5)	◊ ◊	32 (0.2)	336 (2.0)	◊ ◊	3 (0.0)	318 (8.4)	◊ ◊
Romania	47 (2.2)	515 (6.5)	−3 (3.4)	5 (1.9)	498 (14.8)	1 (2.8)	48 (2.4)	462 (8.0)	2 (3.6)
Russian Federation	63 (2.0)	581 (3.4)	8 (3.2) ▲	6 (1.3)	563 (8.8)	4 (2.3)	31 (2.2)	532 (6.1)	−12 (3.3) ▼
Scotland r	32 (3.5)	517 (4.8)	−2 (5.6)	36 (4.3)	539 (5.3)	−3 (6.9)	32 (3.9)	528 (6.8)	6 (5.9)
Singapore	100 (0.0)	558 (2.9)	0 (0.0)	0 (0.0)	~ ~	0 (0.0)	0 (0.0)	~ ~	0 (0.0)
Slovak Republic	52 (3.0)	544 (2.9)	2 (4.8)	8 (2.5)	537 (8.0)	−2 (3.5)	40 (3.3)	512 (5.9)	0 (4.5)
Slovenia	36 (4.2)	529 (3.6)	−4 (5.3)	37 (4.0)	520 (3.1)	10 (5.5)	27 (3.7)	512 (3.6)	−6 (4.7)
South Africa	17 (1.8)	350 (19.5)	◊ ◊	21 (2.2)	381 (14.9)	◊ ◊	62 (2.0)	261 (3.8)	◊ ◊
Spain	58 (4.3)	524 (3.4)	◊ ◊	20 (3.3)	497 (6.8)	◊ ◊	21 (3.4)	498 (7.1)	◊ ◊
Sweden	27 (4.1)	549 (3.8)	12 (5.1) ▲	55 (4.1)	549 (3.4)	−12 (5.6) ▼	18 (2.8)	550 (4.7)	0 (4.5)
Trinidad and Tobago	19 (2.5)	470 (13.0)	◊ ◊	50 (3.7)	441 (7.0)	◊ ◊	32 (3.0)	408 (8.9)	◊ ◊
United States	28 (3.5)	524 (4.4)	−5 (4.9)	47 (3.9)	550 (3.2)	13 (6.1) ▲	25 (2.7)	539 (9.1)	−8 (4.2)
International Avg.	43 (0.5)	508 (1.0)		24 (0.5)	501 (1.4)		33 (0.5)	483 (1.1)	

▲ Percent in 2006 significantly higher
▼ Percent in 2006 significantly lower

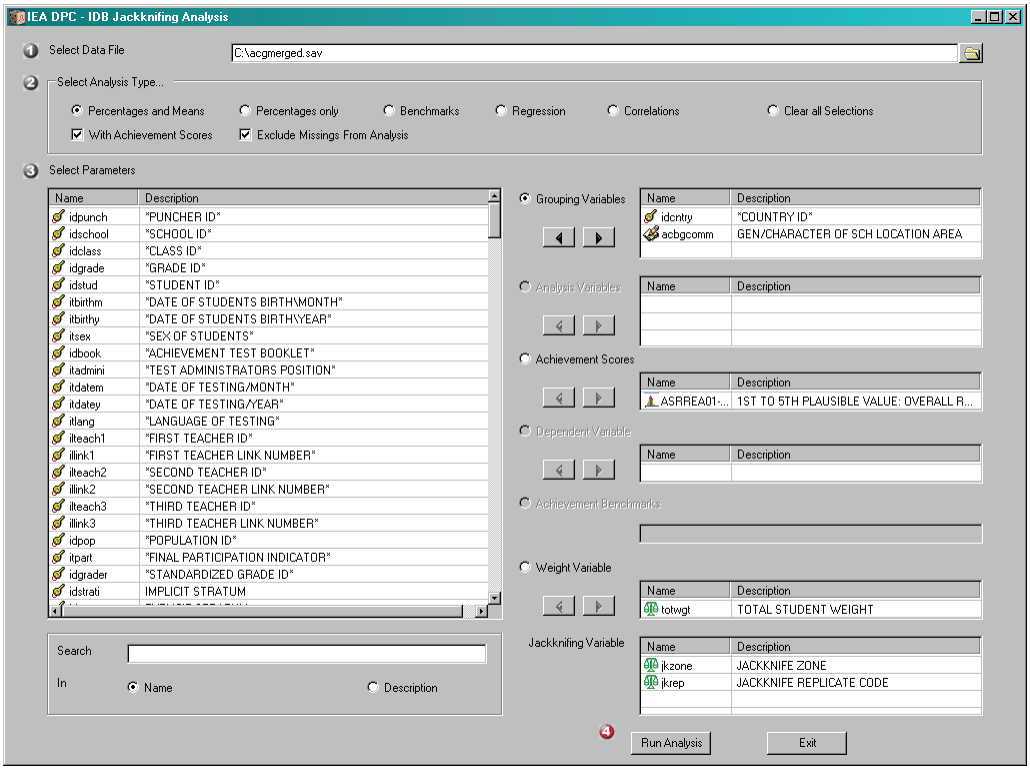
SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

For this analysis there is no need to recode the data. We therefore proceed to conduct the analysis with the IDB Analyzer analysis module selecting the merged school and student background variables file called ACGMERGED.SAV and following the steps for analysis:

1) Select **Percentages and Means** under Analysis Type. By default, the program will exclude those with missing grouping variables from the analysis.
2) Check the box next to **With Achievement Scores**.
3) Define the variables for the analysis:
4) Add the variable ACBGCOMM as a second **Grouping Variable**.
5) Next, define the achievement scores to be used for the analysis. To activate this section, you will need to press the radio button **Achievement Scores**. This time you will need to select variable ASRREA01-05 from the list of variables and move it to the analysis variables field by pressing the right arrow button in this section.
6) The **Weight Variable** is automatically defined by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. Additionally the **Jackknifing Variables** JKZONE and JKREP are selected.
7) Press the button **Run Analysis** in order to perform your analysis. The program will alert you as soon as the analysis has been completed and the results will be displayed in the output window.

Exhibit 3.23 shows the setup for this analysis and Exhibit 3.24 shows the results.

Exhibit 3.23 Example IDB Analyzer Setup for School Variable Analysis



In this example, each country's results are presented on three lines, one for each value of the ACBGCOMM variable. The countries are identified in the first column and the second column describes the categories of ACBGCOMM being reported. All other columns are identical in nature as in the previous two examples.

From the first three lines in the results, we see that in Austria, 31.40 percent of students come from schools in urban areas, 20.26 percent come from schools in suburban areas, and 48.34 percent from rural. We also see that the estimated mean reading achievement of students in urban areas is 528.62 (with a standard error of 4.43), whereas the estimated mean achievement of students in suburban and rural areas are, respectively, 541.75 (standard error of 5.46) and 543.12 (standard error of 2.51).

Notice that since we selected the option **Exclude Missing from Analysis**, Chinese Taipei has dropped from the analysis and does not appear in the output because the variable ACBGCOMM was set to missing for all cases in this country, as the questionnaire item was not administered.

Exhibit 3.24 Output for Example School Variable Analysis

COUNTRY ID	GEN/CHARACTE OF SCH LOCATION AREA	N of Cases	Sum of TOTWGT	Percent	Percent (s.e.)	ASRREA0 (Mean)	ASRREA0 (s.e.)	Std.Dev.	Std.Dev. (s.e.)
AUSTRIA	URBAN	1848	25951	31.40	3.39	528.62	4.43	67.20	1.71
	SUBURBAN	1235	16743	20.26	3.24	541.75	5.46	66.26	2.63
	RURAL	1949	39956	48.34	3.69	543.12	2.51	59.50	1.73
BULGARIA	URBAN	2976	42641	70.41	3.03	557.48	5.21	80.71	2.63
	SUBURBAN	221	3158	5.22	1.75	549.62	10.83	72.05	3.66
	RURAL	467	14760	24.37	2.46	516.24	10.84	84.44	5.29
DENMARK	URBAN	1304	19917	33.49	4.06	545.08	3.41	69.69	1.93
	SUBURBAN	1406	17650	29.68	3.44	555.49	3.88	67.07	2.16
	RURAL	980	21899	36.83	3.96	541.84	4.36	71.16	2.53
				.....					
SPAIN	URBAN	2349	208013	58.36	4.26	524.21	3.40	68.20	1.52
	SUBURBAN	732	72090	20.23	3.33	496.57	6.76	70.96	3.03
	RURAL	638	76330	21.42	3.37	498.21	7.07	72.00	4.68
SWEDEN	URBAN	1368	27171	26.99	4.11	548.74	3.76	62.94	1.79
	SUBURBAN	2507	55264	54.89	4.08	548.99	3.35	64.64	1.68
	RURAL	479	18242	18.12	2.77	550.01	4.71	61.27	3.84
TRINIDAD AND TOBAGO	URBAN	907	3162	18.74	2.55	469.93	13.03	102.29	5.88
	SUBURBAN	1954	8380	49.66	3.68	441.49	7.04	102.50	3.41
	RURAL	992	5333	31.60	3.01	407.74	8.87	97.51	3.51
UNITED STATES	URBAN	1461	908132	27.50	3.50	524.07	4.45	72.57	2.77
	SUBURBAN	2290	1563272	47.34	3.94	549.59	3.22	70.96	1.34
	RURAL	1349	830812	25.16	2.72	538.95	9.13	78.09	2.63

3.8 Performing Analyses with Home Background Variables

Our example of a home background analysis investigates the relationship between the frequency of parents' reading for enjoyment and their children's reading achievement. These results are presented in Exhibit 3.9 of the *PIRLS 2006 International Report* and are duplicated here in Exhibit 3.25.

The first step in our analysis is to identify the variables of interest in the appropriate files and review the documentation on specific national adaptations to the questions of interest (Supplements 1 and 2). We observe that the variable ASBHRRE in the home background data files contains categorical information on the frequency of parents' reading for enjoyment. The reading achievement scores (ASRREA01-5) are available in the student background data file. Since we require variables from both the home and student background data files we need to use the merged home and student data file created earlier and called ASHMERGED.SAV.

Exhibit 3.25 Example Exhibit of Home Background Variable Analysis Taken From the PIRLS 2006 International Report (Exhibit 3.9)

Exhibit 3.9 **Parents Reading for Enjoyment with Trends**

PIRLS 2006
4th Grade

Countries		Every Day or Almost Every Day			Once or Twice a Week			Twice a Month or Less		
		2006 Percent of Students	Average Achievement	Difference in Percent from 2001	2006 Percent of Students	Average Achievement	Difference in Percent from 2001	2006 Percent of Students	Average Achievement	Difference in Percent from 2001
Norway		66 (1.0)	505 (2.5)	7 (1.6) ▲	23 (0.9)	493 (4.2)	−4 (1.4) ▼	11 (0.7)	489 (6.5)	−3 (0.9) ▼
Sweden		64 (0.9)	556 (2.4)	−3 (1.3) ▼	21 (0.8)	546 (2.9)	−1 (1.1)	15 (0.7)	540 (4.6)	5 (0.9) ▲
Scotland	s	63 (1.3)	550 (4.0)	1 (1.8)	23 (0.9)	531 (5.3)	−3 (1.5)	14 (1.1)	525 (7.0)	2 (1.4)
Iceland	r	61 (0.9)	523 (1.9)	−1 (1.2)	21 (0.8)	513 (2.9)	−4 (1.1) ▼	18 (0.6)	498 (3.3)	5 (0.8) ▲
New Zealand	s	60 (1.1)	557 (2.4)	4 (1.6) ▲	25 (0.9)	539 (3.6)	−4 (1.4) ▼	15 (0.7)	528 (4.1)	0 (1.1)
Netherlands	s	59 (1.2)	562 (1.8)	−1 (1.8)	25 (0.9)	548 (2.9)	−3 (1.3) ▼	16 (0.9)	543 (2.6)	4 (1.2) ▲
Trinidad and Tobago		59 (0.9)	441 (5.1)	◊ ◊	31 (1.0)	436 (6.1)	◊ ◊	10 (0.7)	434 (9.2)	◊ ◊
Germany	r	58 (0.9)	561 (2.4)	7 (1.3) ▲	28 (0.7)	544 (2.9)	−3 (1.0) ▼	14 (0.6)	535 (4.5)	−3 (0.9) ▼
Denmark		57 (1.0)	554 (2.6)	◊ ◊	27 (0.9)	544 (3.0)	◊ ◊	16 (0.7)	534 (5.0)	◊ ◊
Latvia		57 (1.3)	547 (2.7)	3 (2.0)	32 (1.1)	537 (3.4)	−3 (1.5)	11 (0.6)	532 (4.2)	0 (1.2)
Canada, British Columbia	r	57 (1.1)	569 (3.2)	◊ ◊	29 (1.1)	560 (3.5)	◊ ◊	15 (0.7)	547 (4.7)	◊ ◊
Canada, Alberta	r	55 (1.2)	571 (2.8)	◊ ◊	29 (1.0)	561 (3.5)	◊ ◊	17 (0.7)	554 (3.7)	◊ ◊
Luxembourg		54 (0.7)	573 (1.3)	◊ ◊	29 (0.7)	544 (2.4)	◊ ◊	18 (0.6)	543 (2.4)	◊ ◊
Canada, Nova Scotia		53 (0.9)	556 (2.1)	◊ ◊	27 (0.8)	539 (2.8)	◊ ◊	19 (0.7)	528 (3.2)	◊ ◊
Austria		53 (1.1)	549 (2.3)	◊ ◊	30 (0.9)	533 (3.2)	◊ ◊	17 (0.7)	524 (2.9)	◊ ◊
Lithuania		51 (1.1)	542 (2.0)	−1 (1.5)	36 (1.0)	532 (2.1)	−1 (1.4)	12 (0.6)	534 (2.6)	2 (0.9)
Belgium (French)		51 (1.3)	512 (2.9)	◊ ◊	32 (0.9)	496 (3.3)	◊ ◊	17 (0.8)	483 (4.4)	◊ ◊
Canada, Ontario	r	51 (1.0)	563 (3.4)	−2 (1.6)	32 (1.1)	552 (3.0)	1 (1.6)	18 (0.9)	547 (4.5)	1 (1.3)
France		50 (1.0)	535 (2.5)	−1 (1.4)	32 (0.8)	518 (2.6)	−2 (1.2)	19 (0.8)	505 (3.0)	3 (1.1) ▲
Spain	s	50 (1.2)	531 (3.0)	◊ ◊	33 (1.0)	512 (3.6)	◊ ◊	18 (0.7)	507 (4.5)	◊ ◊
Singapore		50 (0.7)	569 (3.0)	14 (1.1) ▲	33 (0.7)	552 (3.1)	−3 (0.9) ▼	17 (0.5)	546 (4.3)	−11 (0.9) ▼
Canada, Quebec	r	49 (1.4)	542 (3.2)	1 (1.9)	32 (1.2)	536 (3.5)	−2 (1.7)	18 (0.9)	518 (4.1)	1 (1.4)
Italy		49 (1.0)	564 (3.0)	2 (1.4)	32 (1.0)	547 (3.7)	−4 (1.2) ▼	19 (0.9)	535 (4.2)	2 (1.2)
Hungary		49 (1.0)	561 (3.2)	−2 (1.4)	37 (0.8)	546 (3.7)	1 (1.2)	14 (0.6)	544 (4.4)	1 (0.9)
Macedonia, Rep. of	r	48 (1.1)	459 (4.6)	7 (1.7) ▲	41 (0.9)	441 (4.5)	0 (1.4)	11 (0.9)	423 (8.4)	−7 (1.5) ▼
South Africa	r	48 (0.7)	322 (7.5)	◊ ◊	36 (0.5)	301 (5.2)	◊ ◊	16 (0.5)	276 (6.3)	◊ ◊
Israel		47 (1.4)	538 (5.0)	x x	39 (1.2)	520 (4.4)	x x	15 (0.7)	508 (6.5)	x x
Slovak Republic		46 (1.0)	544 (2.4)	−2 (1.5)	39 (1.0)	529 (3.2)	0 (1.4)	14 (0.8)	504 (7.7)	3 (1.0) ▲
Slovenia		45 (0.8)	530 (2.4)	4 (1.3) ▲	36 (0.8)	520 (2.6)	−6 (1.2) ▼	18 (0.6)	510 (3.3)	2 (0.9) ▲
Poland		45 (0.9)	529 (3.0)	◊ ◊	41 (0.8)	517 (2.8)	◊ ◊	14 (0.6)	504 (4.5)	◊ ◊
Qatar	r	44 (0.7)	361 (2.3)	◊ ◊	39 (0.7)	357 (2.3)	◊ ◊	17 (0.5)	348 (3.8)	◊ ◊
Kuwait	r	44 (1.2)	339 (5.4)	◊ ◊	36 (0.9)	337 (5.4)	◊ ◊	20 (0.9)	335 (5.9)	◊ ◊
Russian Federation		42 (1.0)	573 (3.7)	−7 (1.6) ▼	41 (0.8)	560 (3.9)	6 (1.4) ▲	17 (0.6)	557 (3.6)	1 (1.1)
Belgium (Flemish)		40 (0.9)	558 (2.2)	◊ ◊	33 (0.8)	545 (2.4)	◊ ◊	27 (1.0)	538 (2.4)	◊ ◊
Georgia		39 (1.3)	486 (3.4)	◊ ◊	44 (1.0)	467 (3.2)	◊ ◊	17 (1.2)	450 (7.6)	◊ ◊
Bulgaria		39 (1.4)	565 (4.2)	−14 (2.1) ▼	34 (1.1)	547 (4.7)	6 (1.6) ▲	27 (2.0)	532 (7.6)	8 (2.6) ▲
Hong Kong SAR		36 (0.8)	574 (2.3)	6 (1.4) ▲	39 (0.7)	561 (2.6)	10 (1.0) ▲	25 (0.6)	558 (3.4)	−15 (1.3) ▼
Chinese Taipei		35 (0.9)	547 (2.5)	◊ ◊	42 (0.7)	536 (2.5)	◊ ◊	23 (0.7)	520 (2.4)	◊ ◊
Moldova, Rep. of		35 (1.1)	509 (3.3)	−1 (1.9)	46 (1.1)	498 (3.5)	6 (1.7) ▲	20 (1.5)	491 (5.9)	−5 (1.9) ▼
Indonesia		31 (1.2)	413 (4.6)	◊ ◊	44 (1.3)	405 (4.5)	◊ ◊	25 (1.3)	395 (4.9)	◊ ◊
Romania		27 (1.1)	514 (4.7)	11 (1.6) ▲	43 (1.0)	497 (4.8)	11 (1.7) ▲	30 (1.4)	459 (8.0)	−22 (2.4) ▼
Morocco		24 (1.1)	346 (6.4)	– –	34 (1.4)	329 (5.3)	– –	42 (1.9)	308 (10.9)	– –
Iran, Islamic Rep. of		24 (1.1)	441 (4.9)	2 (1.5)	41 (1.1)	426 (3.4)	6 (1.8) ▲	35 (1.6)	402 (4.3)	−8 (2.2) ▼
England		x x	x x	x x	x x	x x	x x	x x	x x	x x
United States		– –	– –	– –	– –	– –	– –	– –	– –	– –
International Avg.		47 (0.2)	512 (0.6)		34 (0.2)	498 (0.6)		18 (0.2)	487 (0.9)	

▲ Percent in 2006 significantly higher

▼ Percent in 2006 significantly lower

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

Before we can proceed with the analysis, we will need to recode ASBHRRE into a new variable, NEWHRRE, that combines values 3 and 4 into the value 3 to create a category labeled "Twice a Month or Less". Exhibit 3.27 below shows an example of the SPSS code used to perform such recode.

Exhibit 3.26 Example SPSS Program to Recode Variables for Home Background Variable Analysis

```
get file = " "<datpath>ashmerged.sav".

* Compute new variable newHRRE from ASBHRRE .

recode asbhrre (1=1) (2=2) (3,4=3) into newHRRE.

variable labels newHRRE 'Recoded ASBHRRE Readhome for enjoyment'.

value labels newHRRE
   1 'Every/Almost every day'
   2 'Once/Twice per week'
   3 'Twice a month or less'.

save outfile = "<datpath>ashmerged.sav".
```

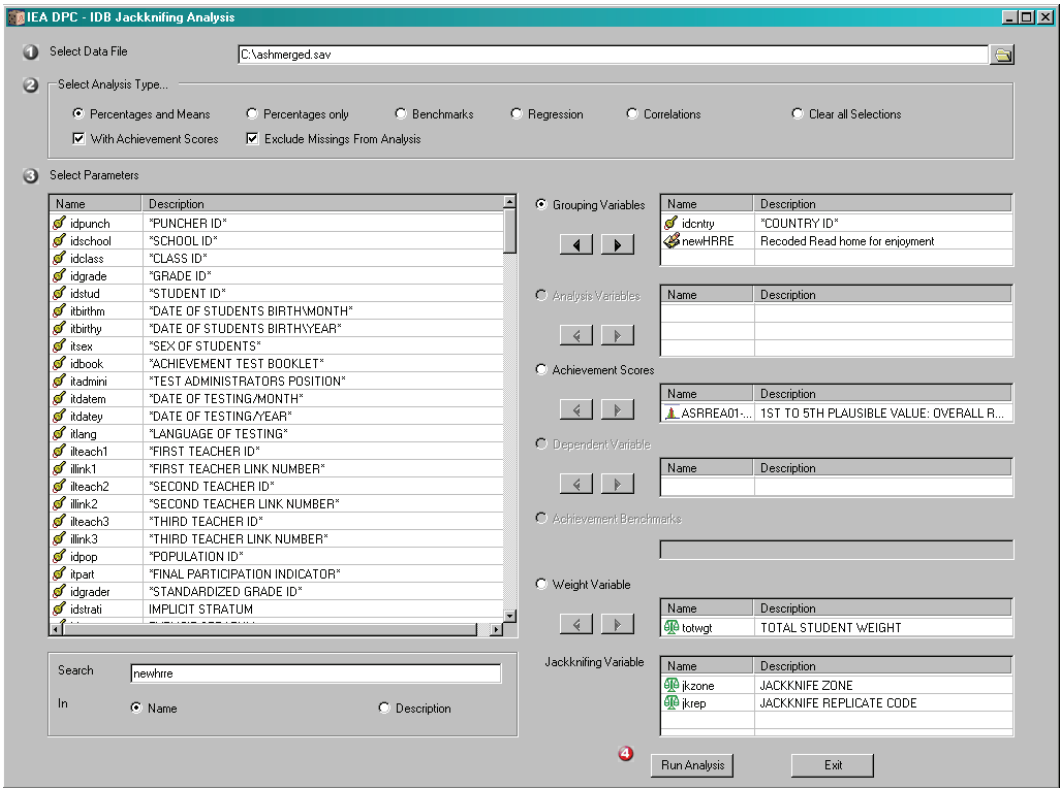
Once the recode is completed, we can open the analysis module of the IDB Analyzer and select the recoded data file ASHMERGED.SAV. Exhibit 3.27 below shows the analysis module display after following the analysis steps:

1) Select **Percentages and Means** as the analysis type. By default, the program will exclude those with missing grouping variables from the analysis.
2) Check the box next to **With Achievement Scores**.
3) Add the variable NEWHRRE as a second **Grouping Variable**.
4) Next, define the achievement scores to be used for the analysis. To activate this section, you will need to press the radio button **Achievement Scores**. This time you will need to select variable ASRREA01-05 from the list of variables and move it to the analysis variables field by pressing the right arrow button in this section.

5) The **Weight Variable** is automatically defined by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. Additionally the **Jackknifing Variables** JKZONE and JKREP are selected.

6) Press the button **Run Analysis** in order to perform your analysis. The program will alert you as soon as the analysis has been completed and the results will be displayed in the output window.

Exhibit 3.27 Example Analysis Setup of Home Background Variable Analysis



The results of the analysis are presented in Exhibit 3.28. In this example, each country's results are presented on three lines, one for each value of the NEWHRRE variable. The countries are identified in the first column and the second column describes the categories of NEWHRRE being reported. All other columns are identical in nature as in the previous example.

From the first three lines in Exhibit 3.28, we see that in Austria, 52.91 percent of students have parents who read for enjoyment every day or almost every day, 30.18 percent have parents who read for enjoyment once or twice a week, and

16.91 percent twice a month or less. We also see that the estimated mean reading achievement of students whose parents read for enjoyment every day or almost every day is 548.84 (with a standard error of 2.30), the estimated mean achievement of students whose parents read for enjoyment once or twice a week is 532.70 (standard error of 3.16), and it is 523.95 (standard error of 2.95) for students whose parents read for enjoyment twice a month or less.

Exhibit 3.28 Output of Example Home Background Variable Analysis

COUNTRY ID	Recoded Read home for enjoyment	N of Cases	Sum of TOTWGT	Percent	Percent (s.e.)	ASRREA0 (Mean)	ASRREA0 (s.e.)	Std.Dev.	Std.Dev. (s.e.)
AUSTRIA	Every/Almost every day	2573	41727	52.91	1.06	548.84	2.30	61.13	1.72
	Once/Twice per week	1440	23805	30.18	.86	532.70	3.16	64.36	2.00
	Twice a month or less	787	13337	16.91	.66	523.95	2.95	63.87	2.24
BULGARIA	Every/Almost every day	1540	23568	38.84	1.44	564.81	4.24	76.83	3.30
	Once/Twice per week	1294	20638	34.01	1.14	547.33	4.69	78.55	2.82
	Twice a month or less	871	16478	27.15	1.98	532.46	7.62	84.23	3.44
				.....					
SPAIN	Every/Almost every day	1287	119919	49.55	1.23	530.99	3.00	68.37	2.14
	Once/Twice per week	825	79606	32.89	1.03	512.14	3.65	69.72	3.28
	Twice a month or less	443	42493	17.56	.74	507.20	4.51	68.63	3.29
SWEDEN	Every/Almost every day	2679	61207	64.22	.88	555.81	2.44	61.33	1.19
	Once/Twice per week	862	20167	21.16	.77	546.35	2.95	63.40	2.43
	Twice a month or less	563	13929	14.62	.74	539.87	4.56	66.95	3.82
TRINIDAD AND TOBAGO	Every/Almost every day	2029	8986	58.81	.92	441.41	5.08	101.52	3.23
	Once/Twice per week	1126	4755	31.12	.97	436.29	6.08	102.91	3.75
	Twice a month or less	363	1539	10.07	.66	433.75	9.22	112.47	6.12

References

IEA. (2007). *International database analyzer* (version 1.4.0.5). Hamburg, Germany: IEA Data Processing and Research Center.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M., & Foy, P. (2007). *PIRLS 2006 international report: IEA's progress in international reading literacy study in primary schools in 40 countries*. Chestnut Hill, MA: Boston College.

SPSS Inc. (2005). *SPSS for Windows* (version 14.0). Chicago, IL: SPSS Inc.

Chapter 4

Performing Analyses with the PIRLS Data Using SPSS

4.1 Overview

Whereas Chapter 3 describes how to conduct simple analyses of the PIRLS 2006 international database by using IEA's IDB Analyzer in conjunction with SPSS (SPSS, 2005), this chapter describes how to analyze the data using SPSS directly, making use of SPSS programs and macros provided with the database. All analyses in this chapter will be performed by SPSS programs with the SPSS data files as input. The analyses presented here are simple in nature, and are designed primarily to familiarize you with the various data files and their structure, as well as the variables to be used in most analyses. In general, the programs compute percentages of students in specified subgroups, average reading achievement in those subgroups, and the proper standard errors for these statistics. Additionally, some programs compute regression coefficients and their standard errors.

The example analyses, using student, home, teacher and school data, replicate some of the analyses that are included in the *PIRLS 2006 International Report* (Mullis, Martin, Kennedy, & Foy, 2007). You are encouraged to practice analyzing the PIRLS data by replicating some of the exhibits presented in the international report.

All SPSS programs presented in this chapter also are available in the PIRLS database. They can be adapted to perform a variety of analyses, provided you have some basic knowledge of the SPSS language. With a little experience and some practice with these programs, you will be able to make the necessary modifications to obtain the desired results. The example SPSS programs invoke SPSS macros that will be described in this chapter. Although you will be expected to modify the example programs, there is no need to make any changes within the SPSS macros.

4.2 SPSS Programs and Macros

The PROGRAMS subfolder on the DVD that accompanies this User Guide includes a number of SPSS programs that may be used to process the SPSS data files, compute survey results, and carry out example analyses. This chapter gives detailed instructions on how to adapt and make use of them. The following programs are available:

ASASCRR2.SPS

This SPSS program that can be used to convert student responses to a test item to its correctness score (e.g., "0" for incorrect, "1" for partially correct, etc.).

JOIN.SPS

This SPSS program combines files of the same type from more than one country.

JACKGEN.SPS (and SAMPLEJACKGEN.SPS)

This SPSS macro program is used to compute weighted percentages of students within defined subgroups, along with their means on a specified continuous variable. This macro generates replicate weights and computes standard errors using the jackknife repeated replication (JRR) methodology. The analysis variable can be any continuous variable. When computing mean achievement scores based on plausible values, you will need to use the macro JACKPV.SPS.

JACKPV.SPS (and SAMPLEJACKGPV.SPS)

This SPSS macro program is used to compute weighted percentages of students within defined subgroups, along with their mean achievement using the available plausible values. This macro generates replicate weights and computes standard errors using the jackknife repeated replication (JRR) and multiple imputation methodologies. This macro should be used when multiple plausible values are used in an analysis.

JACKREG.SPS (and SAMPLEJACKREG.SPS)

This SPSS macro program is used to compute weighted regression coefficients and their standard errors within defined subgroups. This macro can be used with any analysis variable, but is not appropriate for analyzing achievement with plausible values.

JACKREGP.SPS (and SAMPLEJACKREGP.SPS)

This SPSS macro program is used to compute weighted regression coefficients and their standard errors within defined subgroups when using plausible values as achievement scores.

Each of the four SPSS macros above has a corresponding sample program that calls its respective macro and prints out the results. These sample programs are discussed later in this chapter.

EXAMPLE1.SPS, EXAMPLE2.SPS, EXAMPLE3.SPS, EXAMPLE4.SPS, EXAMPLE5.SPS

These are the programs used in the example analyses presented later in this chapter.

4.3 Scoring the Items

Two types of items were administered as part of the PIRLS assessment. There were multiple-choice items, where students were asked to select one of four options as the correct response. Numbers 1 through 4 are used to represent response options A through D, respectively, in the achievement data files. There also were constructed-response items, where students were asked to construct a written response to a question, rather than choosing an answer from a list of options. Constructed-response items were worth a total of one, two, or three points. Scorers from the national centers were trained to use the scoring guides described in Chapter 2 to score the answers to these questions. The numbers 0 through 3 are used to represent the scored responses to these items and also represent their point values. For both types of items, special codes are set aside to represent missing data either as "Not Administered", "Omitted", or "Not Reached".

Responses to multiple-choice items may need to be converted to their appropriate score levels ("1" for correct and "0" for incorrect), as well as responses coded to the special missing codes, in order to carry out specific item-level analyses.

For this purpose, the DVD includes an SPSS program (ASASCRR2.SPS) which allows you to recode the items from the achievement data files to their score level. The program consists of a macro called SCOREIT and the necessary call to this macro so that all the items in the specified data files are scored. This macro will convert the response option codes for multiple-choice items to dichotomous score levels (0 or 1) based on each item's scoring key. It will also convert the special missing codes as either incorrect (0) or missing. By default, the not administered response code is left as missing and the omitted and not reached response codes as incorrect. These default settings can be modified within the SCOREIT macro, depending on the requirements of your item-level analyses. For example, not reached responses were treated as missing for the purpose of calibrating the PIRLS items, whereas they were treated as incorrect when scoring the results of individual countries and deriving achievement scores for students.

To use the SCOREIT macro, you will need to adapt the program code in the ASASCRR2 program. You should do the following:

1) Open the SPSS program file ASASCRR2.SPS
2) Specify the path where the SPSS data files are located in the "LET! LIBDAT!" statement
3) List all the countries of interest in the parameter "COUNTRY" By default, all PIRLS 2006 countries are listed.
4) Submit the edited code for processing

The program recodes the items and saves the results in SPSS data files that have "ASC" instead of "ASA" as the first three characters. If you would like not reached responses to be treated as missing rather than incorrect, you should replace the following statement (which appears twice in the program):

```
(!NR    = 0      )
```

with this statement:

```
(!NR    = SYSMIS )
```

Exhibit 4.1 shows a condensed version of the SPSS program that scores the items.

Exhibit 4.1 Example of ASASCRR2 Program For Converting Item Response Codes to Their Score Level

```
DEFINE SCOREIT (TYPE  = !CHAREND('/') /
                ITEM  = !CHAREND('/') /
                RIGHT = !CHAREND('/') /
                NR    = !CHAREND('/') /
                NA    = !CHAREND('/') /
                OM    = !CHAREND('/') /
                OTHER = !CHAREND('/') ) .

   .  .  .

!ENDDEFINE .

DEFINE DOIT (COUNTRY = !CHAREND('/') ) .

!LET !LIBDAT = !UNQUOTE("D:\PIRLS2006\Data\SPSS_Data\") .

   .  .  .

SCOREIT TYPE = MC / ITEM = < List of multiple-choice items where A is correct > .
SCOREIT TYPE = MC / ITEM = < List of multiple-choice items where B is correct > .
SCOREIT TYPE = MC / ITEM = < List of multiple-choice items where C is correct > .
SCOREIT TYPE = MC / ITEM = < List of multiple-choice items where D is correct > .
SCOREIT TYPE = CR / ITEM = < List of constructed-response items > .

!ENDDEFINE .

DOIT COUNTRY = < List of PIRLS 2006 countries > .
```

4.4 Joining the Files

The PIRLS database DVD contains separate data files for each country. The DVD provides an SPSS program called JOIN.SPS that joins individual country data files of a particular type into a single aggregated data file, facilitating joint analyses involving more than one country. This program, however, can only join SPSS data files of the same type, (e.g., all ASG data files). The JOIN program can be used for the following data file types: ACG, ASA, ASC, ASG, ASH, AST, and ATG.

To create an SPSS data file with more than one country's data, you should do the following:

1) Open the SPSS program file JOIN.SPS
2) At the beginning of the program, specify the data file type in the parameter "TYPE"
3) Specify the sorting variables in the parameter "SORTVARS"
   By default the variables IDCNTRY and IDSTUD are specified for the ASA, ASC, ASG and ASH data files. Consult Exhibit 2.10 to determine which sorting variables need to be specified for each data file type.

4) Specify the path where the SPSS data files are located in the "PATH" statement

5) List all the countries of interest in the parameter "COUNTRY" By default, all PIRLS 2006 countries are listed

6) Submit the edited code for processing

An example of the JOIN program is displayed in Exhibit 4.2. It joins the student background data files of all countries. All country data files are located in the folder "D:\PIRLS2006\Data\SPSS_Data" for the sake of this example. The resulting data file, ASGALLR2, will be saved in this folder as well. The joined data file will be sorted by the identification variables IDCNTRY and IDSTUD.

Exhibit 4.2 Example of JOIN Program Used to Join SPSS Data Files for More Than One Country

```
DEFINE JOIN (COUNTRY = !CHAREND('/') /
             TYPE    = !CHAREND('/') /
             SORTVAR = !CHAREND('/') /
             PATH    = !CHAREND('/') ) .

ADD FILES
   !DO !CTY !IN (!COUNTRY)
      !CONCAT('  FILE = "',!PATH,!TYPE,!CTY,'r2.sav"') /
   !DOEND .

SORT CASES BY !SORTVAR .

SAVE OUTFILE = !QUOTE(!CONCAT(!PATH,!TYPE,'allr2.sav')) .

!ENDDEFINE .

JOIN COUNTRY = < List of PIRLS 2006 countries > /
     TYPE    = ASG /
     SORTVAR = IDCNTRY IDSTUD /
     PATH    = D:\PIRLS2006\Data\SPSS_Data\ .
```

4.5 SPSS Macros to Compute Percentages, Means, Regression Coefficients, and their Standard Errors

This section describes the four SPSS macros needed to compute specific statistics with their correct standard errors, along with sample SPSS programs to demonstrate their use. Each SPSS macro serves a specific analytical purpose. Making use of these macros for analyzing the PIRLS 2006 data is the best way to ensure the analyses will be done properly. Sampling weights must be used to analyze the PIRLS data, and standard errors must be computed using the jackknife repeated replication (JRR) method. Furthermore, all achievement scores are based on sets of plausible values that take into account the measurement error arising from the test design and the IRT scaling methodology. These plausible values will be processed by the macros to produce accurate results.

The sample SPSS programs presented in this section all use an SPSS data file called ASGALLR2 as input, which contains the student background data files of all participating countries. In all sample programs, <datpath> must be edited to specify the folder where the ASGALLR2 file is located.

Computing Means and Their Standard Errors (JACKGEN)

The JACKGEN macro is used to compute percentages and means of continuous variables with their JRR standard errors. We will demonstrate its use with a sample SPSS program that calls the macro JACKGEN to compute the percentages of students within specified subgroups and their mean on a variable of choice. The macro also computes the appropriate standard errors for the percentages and means. However, this macro is not appropriate for analyzing achievement means based on plausible values; the JACKPV macro should be used for this purpose.

The JACKGEN macro is a self-contained program, located in the program file JACKGEN.SPS, and should not be modified. It essentially computes sets of replicate weights using the sampling and weighting variables, aggregates the data by subgroups using the replicate weights, and then computes and stores the desired statistics in an SPSS working file called FINAL.

The macro JACKGEN is included in an SPSS program file by issuing the following command:

```
INCLUDE "<macpath>JACKGEN.SPS" .
```

where <macpath> points to the specific FOLDER where the SPSS macro program JACKGEN.SPS is located. The macro requires that several parameters be specified as input when it is invoked. These parameters are:

WGT	The sampling weight to be used in the analysis. Generally, TOTWGT should be used. TCHWGT should be used when analyzing teacher data from the AST and ATG files.
JKZ	The variable that captures the assignment of cases to sampling zones. The name of this variable in all PIRLS data files is JKZONE.
JKR	The variable that captures whether the case is to be dropped or have its weight doubled for each set of replicate weights. The name of this variable in all PIRLS data files is JKREP.
NJKZ	The number of replicate weights to be generated when computing the JRR standard errors. The value of NJKZ should be set to 75, the maximum possible value across all participating countries.
CVAR	The list of variables that are to be used to define the subgroups. The list can consist of one or more variables. We recommend that you always include IDCNTRY as the first classification variable.
DVAR	The variable for which means are to be computed. Only one variable can be listed and it should be a continuous variable. Plausible values of achievement scores should not be specified here.
INFILE	The name of the data file that contains the data being analyzed. If the path location is included as part of the file name, the name of the file must be enclosed in quotes. It is important to emphasize that this data file must include only those cases that are of interest in the analysis. If you want to have specific cases excluded from the analysis, for example students with missing data, this should be done prior to invoking the macro.

The JACKGEN macro is invoked by an SPSS program using the conventional SPSS notation for invoking macros. This involves listing the macro name followed by the list of parameters in parenthesis, each separated by a forward slash. For example, the macro invoked using the following statement:

```
JACKGEN INFILE = ASGALLR2       /
        CVAR   = IDCNTRY ITSEX  /
        DVAR   = ASDAGE         /
        NJKZ   = 75             /
        JKZ    = JKZONE         /
        JKR    = JKREP          /
        WGT    = TOTWGT         .
```

will compute the mean age (ASDAGE) for boys and girls (ITSEX) and their standard errors within each country (IDCNTRY), using the weighting variable TOTWGT. It will also compute the percentages of boys and girls and their standard errors within each country. The data will be read from the data file ASGALLR2 and the standard errors will be computed based on 75 replicate weights.

The results of the JACKGEN macro are stored in an SPSS working file called FINAL which is stored in the default folder being used by SPSS. The following variables are contained in this results file:

Classification Variables

All classification variables are kept in the results file. In our example, there are two classification variables: IDCNTRY and ITSEX. There is one record in the results file for each subgroup defined by the categories of the classification variables.

N

This variable contains the number of valid cases for each subgroup defined by the classification variables. In our example, it is the number of boys and girls with valid data in each country's sample.

Weight Variable

The weight variable contains the sum of weights within each subgroup defined by the classification variables. In our example, this variable is called TOTWGT since we specified TOTWGT as the weighting variable. When TOTWGT is used (or TCHWGT when analyzing teacher data), this variable will be an estimate of the total population within each subgroup.

MNX

This variable contains the estimated means of the specified analysis variable by subgroup.

MNX_SE

This variable contains the JRR standard errors of the estimated means by subgroup.

PCT

This variable contains the estimated percentages of students in each subgroup for the last classification variable listed. In our example it is the percentage of boys and girls within each country.

PCT_SE

This variable contains the JRR standard errors of the estimated percentages.

The contents of the FINAL file can be printed using the SPSS REPORT procedure. The sample SPSS program that invokes the JACKGEN macro and a printout of the results are presented in Exhibit 4.3. This program is available on the database DVD in the file called SAMPLEJACKGEN.SPS. It produces the mean ages for boys and girls in all countries, although the exhibit shows the results only for the first five countries.

Exhibit 4.3 Sample SPSS Program Invoking the SPSS Macro JACKGEN and Output File

```
GET FILE = "<datpath>ASGALLR2.SAV" .

SELECT IF (NOT(MISSING(ITSEX)) AND NOT(MISSING(ASDAGE))) .

VALUE LABELS
 IDCNTRY < list country formats > .

VALUE LABELS
 ITSEX
     1 'GIRL'
     2 'BOY' .

SAVE OUTFILE = ASGALLR2 .

INCLUDE "<macpath>JACKGEN.SPS" .

JACKGEN INFILE = ASGALLR2       /
        CVAR   = IDCNTRY ITSEX  /
        DVAR   = ASDAGE         /
        NJKZ   = 75             /
        JKZ    = JKZONE         /
        JKR    = JKREP          /
        WGT    = TOTWGT         .

PRINT FORMATS IDCNTRY ITSEX N (F6.0)
              TOTWGT (F10.0) MNX MNX_SE PCT PCT_SE (F6.2) .

REPORT FORMAT = LIST AUTOMATIC
  / VAR = IDCNTRY (LABEL) ITSEX (LABEL) N TOTWGT MNX MNX_SE PCT PCT_SE .
```

COUNTRY ID	*SEX OF STUDENTS*	N	TOTWGT	MNX	MNX_SE	PCT	PCT_SE
AUSTRIA	GIRL	2501	41134	10.30	.01	49.46	.70
AUSTRIA	BOY	2566	42035	10.37	.01	50.54	.70
BULGARIA	GIRL	1906	31226	10.87	.02	49.42	1.01
BULGARIA	BOY	1951	31957	10.88	.02	50.58	1.01
CHINESE TAIPEI	GIRL	2187	145288	10.09	.01	47.72	.51
CHINESE TAIPEI	BOY	2402	159201	10.10	.01	52.28	.51
DENMARK	GIRL	2054	32633	10.85	.01	51.61	.91
DENMARK	BOY	1947	30598	10.97	.01	48.39	.91
FRANCE	GIRL	2121	358440	10.00	.02	48.47	.73
FRANCE	BOY	2281	381144	10.04	.02	51.53	.73

From the first two lines of the results shown in Exhibit 4.3, we see that in Austria there are 2,501 girls in the sample representing 41,134 girls in the whole population. The mean age for girls at the fourth grade in Austria is estimated to be 10.30 with a standard error of 0.01. Girls made up 49.46 percent of Austria's fourth grade student population. Conversely, Austria sampled 2,566 boys representing 42,035 boys in the whole population. The estimated mean age for boys at the fourth grade in Austria is 10.37 with a standard error of 0.01. Boys made up 50.54 percent of Austria's fourth grade student population.

Computing Achievement Means and Their Standard Errors (JACKPV)

The JACKPV macro computes percentages and mean achievement scores based on plausible values with their JRR standard errors. A second sample program demonstrates the use of the JACKPV macro, which computes the percentages of students within specified subgroups and their mean achievement scores. The SPSS macro also computes the appropriate standard errors for those percentages and achievement means.

The JACKPV macro is a self-contained program, located in the program file JACKPV.SPS, and should not be modified. It essentially computes sets of replicate weights using the sampling and weighting variables, aggregates the data by subgroups using the replicate weights, and then computes and stores the desired statistics in an SPSS working file called FINAL. The macro aggregates data across all plausible values to obtain the correct results.

The SPSS macro JACKPV is included in an SPSS program by issuing the following command:

```
INCLUDE "<macpath>JACKPV.SPS" .
```

where <macpath> points to the specific FOLDER where the SPSS macro program JACKPV.SPS is located. The macro requires that several parameters be specified as input when it is invoked. These parameters are:

WGT	The sampling weight to be used in the analysis. Generally, TOTWGT should be used. TCHWGT should be used when analyzing teacher data from the AST and ATG files.
JKZ	The variable that captures the assignment of cases to sampling zones. The name of this variable in all PIRLS data files is JKZONE.
JKR	The variable that captures whether the case is to be dropped or have its weight doubled for each set of replicate weights. The name of this variable in all PIRLS data files is JKREP.
NJKZ	The number of replicate weights to be generated when computing the JRR standard errors. The value of NJKZ should be set to 75, the maximum possible value across all participating countries.
CVAR	The list of variables that are to be used to define the subgroups. The list can consist of one or more variables. We recommend that you

always include IDCNTRY as the first classification variable.

ROOTPV — The variable root used to identify the set of plausible values for the achievement score of interest. It corresponds to the first 7 characters of the plausible values variable name. For example, the root of the overall reading plausible values is ASRREA0, the root of the literary purpose plausible values is ASRLIT0.

NPV — The number of plausible values that will be used for the analysis. Generally, you will want to use all five plausible values for analysis.

INFILE — The name of the data file that contains the data being analyzed. If the path location is included as part of the file name, the name of the file must be enclosed in quotes. It is important to emphasize that this data file must include only those cases that are of interest in the analysis. If you want to have specific cases excluded from the analysis, for example students with missing data, this should be done prior to invoking the macro.

The JACKPV macro is invoked by an SPSS program using the conventional SPSS notation for invoking macros. This involves listing the macro name followed by the list of parameters in parenthesis, each separated by a forward slash. For example, the macro invoked using the following statement:

```
JACKPV INFILE = ASGALLR2      /
       CVAR   = IDCNTRY ITSEX /
       ROOTPV = ASRREA0       /
       NPV    = 5             /
       NJKZ   = 75            /
       JKZ    = JKZONE        /
       JKR    = JKREP         /
       WGT    = TOTWGT        .
```

will compute the mean reading achievement (ASRREA01 through ASRREA05) for boys and girls (ITSEX) within each country (IDCNTRY) and their standard errors, using the weighting variable TOTWGT. The macro uses all five plausible values to compute these statistics. It will also compute the percentages of boys and girls within each country, and their standard errors. The data will be read from the data file ASGALLR2 and the standard errors will be computed based on 75 replicate weights.

The results of the JACKPV macro are stored in an SPSS working file called FINAL which is stored in the default folder being used by SPSS. The following variables are contained in this results file:

Classification Variables

All classification variables are kept in the results file. In our example, there are two classification variables: IDCNTRY and ITSEX. There is one record in the results file for each subgroup defined by the categories of the classification variables.

N

This variable contains the number of valid cases for each subgroup defined by the classification variables. In our example, it is the number of boys and girls with valid data in each country's sample.

Weight Variable

The weight variable contains the sum of weights within each subgroup defined by the classification variables. In our example, this variable is called TOTWGT since we specified TOTWGT as the weighting variable. When TOTWGT is used (also TCHWGT when analyzing teacher data), this variable will be an estimate of the total population within each subgroup.

MNPV

This variable contains the estimated mean achievement by subgroup, based on the plausible values.

MNPV_SE

This variable contains the JRR standard errors of the estimated mean achievement by subgroup, based on the plausible values.

PCT

This variable contains the estimated percentages of students in each subgroup for the last classification variable listed. In our example it is the percentage of boys and girls within each country.

PCT_SE

This variable contains the JRR standard errors of the estimated percentages.

The contents of the FINAL file can be printed using the SPSS REPORT procedure. The sample SPSS program that invokes the JACKPV macro and a printout of the results are presented in Exhibit 4.4. This program is available on the database DVD in the file called SAMPLEJACKPV.SPS. It produces the mean reading achievement for boys and girls in all countries, although Exhibit 4.4 gives the results only for the first five countries.

Exhibit 4.4 Sample SPSS Program Invoking the SPSS Macro JACKPV and Output File

```
GET FILE = "<datpath>ASGALLR2.SAV" .

SELECT IF NOT(MISSING(ITSEX)) .

VALUE LABELS
 IDCNTRY < list country formats > .

VALUE LABELS
 ITSEX
     1 'GIRL'
     2 'BOY' .

SAVE OUTFILE = ASGALLR2 .

INCLUDE "<macpath>JACKPV.SPS" .

JACKPV INFILE = ASGALLR2      /
       CVAR   = IDCNTRY ITSEX /
       ROOTPV = ASRREA0       /
       NPV    = 5             /
       NJKZ   = 75            /
       JKZ    = JKZONE        /
       JKR    = JKREP         /
       WGT    = TOTWGT        .

PRINT FORMATS IDCNTRY ITSEX N (F6.0)
              TOTWGT (F10.0) MNPV MNPV_SE PCT PCT_SE (F6.2) .

REPORT FORMAT = LIST AUTOMATIC
  / VAR = IDCNTRY (LABEL) ITSEX (LABEL) N TOTWGT MNPV MNPV_SE PCT PCT_SE .
```

COUNTRY ID	*SEX OF STUDENTS*	N	TOTWGT	MNPV	MNPV_SE	PCT	PCT_SE
AUSTRIA	GIRL	2501	41134	543.26	2.32	49.46	.70
AUSTRIA	BOY	2566	42035	533.44	2.64	50.54	.70
BULGARIA	GIRL	1909	31329	557.52	4.41	49.44	1.01
BULGARIA	BOY	1954	32042	536.78	5.00	50.56	1.01
CHINESE TAIPEI	GIRL	2187	145288	542.08	2.22	47.72	.51
CHINESE TAIPEI	BOY	2402	159201	529.25	2.27	52.28	.51
DENMARK	GIRL	2054	32633	552.94	2.84	51.61	.91
DENMARK	BOY	1947	30598	539.31	2.72	48.39	.91
FRANCE	GIRL	2121	358440	527.31	2.42	48.46	.73
FRANCE	BOY	2282	381280	516.22	2.41	51.54	.73

From the first two lines of the results presented in Exhibit 4.4, we see that in Austria the mean reading achievement of girls is estimated to be 543.26 with a standard error of 2.32. The mean reading achievement of boys in Austria is estimated to be 533.44 with a standard error of 2.64.

Computing Regression Coefficients and Their Standard Errors (JACKREG)

The JACKREG macro is used to perform a multiple linear regression between a dependent variable and a set of independent variables. A third sample program demonstrates the use of the JACKREG macro, which computes the regression coefficients and their JRR standard errors. This macro is not appropriate for regression analyses using achievement scores and their plausible values as dependent variables. The JACKREGP macro should be used for this purpose.

The JACKREG macro is a self-contained program, located in the program file JACKREG.SPS, and should not be modified. It essentially computes sets of replicate weights using the sampling and weighting variables, performs a multiple linear regression by subgroups and aggregates the data using the replicate weights, and then computes and stores the desired statistics in an SPSS working file called REG.

The SPSS macro JACKREG is included in an SPSS program file by issuing the following command:

```
INCLUDE "<macpath>JACKREG.SPS" .
```

where <macpath> points to the specific FOLDER where the SPSS macro program JACKREG.SPS is located. The macro requires that several parameters be specified as input when it is invoked. These parameters are:

WGT	The sampling weight to be used in the analysis. Generally, TOTWGT should be used. TCHWGT should be used when analyzing teacher data from the AST and ATG files.
JKZ	The variable that captures the assignment of cases to sampling zones. The name of this variable in all PIRLS data files is JKZONE.
JKR	The variable that captures whether the case is to be dropped or have its weight doubled for each set of replicate weights. The name of this variable in all PIRLS data files is JKREP.

NJKZ — The number of replicate weights to be generated when computing the JRR standard errors. The value of NJKZ should be set to 75, the maximum possible value across all participating countries.

CVAR — The list of variables that are to be used to define the subgroups. The list can consist of one or more variables. We recommend that you always include IDCNTRY as the first classification variable.

XVAR — The list of independent variables used as predictors in the regression model. The independent variables can be either continuous or categorical, such as ITSEX for example.

DVAR — The dependent variable to be predicted by the list of independent variables specified in XVAR. Only one variable can be listed and plausible values of achievement scores should not be specified here.

INFILE — The name of the data file that contains the data being analyzed. If the path location is included as part of the file name, the name of the file must be enclosed in quotes. It is important to emphasize that this data file must include only those cases that are of interest in the analysis. If you want to have specific cases excluded from the analysis, for example students with missing data, this should be done prior to invoking the macro.

The JACKREG macro is invoked by an SPSS program using the conventional SPSS notation for invoking macros. This involves listing the macro name followed by the list of parameters in parenthesis, each separated by a forward slash. For example, the macro invoked using the following statement:

```
JACKREG INFILE = ASGALLR2 /
        CVAR   = IDCNTRY  /
        XVAR   = REGSEX   /
        DVAR   = BOOK     /
        NJKZ   = 75       /
        JKZ    = JKZONE   /
        JKR    = JKREP    /
        WGT    = TOTWGT   .
```

will perform a linear regression with the variable REGSEX as a predictor of the number of books in the home (BOOK), using the weighting variable TOTWGT. It will compute the regression coefficients and their standard errors. The data will be read from the data file ASGALLR2 and the standard errors will be computed based on 75 replicate weights.

The results of the JACKREG macro are stored in an SPSS working file called REG which is stored in the default folder being used by SPSS. The following variables are contained in this results file:

Classification Variables

All classification variables are kept in the results file. In our example, there is a single classification variable IDCNTRY. There is one record in the results file for each subgroup defined by the categories of the classification variables.

N

This variable contains the number of valid cases for each subgroup defined by the classification variables. In our example, it is the number of students with valid data in each country's sample.

MULT_RSQ

The squared multiple correlation (R^2) coefficient for the regression model applied in each subgroup.

SS_RES, SS_REG, SS_TOTAL

The residual, regression, and total weighted sums of squares for the regression model applied in each subgroup.

Regression Coefficients and Standard Errors (B## and B##.SE)

The regression coefficients for the predictor variables and the intercept with their respective standard errors. The regression coefficients are numbered sequentially, starting with zero (B00) for the intercept, and based on the order of the predictor variables are specified in the parameter XVAR.

The contents of the REG file can be printed using the SPSS REPORT procedure. The sample SPSS program that invokes the JACKREG macro and a printout of the results are presented in Exhibit 4.5. This program is available on the database DVD in the file called SAMPLEJACKREG.SPS. It performs a linear regression in each country, with the variable REGSEX as a predictor of the number of books in the home (BOOK). The exhibit displays the results for the first five countries.

The regression performed by the sample program uses variables that are transformed within its code. The dependent variable BOOK is derived from the

categorical variable ASBGBOOK, where BOOK is given the midpoint value of the intervals specified by each category of ASBGBOOK. The independent variable REGSEX is a dummy-coded version of ITSEX, such that the value zero represents the girls and the value one represents the boys. By performing this dummy coding, the intercept (B00) will be the estimated number of books in the homes of girls, whereas the regression coefficient B01 will be the estimated number of additional books in the homes of boys. This will also allow us to perform a t-test to determine if the number of books in the home is significantly different between girls and boys.

Exhibit 4.5 Sample SPSS Program Invoking the SPSS Macro JACKREG and Output File

```
GET FILE = "<datpath>ASGALLR2.SAV" .

SELECT IF (NOT(MISSING(ITSEX)) AND NOT(MISSING(ASBGBOOK))) .

RECODE ITSEX (1=0) (2=1) INTO REGSEX .

RECODE ASBGBOOK (1=5) (2=18) (3=63) (4=151) (5=251) INTO BOOK .

VALUE LABELS
 IDCNTRY < list country formats > .

SAVE OUTFILE = ASGALLR2 .

INCLUDE "<macpath>JACKREG.SPS" .

JACKREG INFILE = ASGALLR2 /
        CVAR   = IDCNTRY  /
        XVAR   = REGSEX   /
        DVAR   = BOOK     /
        NJKZ   = 75       /
        JKZ    = JKZONE   /
        JKR    = JKREP    /
        WGT    = TOTWGT   .

PRINT FORMATS IDCNTRY N (F6.0) MULT_RSQ (F5.3) SS_TOTAL SS_REG (F10.0)
              B00 B00.SE B01 B01.SE (F6.2) .

REPORT FORMAT = LIST AUTOMATIC
  / VAR = IDCNTRY (LABEL) N MULT_RSQ SS_TOTAL SS_REG B00 B00.SE B01 B01.SE .


*COUNTRY ID*     N    MULT_RSQ  SS_TOTAL    SS_REG    B00   B00.SE  B01   B01.SE
_____________ ______ ________ __________ __________ ______ ______ ______ ______

AUSTRIA          4923    .001   482333720     363379  82.81   2.46  -4.24   3.04
BULGARIA         3723    .000   483282536        871  85.55   3.97   -.24   3.99
CHINESE TAIPEI   4465    .000  2068367933     453550  83.65   1.96  -2.48   2.69
DENMARK          3917    .000   418862800       8445 102.15   2.61    .74   3.44
FRANCE           4142    .000  5407360619     399107 106.40   2.99   1.52   2.70
```

From the first line of the results displayed in Exhibit 4.5, we see that in Austria the estimated average number of books in the homes of fourth grade girls (B00) is 82.81, with a standard error of 2.46. The Austrian fourth grade boys have an estimated 4.24 fewer books in the home (B01) than girls. With an estimated standard error of 3.04, this difference is not statistically significant at a 95% confidence level.

Computing Regression Coefficients and Their Standard Errors with Achievement Scores (JACKREGP)

The JACKREGP macro is used to perform a multiple linear regression between a set of plausible values as the dependent variable and a set of independent variables. It computes the regression coefficients and their JRR standard errors, taking into account the plausible values. We present a fourth sample program to demonstrate the use of the JACKREGP macro.

The JACKREGP macro is a self-contained program, located in the program file JACKREGP.SPS, and should not be modified. It essentially computes sets of replicate weights using the sampling and weighting variables, performs a multiple linear regression by subgroups and aggregates the data using the replicate weights, and then computes and stores the desired statistics in an SPSS working file called REG.

The SPSS macro JACKREGP is included in an SPSS program file by issuing the following command:

```
INCLUDE "<macpath>JACKREG.SPS" .
```

where <macpath> points to the specific FOLDER where the SPSS macro program JACKREGP.SPS is located. The macro requires that several parameters be specified as input when it is invoked. These parameters are:

WGT	The sampling weight to be used in the analysis. Generally, TOTWGT should be used. TCHWGT should be used when analyzing teacher data from the AST and ATG files.
JKZ	The variable that captures the assignment of cases to sampling zones. The name of this variable in all PIRLS data files is JKZONE.

JKR	The variable that captures whether the case is to be dropped or have its weight doubled for each set of replicate weights. The name of this variable in all PIRLS data files is JKREP.
NJKZ	The number of replicate weights to be generated when computing the JRR standard errors. The value of NJKZ should be set to 75, the maximum possible value across all participating countries.
CVAR	The list of variables that are to be used to define the subgroups. The list can consist of one or more variables. We recommend that you always include IDCNTRY as the first classification variable.
XVAR	The list of independent variables used as predictors in the regression model. The independent variables can be either continuous or categorical, such as ITSEX for example.
ROOTPV	The variable root used to identify the set of plausible values for the achievement score of interest. It corresponds to the first 7 characters of the plausible values variable name. For example, the root of the overall reading plausible values is ASRREA0, the root of the literary purpose plausible values is "ASRLIT0."
NPV	The number of plausible values that will be used for the analysis. Generally, you will want to use all five plausible values for analysis.
INFILE	The name of the data file that contains the data being analyzed. If the path location is included as part of the file name, the name of the file must be enclosed in quotes. It is important to emphasize that this data file must include only those cases that are of interest in the analysis. If you want to have specific cases excluded from the analysis (e.g., students with missing data), this should be done prior to invoking the macro.

The JACKREGP macro is invoked by an SPSS program using the conventional SPSS notation for invoking macros. This involves listing the macro name followed by the list of parameters in parenthesis, each separated by a forward slash. For example, the macro invoked using the following statement:

```
JACKREGP INFILE = ASGALLR2 /
         CVAR   = IDCNTRY  /
         XVAR   = REGSEX   /
         ROOTPV = ASRREA0  /
         NPV    = 5        /
         NJKZ   = 75       /
         JKZ    = JKZONE   /
         JKR    = JKREP    /
         WGT    = TOTWGT   .
```

will perform a linear regression with the variable REGSEX as a predictor of reading achievement based on its five plausible values (ASRREA01 through ASRREA05), using the weighting variable TOTWGT. It will compute the regression coefficients and their standard errors. The data will be read from the data file ASGALLR2 and the standard errors will be computed based on 75 replicate weights.

The results of the JACKREGP macro are stored in an SPSS working file called REG which is stored in the default folder being used by SPSS. The following variables are contained in this results file:

Classification Variables

All classification variables are kept in the results file. In our example, there is a single classification variable IDCNTRY. There is one record in the results file for each subgroup defined by the categories of the classification variables.

N

This variable contains the number of valid cases for each subgroup defined by the classification variables. In our example, it is the number of students with valid data in each country's sample.

MULT_RSQ

The squared multiple correlation (R^2) coefficient for the regression model applied in each subgroup.

SS_RES, SS_REG, SS_TOTAL

The residual, regression, and total weighted sums of squares for the regression model applied in each subgroup.

Regression Coefficients and Standard Errors (B## and B##.SE)

The regression coefficients for the predictor variables and the intercept with their respective standard errors. The regression coefficients are numbered sequentially, starting with zero (B00) for the intercept, and based on the order of the predictor variables are specified in the parameter XVAR.

The contents of the REG file can be printed using the SPSS REPORT procedure. The sample SPSS program invoking the JACKREGP macro and a printout of the results are presented in Exhibit 4.6. This program is available on the database DVD in the file called SAMPLEJACKREGP.SPS. It performs a linear regression in each country, with the variable REGSEX as a predictor of reading achievement. The exhibit displays the results for the first five countries.

The regression performed by our sample program uses the variable REGSEX that was defined in our previous example. By using REGSEX, the intercept (B00) will be the estimated mean reading achievement of girls, whereas the regression coefficient B01 will be the estimated difference in the mean reading achievement of boys. This will allow us to perform a t-test to determine if reading achievement is significantly different between girls and boys.

Exhibit 4.6 Sample SPSS Program Invoking the SPSS Macro JACKREGP and Output File

```
GET FILE = "<datpath>ASGALLR2.SAV" .

SELECT IF NOT(MISSING(ITSEX)) .

RECODE ITSEX (1=0) (2=1) INTO REGSEX .

VALUE LABELS
 IDCNTRY < list country formats > .

SAVE OUTFILE = ASGALLR2 .

INCLUDE "<macpath>JACKREGP.SPS" .

JACKREGP INFILE = ASGALLR2 /
         CVAR   = IDCNTRY  /
         XVAR   = REGSEX   /
         ROOTPV = ASRREA0  /
         NPV    = 5        /
         NJKZ   = 75       /
         JKZ    = JKZONE   /
         JKR    = JKREP    /
         WGT    = TOTWGT   .

PRINT FORMATS IDCNTRY N (F6.0) MULT_RSQ (F5.3) SS_TOTAL SS_REG (F10.0)
              B00 B00.SE B01 B01.SE (F6.2) .

REPORT FORMAT = LIST AUTOMATIC
  / VAR = IDCNTRY (LABEL) N MULT_RSQ SS_TOTAL SS_REG B00 B00.SE B01 B01.SE .


*COUNTRY ID*     N    MULT_RSQ  SS_TOTAL    SS_REG    B00   B00.SE  B01   B01.SE
_____________ _____ ________ __________ __________ ______ ______ ______ ______

AUSTRIA          5067    .006   337056418    2026407 543.26   2.32  -9.82   2.35
BULGARIA         3863    .016   433266259    6824206 557.52   4.41 -20.74   3.83
CHINESE TAIPEI   4589    .010  1252797359   12579053 542.08   2.22 -12.83   1.86
DENMARK          4001    .010   307300169    2951463 552.94   2.84 -13.63   3.25
FRANCE           4403    .007  3279772121   22867710 527.31   2.42 -11.09   2.54
```

From the first line of the results shown in Exhibit 4.6, we see that in Austria the estimated average reading achievement of fourth grade girls (B00) is 543.26, with a standard error of 2.32. Note that these are the same results obtained from the JACKPV sample program. The Austrian fourth grade boys have an estimated average reading achievement 9.82 points (B01) lower than girls. With an estimated standard error of 2.35, this difference is indeed statistically significant at a 95% confidence level.

4.6 Performing Analyses with Student-level Variables

Many analyses of the PIRLS data can be undertaken using student-level data. Examples in the previous sections illustrate the functioning of the SPSS macros. This section presents examples of actual analyses used to produce the exhibits in the *PIRLS 2006 International Report*, using SPSS programs provided on the DVD.

Student-Level Analysis Without Achievement Scores

In our first example, we wish to replicate the analysis of students' reports on the number of hours spent reading stories or articles in books or magazines outside of school. The results, presented in Exhibit 4.5 of the *PIRLS 2006 International Report*, are reproduced here in Exhibit 4.7. This example will focus on the results presented in the fourth data column—the average number of hours spent reading stories or articles in books or magazines outside of school, overall. Since we want to report the average number of hours, which does not require any plausible values, we will use the macro JACKGEN.

We need to undertake a number of steps to replicate the results in this exhibit. After reviewing the *Student Questionnaire* and codebook for the student background data files, we observe that the variable ASBGTSP4 contains categorical information on the number of hours spent reading stories or articles in books or magazines outside of school (see Supplement 1 for a copy of the questionnaire), and this variable is found in the student background data files. Our next step is to review the documentation of national adaptations to the questionnaires to ensure that there were no deviations listed for this variable (see Supplement 2). If no national adaptations were made, we can continue with our analysis without any modifications.

We then proceed to read from the student background data files our variable of interest (ASBGTSP4), the student sampling weight (TOTWGT), the variables that contain the jackknife replication information (JKZONE and JKREP), and the variable containing the country identification code (IDCNTRY). In this analysis, we will use the data for all available countries. We used the JOIN program, described earlier in this chapter, to join the ASG files for all countries into a single file called ASGALLR2.

The SPSS program used to perform this first example is presented in Exhibit 4.8 and is included on the DVD under the name EXAMPLE1.SPS. The results obtained from this program are displayed in Exhibit 4.9. Note that one of the steps in this program is to select only those students who have non-missing data in our

variable of interest ASBGTSP4. Furthermore, the variable ASBGTSP4 is recoded to have the mid-points of the categorical intervals as its values.

Exhibit 4.7 Example Student-Level Analysis Without Achievement Scores Taken from the PIRLS 2006 International Report (Exhibit 4.5)

Exhibit 4.5 **Students Read Stories or Articles Outside of School** — **PIRLS 2006** 4th Grade

Countries	Average Number of Hours on a Typical Day Spent Reading					
	Stories or Articles on the Internet			Stories or Articles in Books or Magazines		
	Overall	Girls	Boys	Overall	Girls	Boys
Austria	0.8 (0.02)	0.8 (0.03)	0.9 (0.03) ▲	1.6 (0.03)	1.8 (0.04) ▲	1.3 (0.04)
Belgium (Flemish)	0.6 (0.03)	0.5 (0.03)	0.6 (0.04) ▲	0.6 (0.02)	0.7 (0.03) ▲	0.5 (0.02)
Belgium (French)	1.2 (0.04)	1.2 (0.05)	1.3 (0.05)	1.2 (0.03)	1.3 (0.04) ▲	1.1 (0.04)
Bulgaria	1.0 (0.05)	0.9 (0.05)	1.1 (0.06) ▲	1.5 (0.05)	1.6 (0.07) ▲	1.3 (0.05)
Canada, Alberta	0.8 (0.03)	0.8 (0.04)	0.9 (0.04)	1.3 (0.03)	1.4 (0.04) ▲	1.2 (0.04)
Canada, British Columbia	0.8 (0.03)	0.7 (0.04)	0.8 (0.04) ▲	1.3 (0.04)	1.4 (0.05) ▲	1.2 (0.05)
Canada, Nova Scotia	0.9 (0.03)	0.9 (0.04)	1.0 (0.04) ▲	1.3 (0.03)	1.5 (0.04) ▲	1.2 (0.04)
Canada, Ontario	0.9 (0.04)	0.8 (0.05)	0.9 (0.04)	1.4 (0.05)	1.5 (0.06) ▲	1.3 (0.06)
Canada, Quebec	1.0 (0.04)	1.0 (0.04)	1.1 (0.06)	1.3 (0.04)	1.5 (0.06) ▲	1.2 (0.04)
Chinese Taipei	1.0 (0.03)	1.1 (0.04) ▲	0.9 (0.03)	1.2 (0.03)	1.4 (0.04) ▲	1.1 (0.04)
Denmark	0.6 (0.03)	0.5 (0.03)	0.7 (0.04) ▲	1.0 (0.03)	1.1 (0.04) ▲	0.9 (0.04)
England	0.9 (0.03)	0.8 (0.04)	0.9 (0.04)	1.2 (0.03)	1.4 (0.05) ▲	1.1 (0.04)
France	0.9 (0.03)	0.9 (0.04)	0.9 (0.05)	1.2 (0.03)	1.2 (0.04) ▲	1.1 (0.04)
Georgia	0.9 (0.05)	0.8 (0.06)	0.9 (0.06)	1.5 (0.05)	1.6 (0.07) ▲	1.4 (0.06)
Germany	0.6 (0.02)	0.6 (0.03)	0.6 (0.03) ▲	1.5 (0.03)	1.7 (0.04) ▲	1.4 (0.05)
Hong Kong SAR	1.1 (0.03)	1.1 (0.04)	1.0 (0.04)	1.0 (0.03)	1.1 (0.04) ▲	1.0 (0.04)
Hungary	0.7 (0.03)	0.7 (0.04)	0.7 (0.03)	1.3 (0.04)	1.4 (0.06) ▲	1.1 (0.04)
Iceland	0.6 (0.02)	0.6 (0.02)	0.7 (0.03) ▲	0.8 (0.02)	0.9 (0.03) ▲	0.7 (0.03)
Indonesia	1.3 (0.05)	1.2 (0.05)	1.3 (0.06)	1.6 (0.04)	1.7 (0.06)	1.6 (0.05)
Iran, Islamic Rep. of	0.3 (0.03)	0.3 (0.05)	0.4 (0.05)	1.5 (0.05)	1.5 (0.06)	1.4 (0.08)
Israel	1.5 (0.04)	1.4 (0.05)	1.5 (0.05) ▲	1.4 (0.04)	1.6 (0.05) ▲	1.3 (0.04)
Italy	0.7 (0.03)	0.6 (0.04)	0.8 (0.04) ▲	1.3 (0.04)	1.4 (0.05) ▲	1.1 (0.05)
Kuwait	2.1 (0.06)	2.0 (0.07)	2.1 (0.08)	2.1 (0.05)	2.2 (0.07)	2.1 (0.07)
Latvia	1.0 (0.04)	0.9 (0.04)	1.1 (0.05) ▲	1.2 (0.03)	1.4 (0.05) ▲	1.0 (0.04)
Lithuania	0.9 (0.03)	0.9 (0.04)	1.0 (0.03)	1.4 (0.03)	1.7 (0.04) ▲	1.2 (0.04)
Luxembourg	0.5 (0.01)	0.5 (0.02)	0.6 (0.02) ▲	0.9 (0.02)	1.0 (0.02) ▲	0.8 (0.03)
Macedonia, Rep. of	1.8 (0.08)	1.7 (0.09)	1.8 (0.08) ▲	2.6 (0.07)	2.7 (0.08) ▲	2.4 (0.07)
Moldova, Rep. of	1.0 (0.06)	0.9 (0.08)	1.0 (0.06)	1.8 (0.05)	1.9 (0.07) ▲	1.7 (0.06)
Morocco	1.3 (0.08)	1.3 (0.09)	1.4 (0.08)	1.3 (0.07)	1.4 (0.09)	1.3 (0.08)
Netherlands	0.5 (0.02)	0.5 (0.02)	0.5 (0.03)	0.8 (0.02)	0.9 (0.04) ▲	0.6 (0.03)
New Zealand	0.9 (0.03)	0.9 (0.04)	1.0 (0.04) ▲	1.4 (0.04)	1.6 (0.05) ▲	1.3 (0.04)
Norway	0.6 (0.03)	0.5 (0.04)	0.6 (0.03)	0.9 (0.04)	0.9 (0.05) ▲	0.8 (0.05)
Poland	0.9 (0.03)	0.8 (0.03)	1.0 (0.04) ▲	1.5 (0.03)	1.7 (0.04) ▲	1.3 (0.05)
Qatar	2.3 (0.03)	2.2 (0.04)	2.4 (0.04) ▲	2.2 (0.03)	2.3 (0.04) ▲	2.1 (0.04)
Romania	0.9 (0.06)	0.8 (0.06)	1.0 (0.07) ▲	1.6 (0.05)	1.8 (0.07) ▲	1.5 (0.06)
Russian Federation	0.5 (0.03)	0.4 (0.02)	0.6 (0.04) ▲	1.5 (0.04)	1.6 (0.05) ▲	1.3 (0.04)
Scotland	0.9 (0.04)	0.9 (0.05)	0.9 (0.05)	1.2 (0.03)	1.4 (0.05) ▲	1.1 (0.05)
Singapore	1.1 (0.03)	1.1 (0.03)	1.0 (0.04)	1.4 (0.02)	1.6 (0.04) ▲	1.2 (0.03)
Slovak Republic	0.7 (0.03)	0.7 (0.04)	0.8 (0.04) ▲	1.5 (0.04)	1.7 (0.05) ▲	1.3 (0.05)
Slovenia	0.7 (0.03)	0.6 (0.03)	0.8 (0.04) ▲	1.0 (0.02)	1.1 (0.03) ▲	0.9 (0.03)
South Africa	2.1 (0.07)	2.1 (0.07)	2.1 (0.07)	2.7 (0.06)	2.7 (0.06) ▲	2.6 (0.07)
Spain	0.9 (0.03)	0.8 (0.04)	1.0 (0.05) ▲	1.2 (0.03)	1.2 (0.05)	1.2 (0.05)
Sweden	0.5 (0.02)	0.4 (0.02)	0.6 (0.04) ▲	0.7 (0.02)	0.7 (0.03) ▲	0.6 (0.03)
Trinidad and Tobago	1.5 (0.07)	1.5 (0.09)	1.5 (0.07)	1.7 (0.06)	1.9 (0.08) ▲	1.6 (0.07)
United States	1.0 (0.05)	1.1 (0.06)	1.0 (0.05)	1.4 (0.04)	1.6 (0.04) ▲	1.2 (0.06)
International Avg.	1.0 (0.01)	0.9 (0.01)	1.0 (0.01) ▲	1.4 (0.01)	1.5 (0.01) ▲	1.3 (0.01)

▲ Average significantly higher than other gender

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

In general, to perform student-level analyses of this type using the student background data files, you should do the following:

1) Identify the variables of interest in the student background data files and note any specific national adaptations to the variables
2) Retrieve the relevant variables from the student background data files, including analysis and classification variables, identification variables, sampling and weighting variables, and any other variables used in the selection of cases
3) Perform any necessary variable transformations or recodes
4) Use the macros JACKGEN and JACKREG with the appropriate parameters
5) Specify the location of the data files (<datpath>) and the macros (<macpath>)
6) Print out the results file

Exhibit 4.8 Example SPSS Program to Perform Student-Level Analysis without Achievement Scores (EXAMPLE1.SPS)

```
GET FILE = "<datpath>ASGALLR2.SAV" .

SELECT IF NOT(MISSING(ASBGTSP4)) .

RECODE ASBGTSP4 (1=5.5) (2=4) (3=2) (4=0.5) (5=0) (ELSE=SYSMIS) INTO ASBGTSP4.

VALUE LABELS
 IDCNTRY < list country formats > .

SAVE OUTFILE = ASGALLR2 .

INCLUDE "<macpath>JACKGEN.SPS" .

JACKGEN INFILE = ASGALLR2 /
        CVAR   = IDCNTRY  /
        DVAR   = ASBGTSP4 /
        NJKZ   = 75       /
        JKZ    = JKZONE   /
        JKR    = JKREP    /
        WGT    = TOTWGT   .

PRINT FORMATS IDCNTRY N (F6.0)
              TOTWGT (F10.0) MNX MNX_SE PCT PCT_SE (F6.2) .

REPORT FORMAT = LIST AUTOMATIC
  / VAR = IDCNTRY (LABEL) N TOTWGT MNX MNX_SE PCT PCT_SE .
```

Exhibit 4.9 Output for Example Student-Level Analysis without Achievement Scores (EXAMPLE 1)

COUNTRY ID	N	TOTWGT	MNX	MNX_SE	PCT	PCT_SE
AUSTRIA	4967	81526	1.57	.03	.50	.02
BELGIUM FLEMISH	4446	65646	.63	.02	.40	.01
BELGIUM FRENCH	4479	46885	1.23	.03	.29	.01
BULGARIA	3739	61273	1.46	.05	.38	.01
CANADA ONTARIO	3888	136297	1.41	.05	.84	.03
CANADA QUEBEC	3626	76016	1.35	.04	.47	.01
CANADA ALBERTA	4113	35481	1.30	.03	.22	.01
CANADA BRITISH COLUMBIA	4045	41663	1.28	.04	.26	.00
CANADA NOVA SCOTIA	4297	9380	1.32	.03	.06	.00
CHINESE TAIPEI	4462	295892	1.24	.03	1.82	.03
DENMARK	3939	62239	1.03	.03	.38	.01
ENGLAND	3952	540722	1.24	.04	3.33	.09
FRANCE	4269	718044	1.15	.03	4.42	.11
GEORGIA	4115	41524	1.53	.05	.26	.01
GERMANY	7340	733674	1.53	.03	4.52	.09
HONG KONG (SAR)	4572	68340	1.02	.03	.42	.01
HUNGARY	4032	103739	1.27	.04	.64	.02
ICELAND	3588	3978	.82	.02	.02	.00
INDONESIA	4505	4008052	1.65	.04	24.69	.60
IRAN	5323	1138032	1.47	.05	7.01	.19
ISRAEL	3689	80600	1.45	.04	.50	.01
ITALY	3520	503243	1.25	.04	3.10	.06
KUWAIT	3392	23650	2.12	.05	.15	.00
LATVIA	4116	19567	1.17	.03	.12	.00
LITHUANIA	4666	32496	1.43	.03	.20	.00
LUXEMBOURG	5072	5140	.88	.02	.03	.00
MACEDONIA	3647	20817	2.58	.07	.13	.00
MOLDOVA	3946	42786	1.77	.05	.26	.01
MOROCCO	3120	543315	1.33	.07	3.35	.10
NETHERLANDS	4106	174450	.76	.02	1.07	.03
NEW ZEALAND	5927	53772	1.44	.04	.33	.01
NORWAY	3732	59819	.88	.04	.37	.01
POLAND	4770	389159	1.48	.03	2.40	.06
QATAR	6098	6516	2.25	.03	.04	.00
ROMANIA	4144	192349	1.63	.05	1.18	.04
RUSSIAN FEDERATION	4673	1212611	1.46	.04	7.47	.17
SCOTLAND	3703	56010	1.24	.03	.35	.01
SINGAPORE	6325	48760	1.40	.02	.30	.00
SLOVAKIA	5302	51588	1.54	.04	.32	.01
SLOVENIA	5276	17423	.98	.02	.11	.00
SOUTH AFRICA	13092	744365	2.67	.06	4.59	.07
SPAIN	4026	384242	1.19	.03	2.37	.05
SWEDEN	4306	100097	.68	.02	.62	.02
TRINIDAD AND TOBAGO	3748	16249	1.74	.06	.10	.00
UNITED STATES	4922	3187176	1.40	.04	19.63	.45

In this example, each country's mean value for the recoded ASBGTSP4 variable is reported for all sampled students. The countries are identified in the first column and the second column reports the number of valid cases. The third column reports the sum of weights of the sampled students, followed by the mean for ASBGTSP4 and its standard error. The last two columns report the weighted percentage of students in the population and its standard error. For this example, the weighted percentages are of little use as they are the proportion each country represents among all participating countries.

From the first line shown in Exhibit 4.9, we see that in Austria valid data were available for 4,967 students and these sampled students represent a population of

81,526 students. On average, Austrian students spend 1.57 hours reading stories or articles in books or magazines outside of school, with a standard error of 0.03.

Student-Level Analysis with Achievement Scores

Our second example replicates another set of results presented in the *PIRLS 2006 International Report*. We are interested in investigating whether students' reading achievement increases as students' reports of the frequency with which they read outside of school increases. These results, presented in Exhibit 4.6 of the *PIRLS 2006 International Report*, are repeated here in Exhibit 4.10. Since the results in this Exhibit are based on plausible values, we need to use the macro JACKPV.

After reviewing the questionnaire and codebook, we observe that the variable ASBGTOC5, found in the ASG data file, contains categorical information on how often students reported reading for fun outside of school. Our next step is to review the documentation of national adaptations to the questionnaires to ensure that there were no deviations listed for this variable (see Supplement 2). If no national adaptations were made, we can continue with our analysis without any modifications.

We then proceed to read from the ASG data files our variable of interest (ASBGTOC5), the five plausible values of reading achievement (ASRREA01 through ASRREA05), the student sampling weight (TOTWGT), the variables that contain the jackknife replication information (JKZONE and JKREP), and the variable containing the country identification code (IDCNTRY). Again, we will use the data of all available countries contained in the file ASGALLR2.

The SPSS program that implements this second example is presented in Exhibit 4.11 and is included on the DVD under the name EXAMPLE2.SPS. The results obtained from this program are displayed in Exhibit 4.12. For the sake of conciseness, only the results of first five and last five countries, as sorted alphabetically, are shown. Note that one of the steps in this program is to select only those students who have non-missing data in our variable of interest ASBGTOC5. A second step consists of combining category 4 of the variable ASBGTOC5 with category 3 to create a category we will label "Twice a Month or Less".

Exhibit 4.10 Example Student-Level Analysis with Achievement Scores Taken From the PIRLS 2006 International Report (Exhibit 4.6)

Exhibit 4.6 **Students Reading for Fun Outside of School with Trends** — **PIRLS 2006** 4th Grade

Countries	Every Day or Almost Every Day			Once or Twice a Week			Twice a Month or Less		
	2006 Percent of Students	Average Achievement	Difference in Percent from 2001	2006 Percent of Students	Average Achievement	Difference in Percent from 2001	2006 Percent of Students	Average Achievement	Difference in Percent from 2001
Russian Federation	58 (1.1)	570 (3.8)	0 (1.8)	28 (0.8)	559 (3.9)	–1 (1.3)	14 (0.8)	556 (3.9)	2 (1.3)
Canada, Alberta	53 (0.9)	575 (2.5)	◊ ◊	23 (0.8)	555 (3.5)	◊ ◊	23 (0.9)	537 (2.7)	◊ ◊
Canada, British Columbia	53 (1.0)	573 (2.9)	◊ ◊	26 (0.8)	554 (2.5)	◊ ◊	21 (0.9)	531 (4.1)	◊ ◊
Germany	53 (0.9)	563 (2.7)	5 (1.2) ▲	24 (0.6)	545 (3.0)	0 (0.9)	24 (0.8)	525 (2.5)	–5 (1.2) ▼
Lithuania	52 (1.2)	545 (2.1)	–1 (1.9)	30 (1.0)	533 (2.2)	–1 (1.6)	17 (0.8)	520 (2.8)	2 (1.2)
Moldova, Rep. of	52 (1.4)	507 (3.2)	2 (2.6)	34 (1.1)	498 (4.2)	0 (1.9)	14 (1.0)	484 (5.1)	–2 (1.6)
France	51 (1.0)	540 (2.5)	2 (1.6)	24 (0.8)	517 (2.3)	–2 (1.2)	25 (0.9)	491 (2.7)	0 (1.3)
Canada, Ontario	49 (1.4)	567 (3.2)	14 (2.0) ▲	25 (1.1)	552 (3.6)	2 (1.4)	26 (1.1)	534 (4.1)	–16 (1.9) ▼
Belgium (French)	49 (1.1)	517 (3.0)	◊ ◊	26 (0.7)	495 (2.9)	◊ ◊	25 (0.9)	473 (3.1)	◊ ◊
Iceland	49 (0.9)	527 (1.9)	–3 (1.2) ▼	23 (0.7)	511 (2.6)	2 (1.0)	28 (0.7)	485 (2.3)	1 (1.0)
Denmark	49 (1.1)	559 (2.9)	◊ ◊	30 (0.8)	540 (2.7)	◊ ◊	21 (0.9)	528 (3.2)	◊ ◊
Canada, Nova Scotia	48 (0.9)	560 (2.6)	◊ ◊	25 (0.8)	541 (2.7)	◊ ◊	27 (0.8)	515 (3.2)	◊ ◊
Bulgaria	47 (1.6)	561 (4.4)	–4 (2.3)	27 (1.0)	555 (5.0)	–2 (1.5)	26 (1.6)	520 (6.5)	5 (2.2) ▲
Canada, Quebec	47 (1.3)	549 (3.0)	1 (1.9)	26 (1.0)	530 (3.8)	1 (1.4)	27 (1.2)	509 (3.2)	–3 (1.7)
South Africa	45 (0.9)	303 (6.4)	◊ ◊	26 (0.5)	314 (6.4)	◊ ◊	28 (0.8)	307 (6.4)	◊ ◊
Austria	45 (1.1)	555 (3.0)	◊ ◊	25 (0.8)	535 (2.4)	◊ ◊	29 (1.0)	516 (2.9)	◊ ◊
Spain	45 (1.1)	525 (2.9)	◊ ◊	27 (0.7)	515 (2.5)	◊ ◊	28 (1.0)	494 (3.7)	◊ ◊
Hungary	44 (1.2)	565 (3.7)	4 (1.7) ▲	30 (0.9)	547 (3.4)	–2 (1.4)	26 (1.1)	532 (4.2)	–2 (1.5)
Indonesia	44 (1.4)	405 (4.7)	◊ ◊	31 (1.1)	414 (4.6)	◊ ◊	25 (1.2)	403 (4.8)	◊ ◊
Macedonia, Rep. of	43 (1.2)	453 (5.7)	–3 (1.9)	31 (1.0)	451 (4.6)	3 (1.4) ▲	25 (1.1)	435 (5.1)	0 (1.8)
Poland	43 (1.3)	538 (2.5)	◊ ◊	29 (1.0)	518 (3.2)	◊ ◊	27 (1.0)	495 (3.0)	◊ ◊
New Zealand	42 (1.1)	562 (2.4)	–1 (1.8)	24 (0.7)	531 (2.5)	0 (1.1)	34 (1.0)	500 (3.0)	1 (1.6)
Belgium (Flemish)	40 (1.1)	563 (2.1)	◊ ◊	29 (0.8)	545 (2.9)	◊ ◊	31 (1.2)	529 (2.3)	◊ ◊
Slovak Republic	39 (1.0)	545 (2.9)	0 (1.5)	33 (0.9)	535 (3.2)	0 (1.3)	27 (1.1)	507 (5.4)	0 (1.6)
Italy	38 (1.3)	573 (3.3)	7 (1.7) ▲	25 (0.7)	554 (3.2)	1 (1.0)	37 (1.3)	529 (3.8)	–7 (1.7) ▼
Luxembourg	38 (0.6)	581 (1.8)	◊ ◊	27 (0.7)	551 (2.0)	◊ ◊	35 (0.5)	537 (1.5)	◊ ◊
Israel	38 (1.2)	538 (4.2)	–6 (1.6) ▼	28 (0.9)	518 (4.4)	1 (1.2)	35 (1.1)	497 (4.1)	5 (1.5) ▲
Slovenia	37 (0.9)	543 (2.5)	–8 (1.6) ▼	33 (0.7)	519 (3.0)	4 (1.1) ▲	30 (0.9)	500 (2.6)	4 (1.5) ▲
Netherlands	36 (1.1)	566 (2.1)	0 (1.6)	22 (0.7)	550 (1.8)	2 (1.1) ▲	42 (1.1)	530 (1.8)	–2 (1.6)
Sweden	36 (1.0)	569 (2.8)	–8 (1.3) ▼	31 (0.9)	549 (3.2)	–1 (1.2)	33 (1.0)	530 (2.6)	9 (1.3) ▲
United States	35 (1.3)	561 (4.3)	1 (1.8)	22 (0.7)	550 (3.3)	0 (1.3)	43 (1.4)	521 (3.3)	–1 (1.8)
Latvia	35 (1.2)	556 (3.0)	–8 (1.7) ▼	31 (0.8)	543 (2.8)	–2 (1.2)	34 (1.2)	524 (2.6)	10 (1.7) ▲
Hong Kong SAR	35 (1.0)	575 (2.6)	14 (1.3) ▲	33 (0.9)	567 (2.7)	–5 (1.2) ▼	32 (1.0)	549 (2.8)	–8 (1.4) ▼
Iran, Islamic Rep. of	33 (1.2)	428 (4.2)	–1 (1.7)	41 (1.2)	429 (3.9)	2 (1.9)	26 (1.0)	406 (5.2)	0 (1.5)
Norway	33 (1.1)	514 (3.4)	–5 (1.5) ▼	30 (1.0)	505 (3.2)	2 (1.3)	37 (1.2)	481 (3.1)	4 (1.7) ▲
Qatar	33 (0.6)	357 (2.2)	◊ ◊	28 (0.5)	367 (2.5)	◊ ◊	39 (0.6)	352 (2.2)	◊ ◊
Scotland	33 (1.1)	555 (4.4)	2 (1.6)	24 (1.0)	533 (3.1)	0 (1.3)	44 (1.5)	505 (2.7)	–2 (2.2)
England	33 (1.2)	575 (4.0)	0 (1.8)	25 (0.8)	537 (3.5)	–1 (1.2)	42 (1.3)	517 (2.9)	1 (2.0)
Kuwait	32 (1.1)	338 (5.5)	◊ ◊	32 (1.0)	342 (5.5)	◊ ◊	36 (1.2)	332 (5.4)	◊ ◊
Trinidad and Tobago	32 (1.2)	450 (6.7)	◊ ◊	25 (1.0)	442 (5.9)	◊ ◊	43 (1.4)	427 (5.6)	◊ ◊
Georgia	29 (1.4)	479 (4.5)	◊ ◊	29 (1.2)	484 (4.0)	◊ ◊	41 (1.4)	461 (4.3)	◊ ◊
Morocco	29 (1.3)	317 (8.2)	–3 (2.3)	34 (1.4)	326 (6.9)	–3 (2.0)	37 (1.9)	331 (7.1)	6 (3.1)
Singapore	27 (0.9)	587 (3.9)	–3 (1.5) ▼	26 (0.6)	564 (3.1)	3 (0.9) ▲	47 (1.0)	540 (2.7)	0 (1.5)
Romania	25 (1.3)	510 (5.5)	–3 (2.2)	26 (1.1)	502 (5.5)	–4 (1.6) ▼	50 (1.6)	478 (5.5)	7 (2.3) ▲
Chinese Taipei	24 (0.7)	553 (2.6)	◊ ◊	31 (0.8)	539 (2.6)	◊ ◊	45 (1.0)	525 (2.2)	◊ ◊
International Avg.	40 (0.2)	516 (0.6)		28 (0.1)	503 (0.6)		32 (0.2)	484 (0.6)	

▲ Percent in 2006 significantly higher
▼ Percent in 2006 significantly lower

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

In general, to perform student-level analyses of this type using the student background data files, you should do the following:

1) Identify the variables of interest in the student background data files and note any specific national adaptations to the variables
2) Retrieve the relevant variables from the student background data files, including plausible values for achievement, classification variables, identification variables, sampling and weighting variables, and any other variables used in the selection of cases
3) Perform any necessary variable transformations or recodes
4) Use the macros JACKPV and JACKREGP with the appropriate parameters
5) Specify the location of the data files (<datpath>) and the macros (<macpath>)
6) Print out the results file

Exhibit 4.11 Example SPSS Program to Perform Student-Level Analysis with Achievement Scores (EXAMPLE2.SPS)

```
GET FILE = "<datpath>ASGALLR2.SAV" .

SELECT IF NOT(MISSING(ASBGTOC5)) .

RECODE ASBGTOC5 (1=1) (2=2) (3,4=3) (ELSE=SYSMIS) INTO ASBGTOC5.

VALUE LABELS
 IDCNTRY < list country formats > .

VALUE LABELS
 ASBGTOC5
     1 'EVERY DAY OR ALMOST EVERY DAY'
     2 'ONCE OR TWICE A WEEK'
     3 'TWICE A MONTH OR LESS' .

SAVE OUTFILE = ASGALLR2 .

INCLUDE "<macpath>JACKPV.SPS" .

JACKPV INFILE = ASGALLR2         /
       CVAR   = IDCNTRY ASBGTOC5 /
       ROOTPV = ASRREA0          /
       NPV    = 5                /
       JKZ    = JKZONE           /
       JKR    = JKREP            /
       NJKZ   = 75               /
       WGT    = TOTWGT           .

PRINT FORMATS IDCNTRY ASBGTOC5 N (F6.0)
              TOTWGT (F10.0) MNPV MNPV_SE PCT PCT_SE (F6.2) .

REPORT FORMAT = LIST AUTOMATIC
  / VAR = ASBGTOC5 (LABEL) N TOTWGT MNPV MNPV_SE PCT PCT_SE
  / BREAK = IDCNTRY (LABEL) .
```

Exhibit 4.12 Output for Example Student-Level Analysis with Achievement Scores (EXAMPLE 2)

COUNTRY ID	GEN/THINGS OUTSIDE SCH/ READING FOR FUN	N	TOTWGT	MNPV	MNPV_SE	PCT	PCT_SE
AUSTRIA	EVERY DAY OR ALMOST EVERY DAY	2281	36917	555.44	3.00	45.16	1.10
AUSTRIA	ONCE OR TWICE A WEEK	1255	20821	535.13	2.43	25.47	.80
AUSTRIA	TWICE A MONTH OR LESS	1445	24016	515.74	2.88	29.38	.98
BELGIUM FLEMISH	EVERY DAY OR ALMOST EVERY DAY	1776	25976	562.65	2.11	39.61	1.06
BELGIUM FLEMISH	ONCE OR TWICE A WEEK	1281	19015	545.24	2.89	28.99	.80
BELGIUM FLEMISH	TWICE A MONTH OR LESS	1386	20591	529.46	2.33	31.40	1.16
BELGIUM FRENCH	EVERY DAY OR ALMOST EVERY DAY	2207	22935	516.84	3.00	49.09	1.10
BELGIUM FRENCH	ONCE OR TWICE A WEEK	1146	12063	494.61	2.93	25.82	.66
BELGIUM FRENCH	TWICE A MONTH OR LESS	1106	11726	472.75	3.14	25.10	.94
BULGARIA	EVERY DAY OR ALMOST EVERY DAY	1796	29070	560.61	4.37	47.24	1.59
BULGARIA	ONCE OR TWICE A WEEK	1033	16313	555.34	5.00	26.51	1.04
BULGARIA	TWICE A MONTH OR LESS	928	16154	520.18	6.50	26.25	1.56
CANADA ALBERTA	EVERY DAY OR ALMOST EVERY DAY	2200	18971	574.88	2.46	53.44	.95
CANADA ALBERTA	ONCE OR TWICE A WEEK	971	8276	554.91	3.51	23.31	.79
CANADA ALBERTA	TWICE A MONTH OR LESS	947	8251	536.85	2.75	23.24	.86
	. . . . .						
SOUTH AFRICA	EVERY DAY OR ALMOST EVERY DAY	6161	345653	303.08	6.41	45.50	.91
SOUTH AFRICA	ONCE OR TWICE A WEEK	3443	199247	314.45	6.35	26.23	.53
SOUTH AFRICA	TWICE A MONTH OR LESS	3783	214845	306.94	6.44	28.28	.75
SPAIN	EVERY DAY OR ALMOST EVERY DAY	1801	170563	525.03	2.88	44.77	1.15
SPAIN	ONCE OR TWICE A WEEK	1078	102245	514.71	2.45	26.84	.70
SPAIN	TWICE A MONTH OR LESS	1117	108198	493.87	3.69	28.40	1.01
SWEDEN	EVERY DAY OR ALMOST EVERY DAY	1572	36017	568.61	2.81	35.97	1.02
SWEDEN	ONCE OR TWICE A WEEK	1333	30843	549.35	3.22	30.80	.86
SWEDEN	TWICE A MONTH OR LESS	1401	33281	530.13	2.56	33.23	.96
TRINIDAD AND TOBAGO	EVERY DAY OR ALMOST EVERY DAY	1215	5218	449.84	6.74	31.70	1.17
TRINIDAD AND TOBAGO	ONCE OR TWICE A WEEK	956	4170	441.71	5.89	25.33	.95
TRINIDAD AND TOBAGO	TWICE A MONTH OR LESS	1619	7072	426.74	5.58	42.97	1.43
UNITED STATES	EVERY DAY OR ALMOST EVERY DAY	1717	1131176	560.82	4.26	35.28	1.25
UNITED STATES	ONCE OR TWICE A WEEK	1095	705135	549.57	3.35	21.99	.74
UNITED STATES	TWICE A MONTH OR LESS	2139	1369921	521.22	3.27	42.73	1.43

In this example, each country's results are presented on three lines, one for each value of the recoded ASBGTOC5 variable. The countries are identified in the first column and the second column presents the label of the category of ASBGTOC5 being reported. The third column reports the number of valid cases, and the fourth the sum of weights of the sampled students. The next two columns report the estimated average reading achievement and its standard error. The last two columns report the weighted percentage of students and its standard error.

From the first three lines in Exhibit 4.12, we see that 45.16 percent of students in Austria reported reading for fun every day or almost every day, 25.47 percent reported reading for fun once or twice a week, and 29.38 percent twice a month or less. We also see that the average reading achievement was 555.44 (standard error of 3.00) for those reporting reading every day or almost every day, 535.13 (standard error of 2.43) for those reading once or twice a week, and 515.74 (standard error of 2.88) for those reading twice a month or less.

4.7 Performing Analyses with Teacher Variables

When considering analyses with teacher-level variables, it is important to realize that the teachers do not constitute representative samples of teachers. Rather, they are the teachers of nationally representative samples of students. Therefore, analyses with teacher data should be made with students as units of analysis, in terms of how many students are taught by teachers with a particular attribute, and not in terms of how many teachers in a country have a particular attribute.

When analyzing teacher data, it is first necessary to link the students to their respective teachers. The student-teacher linkage data files were created to facilitate this linkage. Each student record in a student background data file links to at least one reading teacher in its corresponding teacher background data file. There is usually only one student-teacher link, but occasionally two, and in rare circumstances three.

The student-teacher linkage data files contain one record for each student-teacher combination. For example, if a student is taught by two reading teachers, the AST data file has two records for that student, one for each of his/her reading teachers. Each record in the data file holds the number of reading teachers linked to that student.

Student achievement scores (plausible values), jackknife replication information, and a weighting variable, TCHWGT, appropriate for analyzing teacher-level variables, also are found in the AST data files in order to simplify the merging process for analyses that link teacher variables to student achievement. For such analyses, it is only necessary to merge the teacher background data files with the student-teacher linkage data files. For analyses linking teacher variables to student background variables, it is also necessary to merge the student background data files with the teacher background data files after having been combined with the student-teacher linkage data files.

As our example of a teacher data analysis, we will investigate the age of the PIRLS reading teachers. The results of such an analysis are presented in Exhibit 6.3 of the *PIRLS 2006 International Report* and are reproduced here in Exhibit 4.13. Although the results in this exhibit do not include any achievement, we will use the macro JACKPV to estimate the percentages we want.

Conducting analyses with teacher data requires a few extra steps. As before, we first proceed to identify the variables relevant to the analysis in the appropriate files, and review the documentation for any specific national adaptations to the questions of interest (Supplements 1 and 2). Since we are using a teacher-level variable, we need to use the teacher background data files and the student-teacher linkage data files to find the variables. From the ATG data files, we will need the variable that contains the information on the reading teachers' age (ATBGAGE), the variable that identifies the country (IDCNTRY), and the two identification variables (IDTEACH and IDLINK) that will allow us to link the teacher data to the student data.

We then proceed to read the necessary information from the student-teacher linkage data files. From these files we need the country identification (IDCNTRY) and the two identification variables (IDTEACH and IDLINK) needed to link the student data to the teacher data. We also need the jackknife replication variables (JKZONE and JKREP), the teacher weighting variable (TCHWGT), and the reading achievement plausible values (ASRREA01 through ASRREA05). Although we are only interested in estimating percentages, the reading achievement plausible values are required input for the JACKPV macro. This may be of analytical interest, providing some insight into the relationship between students' reading achievement and the age of their teachers.

Exhibit 4.13 Example Teacher Variable Analysis Taken from the PIRLS 2006 International Report (Exhibit 6.3)

Exhibit 6.3 **Teachers' Gender, Age, and Number of Years Teaching**

PIRLS 2006
4th Grade

Countries	Percentage of Students by Teacher Characteristics						Trends in Number of Years Teaching All Grades		
	Gender		Age						
	Female	Male	29 Years or Under	30-39 Years	40-49 Years	50 Years or Older	2006	Difference from 2001	
Austria	88 (2.1)	12 (2.1)	6 (1.7)	17 (2.5)	39 (3.0)	38 (3.2)	22 (0.7)	◊ ◊	
Belgium (Flemish)	75 (3.0)	25 (3.0)	28 (3.2)	29 (3.3)	29 (2.9)	14 (2.2)	16 (0.6)	◊ ◊	
Belgium (French)	80 (3.0)	20 (3.0)	16 (2.2)	33 (3.2)	37 (3.5)	14 (2.6)	17 (0.6)	◊ ◊	
Bulgaria	94 (1.8)	6 (1.8)	3 (1.4)	26 (2.8)	40 (4.0)	31 (3.9)	21 (0.6)	4 (0.9)	▲
Canada, Alberta	80 (3.2)	20 (3.2)	22 (3.7)	20 (3.3)	26 (3.6)	32 (3.3)	15 (0.8)	◊ ◊	
Canada, British Columbia	r 72 (3.6)	28 (3.6)	r 9 (2.2)	21 (3.0)	32 (4.1)	38 (4.2)	17 (0.8)	◊ ◊	
Canada, Nova Scotia	84 (2.9)	16 (2.9)	12 (2.4)	22 (3.4)	25 (3.0)	40 (3.4)	18 (0.7)	◊ ◊	
Canada, Ontario	75 (4.8)	25 (4.8)	18 (3.9)	42 (5.2)	21 (4.3)	20 (4.3)	12 (0.9)	-4 (1.3)	▼
Canada, Quebec	86 (3.2)	14 (3.2)	12 (2.3)	37 (4.0)	19 (3.5)	32 (3.7)	17 (0.8)	-1 (1.2)	
Chinese Taipei	83 (3.2)	17 (3.2)	24 (3.7)	44 (4.1)	27 (3.5)	5 (1.6)	12 (0.6)	◊ ◊	
Denmark	90 (2.1)	10 (2.1)	16 (2.9)	24 (3.4)	19 (2.9)	41 (4.3)	16 (1.1)	◊ ◊	
England	75 (3.5)	25 (3.5)	30 (3.8)	33 (4.1)	14 (2.9)	23 (3.7)	12 (0.9)	-2 (1.3)	
France	71 (3.3)	29 (3.3)	17 (2.9)	31 (3.0)	34 (3.4)	18 (2.6)	15 (0.7)	-3 (1.1)	▼
Georgia	100 (0.3)	0 (0.0)	9 (1.6)	28 (3.4)	24 (3.4)	39 (3.6)	20 (0.9)	◊ ◊	
Germany	89 (2.4)	11 (2.4)	5 (1.7)	23 (2.9)	22 (3.0)	49 (3.8)	20 (0.9)	-3 (1.1)	▼
Hong Kong SAR	78 (3.7)	22 (3.7)	29 (4.2)	33 (4.1)	19 (3.5)	19 (2.6)	13 (0.7)	0 (1.2)	
Hungary	97 (1.6)	3 (1.6)	6 (2.1)	21 (3.1)	48 (3.7)	24 (3.5)	21 (0.7)	3 (1.0)	▲
Iceland	93 (0.2)	7 (0.2)	13 (0.3)	31 (0.4)	35 (0.4)	22 (0.3)	12 (0.1)	-1 (0.1)	▼
Indonesia	56 (3.9)	44 (3.9)	13 (2.2)	31 (4.1)	39 (3.9)	17 (3.0)	16 (0.8)	◊ ◊	
Iran, Islamic Rep. of	50 (2.3)	50 (2.3)	8 (1.6)	44 (3.6)	41 (3.5)	7 (1.9)	17 (0.5)	3 (0.8)	▲
Israel	92 (1.9)	8 (1.9)	8 (1.8)	41 (4.0)	33 (4.1)	18 (3.4)	16 (0.8)	2 (1.1)	
Italy	98 (1.1)	2 (1.1)	2 (1.3)	14 (2.5)	37 (3.5)	47 (3.6)	22 (0.7)	1 (1.0)	
Kuwait	86 (2.3)	14 (2.3)	36 (3.8)	48 (3.8)	15 (3.1)	1 (0.0)	8 (0.6)	◊ ◊	
Latvia	99 (0.6)	1 (0.0)	7 (2.0)	30 (3.6)	34 (3.6)	29 (3.4)	21 (0.8)	2 (1.3)	
Lithuania	99 (0.5)	1 (0.5)	1 (0.7)	35 (3.4)	40 (3.2)	23 (3.2)	21 (0.6)	1 (1.1)	
Luxembourg	55 (0.2)	45 (0.2)	32 (0.2)	24 (0.1)	16 (0.1)	27 (0.1)	15 (0.0)	◊ ◊	
Macedonia, Rep. of	70 (3.5)	30 (3.5)	5 (1.7)	24 (3.4)	40 (4.1)	31 (4.2)	r 20 (1.0)	0 (1.5)	
Moldova, Rep. of	90 (2.8)	10 (2.8)	8 (2.3)	23 (3.7)	36 (4.3)	33 (4.2)	24 (0.9)	5 (1.2)	▲
Morocco	56 (4.0)	44 (4.0)	18 (3.0)	25 (3.5)	40 (3.4)	17 (2.9)	17 (0.6)	1 (1.0)	
Netherlands	68 (3.4)	32 (3.4)	26 (3.5)	19 (3.0)	19 (3.5)	36 (4.0)	17 (1.0)	0 (1.4)	
New Zealand	77 (2.7)	23 (2.7)	22 (2.1)	28 (2.6)	26 (2.4)	25 (2.4)	12 (0.6)	-1 (1.1)	
Norway	91 (1.9)	9 (1.9)	9 (2.7)	28 (3.3)	29 (3.9)	35 (3.9)	16 (1.0)	-1 (1.3)	
Poland	100 (0.3)	0 (0.3)	2 (1.1)	31 (3.4)	58 (4.0)	8 (2.3)	20 (0.5)	◊ ◊	
Qatar	r 90 (0.1)	10 (0.1)	r 27 (0.2)	54 (0.3)	16 (0.2)	3 (0.1)	11 (0.0)	◊ ◊	
Romania	89 (2.4)	11 (2.4)	18 (3.1)	25 (3.3)	27 (3.5)	30 (3.2)	22 (0.9)	2 (1.2)	
Russian Federation	98 (1.1)	2 (1.1)	6 (1.6)	38 (3.7)	35 (3.5)	21 (2.8)	22 (0.6)	2 (1.0)	▲
Scotland	96 (1.6)	4 (1.6)	26 (3.8)	19 (2.8)	18 (3.5)	37 (4.2)	16 (1.1)	-2 (1.5)	
Singapore	75 (2.2)	25 (2.2)	37 (2.6)	42 (3.0)	16 (2.4)	6 (1.0)	9 (0.4)	-2 (1.0)	▼
Slovak Republic	93 (1.7)	7 (1.7)	13 (2.4)	37 (3.5)	25 (2.7)	25 (3.2)	17 (0.8)	0 (1.2)	
Slovenia	98 (1.1)	2 (1.1)	12 (2.2)	24 (2.9)	44 (3.0)	19 (2.4)	19 (0.7)	0 (1.0)	
South Africa	71 (2.5)	29 (2.5)	4 (1.2)	44 (2.8)	32 (2.9)	20 (2.3)	15 (0.4)	◊ ◊	
Spain	78 (3.6)	22 (3.6)	11 (2.4)	16 (2.6)	25 (3.4)	49 (3.8)	22 (0.9)	◊ ◊	
Sweden	84 (2.8)	16 (2.8)	9 (1.7)	29 (3.3)	24 (2.9)	38 (3.6)	17 (1.0)	1 (1.2)	
Trinidad and Tobago	82 (3.0)	18 (3.0)	11 (2.5)	37 (3.6)	28 (3.3)	24 (3.0)	19 (0.7)	◊ ◊	
United States	85 (2.7)	15 (2.7)	21 (2.8)	27 (2.8)	28 (3.7)	25 (3.4)	12 (0.7)	-3 (1.1)	▼
International Avg.	83 (0.4)	17 (0.4)	15 (0.4)	30 (0.5)	30 (0.5)	25 (0.5)	17 (0.1)	0 (0.0)	

▲ Number of years in 2006 significantly higher ▼ Number of years in 2006 significantly lower

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

The two file types are merged and the merged file is then input to the JACKPV macro. The merging is done using the combination of identification variables IDCNTRY, IDTEACH, and IDLINK. The combination of values for these three variables is unique within the teacher background data files, but is repeated in the student-teacher linkage data files as many times as needed when a specific student is taught by more than one teacher. After the files are merged, the JACKPV macro is invoked and the results can be printed.

For this analysis, we will again use the data for all available countries, making use of an aggregated teacher background data file, ATGALLR2, and an aggregated student-teacher linkage data file, ASTALLR2. These aggregated files were created using the JOIN macro.

The SPSS program that executes this third example is presented in Exhibit 4.14 and is included on the DVD under the name EXAMPLE3.SPS. The results obtained from this program are displayed in Exhibit 4.15, edited to show only the first four and last four countries, alphabetically, for the sake of conciseness. Note that one of the steps in this program is to select only those students who have non-missing data in our variable of interest ATBGAGE. A second step consists of combining categories 1 and 2 and categories 5 and 6 of the variable ATBGAGE in order to match the results presented in Exhibit 4.13, where teachers are categorized into four groups: 29 years or under, 30-39 years old, 40-49 years old, and 50 years or older.

In general, to perform teacher-level analyses using the teacher background data files, you should do the following:

1) Identify the variables of interest in the appropriate teacher background data files and note any specific national adaptations to the variables

2) Retrieve the relevant variables from the teacher background data files, including analysis and classification variables, identification variables (IDCNTRY, IDTEACH, and IDLINK), and any other variables used in the selection of cases

3) Retrieve the relevant variables from the student-teacher linkage data files, including plausible values for achievement, classification variables, identification variables (IDCNTRY, IDSTUD, IDTEACH, and IDLINK), sampling (JKZONE and JKREP) and weighting (TCHWGT) variables, and any other variables used in the selection of cases

4) Merge the teacher background data files with the student-teacher linkage data files using the variables IDCNTRY, IDTEACH and IDLINK

5) Perform any necessary variable transformations or recodes
6) Use the macros JACKGEN and JACKREG, or JACKPV and JACKREGP if plausible values are involved, with the appropriate arguments and parameters
7) Specify the location of the data files (<datpath>) and the macros (<macpath>)
8) Print out the results file

Exhibit 4.14 Example SPSS Program to Perform Teacher Variable Analysis (EXAMPLE3.SPS)

```
GET FILE = "<datpath>ATGALLR2.SAV" .

SORT CASES BY IDCNTRY IDTEACH IDLINK .

SAVE OUTFILE = ATGALLR2 .

GET FILE = "<datpath>ASTALLR2.SAV" .

SORT CASES BY IDCNTRY IDTEACH IDLINK .

SAVE OUTFILE= ASTALLR2 .

MATCH FILES
  / FILE  = ASTALLR2
  / TABLE = ATGALLR2
  / BY IDCNTRY IDTEACH IDLINK .

SELECT IF NOT(MISSING(ATBGAGE)) .

RECODE ATBGAGE (1,2=1) (3=2) (4=3) (5,6=4) (ELSE=SYSMIS) INTO ATBGAGE.

VALUE LABELS
 IDCNTRY < list country formats > .

VALUE LABELS
 ATBGAGE
     1 '29 YEARS OR UNDER'
     2 '30-39 YEARS OLD'
     3 '40-49 YEARS OLD'
     4 '50 YEARS OR OLDER' .

SAVE OUTFILE = MERGED .

INCLUDE "<macpath>JACKPV.SPS" .

JACKPV INFILE = MERGED          /
       CVAR   = IDCNTRY ATBGAGE /
       ROOTPV = ASRREA0         /
       NPV    = 5               /
       JKZ    = JKZONE          /
       JKR    = JKREP           /
       NJKZ   = 75              /
       WGT    = TCHWGT          .

PRINT FORMATS IDCNTRY ATBGAGE N (F6.0)
              TCHWGT (F10.0) MNPV MNPV_SE PCT PCT_SE (F6.2) .

REPORT FORMAT = LIST AUTOMATIC
  / VAR = ATBGAGE (LABEL) N TCHWGT MNPV MNPV_SE PCT PCT_SE
  / BREAK = IDCNTRY (LABEL) .
```

Exhibit 4.15 Output for Example Teacher Variable Analysis (EXAMPLE 3)

COUNTRY ID	GEN/AGE OF TEACHER	N	TCHWGT	MNPV	MNPV_SE	PCT	PCT_SE
AUSTRIA	29 YEARS OR UNDER	320	5289	540.32	6.34	6.41	1.65
AUSTRIA	30-39 YEARS OLD	822	13629	541.44	4.68	16.52	2.48
AUSTRIA	40-49 YEARS OLD	2115	32431	536.95	3.38	39.30	3.01
AUSTRIA	50 YEARS OR OLDER	1812	31169	538.79	3.47	37.77	3.17
BELGIUM FLEMISH	29 YEARS OR UNDER	1202	18077	544.18	3.40	27.71	3.22
BELGIUM FLEMISH	30-39 YEARS OLD	1321	18825	545.51	3.41	28.86	3.34
BELGIUM FLEMISH	40-49 YEARS OLD	1336	19196	554.58	2.08	29.43	2.85
BELGIUM FLEMISH	50 YEARS OR OLDER	551	9139	538.74	5.50	14.01	2.20
BELGIUM FRENCH	29 YEARS OR UNDER	718	7158	493.42	8.30	15.75	2.18
BELGIUM FRENCH	30-39 YEARS OLD	1443	15013	495.70	4.07	33.04	3.17
BELGIUM FRENCH	40-49 YEARS OLD	1569	16708	507.54	3.99	36.77	3.55
BELGIUM FRENCH	50 YEARS OR OLDER	614	6557	496.10	7.24	14.43	2.58
BULGARIA	29 YEARS OR UNDER	100	2135	519.46	21.13	3.47	1.39
BULGARIA	30-39 YEARS OLD	948	15916	541.78	12.08	25.90	2.80
BULGARIA	40-49 YEARS OLD	1603	24384	551.45	5.80	39.68	3.98
BULGARIA	50 YEARS OR OLDER	1089	19013	547.65	7.44	30.94	3.87
		.	.	.	.	.	
SPAIN	29 YEARS OR UNDER	448	39572	499.86	12.17	10.78	2.42
SPAIN	30-39 YEARS OLD	630	57765	523.01	5.52	15.73	2.57
SPAIN	40-49 YEARS OLD	825	91095	511.17	5.23	24.81	3.42
SPAIN	50 YEARS OR OLDER	1921	178735	512.97	3.79	48.68	3.81
SWEDEN	29 YEARS OR UNDER	428	8179	557.13	5.18	8.59	1.65
SWEDEN	30-39 YEARS OLD	1307	27345	550.14	4.12	28.71	3.33
SWEDEN	40-49 YEARS OLD	1083	23098	542.76	3.90	24.25	2.88
SWEDEN	50 YEARS OR OLDER	1697	36636	549.33	3.11	38.46	3.61
TRINIDAD AND TOBAGO	29 YEARS OR UNDER	399	1879	427.77	15.00	11.35	2.54
TRINIDAD AND TOBAGO	30-39 YEARS OLD	1411	6125	428.56	7.69	36.99	3.62
TRINIDAD AND TOBAGO	40-49 YEARS OLD	1012	4592	441.24	9.65	27.73	3.34
TRINIDAD AND TOBAGO	50 YEARS OR OLDER	992	3963	444.38	13.68	23.93	3.00
UNITED STATES	29 YEARS OR UNDER	1011	664097	536.49	8.16	20.72	2.75
UNITED STATES	30-39 YEARS OLD	1376	849591	534.22	4.61	26.51	2.84
UNITED STATES	40-49 YEARS OLD	1252	885575	548.52	4.89	27.63	3.72
UNITED STATES	50 YEARS OR OLDER	1316	805499	535.80	6.93	25.13	3.44

In this example, each country's results are presented on four lines, one for each value of the recoded ATBGAGE variable. The results are presented much in the same manner as in the previous example, where the countries are identified in the first column and the second column describes the categories of ATBGAGE being reported. From the first four lines displayed in Exhibit 4.15, we see that in Austria 6.41 percent of students were taught by teachers 29 years or younger, 16.52 percent by teachers 30 to 39 years old, 39.30 percent by teachers 40 to 49 years old, and 37.77 percent by teachers 50 years or older.

4.8 Performing Analyses with School Variables

PIRLS 2006 has representative samples of schools and so it is possible to compute weighted numbers of schools with particular characteristics for providing reasonable estimates of percentages and means across primary schools in each

country. However, the school samples were designed to optimize the student samples and the student-level estimates. For this reason, it is preferable to analyze school-level variables as attributes of students, rather than as elements in their own right.

For student-weighted analyses, the school-level data are analyzed to make statements about the students attending schools with one characteristic or another, rather than about the schools with one characteristic or another. When school-level variables are analyzed, we recommend that you merge the school background data files with the student background data files, and then use the sampling and weight information contained in the student-level file to make the desired analytical statements.

Our example of a school-level analysis will investigate the percentage of students who attend schools in areas of different population sizes, and their average reading achievement. The results of this analysis are presented in Exhibit 7.1 of the *PIRLS 2006 International Report* and are reproduced here in Exhibit 4.16. We will use the macro JACKPV in order to estimate the percentages and average achievement.

The first step in our analysis is to identify the variables of interest in the appropriate files and review the documentation on specific national adaptations to the questions of interest (Supplements 1 and 2). We observe that the variable ACBGCOMM in the school background data files contains information on the size of the community where the schools are located. Our next step is to review the documentation of national adaptations to the questionnaire to ensure that there were no deviations listed for this variable. If no adaptations were made, we can continue with our analysis without any modifications.

Since we are using a school-level variable, we need to look in the school background data files and link this to the student background data files to get the achievement scores. From the ACG data files, we will need the variable that contains the information on the schools' location (ACBGCOMM) and the identification variables IDCNTRY and IDSCHOOL that will allow us to link the school data to the student data.

We then read the variables of interest from the student background data files. We will need the country and school identification variables (IDCNTRY and IDSCHOOL), necessary to merge the school data to the student data. We also need the jackknife replication variables (JKZONE and JKREP), the student weighting variable (TOTWGT), and the reading achievement plausible values (ASRREA01 through ASRREA05).

Exhibit 4.16 Example School Variable Analysis Taken from the PIRLS 2006 International Report (Exhibit 7.1)

Exhibit 7.1 **Principals' Reports on Their Schools' Locations with Trends** — **PIRLS 2006** 4th Grade

Countries	Urban			Suburban			Rural		
	2006 Percent of Students	Average Achievement	Difference in Percent from 2001	2006 Percent of Students	Average Achievement	Difference in Percent from 2001	2006 Percent of Students	Average Achievement	Difference in Percent from 2001
Austria	31 (3.4)	529 (4.4)	◊ ◊	20 (3.2)	542 (5.5)	◊ ◊	48 (3.7)	543 (2.5)	◊ ◊
Belgium (Flemish)	21 (3.6)	541 (5.4)	◊ ◊	37 (4.7)	546 (3.7)	◊ ◊	42 (4.8)	551 (2.4)	◊ ◊
Belgium (French) r	47 (4.0)	494 (5.5)	◊ ◊	21 (3.9)	496 (5.7)	◊ ◊	33 (3.8)	512 (4.2)	◊ ◊
Bulgaria	70 (3.0)	557 (5.2)	6 (4.1)	5 (1.7)	550 (10.8)	−6 (3.1) ▼	24 (2.5)	516 (10.8)	0 (3.4)
Canada, Alberta	46 (4.5)	559 (3.8)	◊ ◊	26 (3.5)	572 (4.4)	◊ ◊	28 (3.5)	550 (4.1)	◊ ◊
Canada, British Columbia	38 (4.3)	555 (4.3)	◊ ◊	46 (4.7)	565 (3.5)	◊ ◊	17 (3.2)	545 (6.2)	◊ ◊
Canada, Nova Scotia	25 (3.2)	542 (6.5)	◊ ◊	26 (3.2)	551 (3.9)	◊ ◊	50 (3.5)	537 (2.6)	◊ ◊
Canada, Ontario	51 (4.9)	549 (4.1)	8 (6.9)	36 (5.0)	563 (4.3)	1 (7.0)	14 (3.2)	552 (4.9)	−9 (5.1)
Canada, Quebec	51 (4.8)	533 (3.7)	15 (6.7) ▲	28 (4.0)	538 (6.5)	−19 (6.3) ▼	20 (3.6)	528 (5.6)	3 (5.0)
Chinese Taipei	– –	– –	– –	– –	– –	– –	– –	– –	– –
Denmark	33 (4.1)	545 (3.4)	◊ ◊	30 (3.4)	555 (3.9)	◊ ◊	37 (4.0)	542 (4.4)	◊ ◊
England r	45 (4.2)	523 (5.3)	−2 (6.5)	35 (3.9)	553 (5.1)	4 (6.0)	19 (3.7)	564 (5.2)	−1 (5.2)
France	34 (4.0)	522 (4.3)	−2 (5.7)	25 (3.9)	518 (6.3)	−5 (5.7)	41 (3.9)	524 (2.5)	6 (5.3)
Georgia	42 (3.6)	486 (4.5)	◊ ◊	15 (2.7)	465 (8.1)	◊ ◊	43 (2.6)	459 (5.1)	◊ ◊
Germany	37 (3.3)	535 (4.6)	4 (4.4)	19 (3.1)	557 (3.9)	−3 (4.3)	44 (4.0)	555 (2.3)	−1 (5.5)
Hong Kong SAR	58 (4.4)	573 (3.1)	6 (5.3)	37 (4.1)	555 (4.5)	−9 (4.9)	5 (1.9)	540 (11.0)	3 (2.2)
Hungary	28 (2.2)	565 (6.6)	0 (3.3)	40 (2.5)	557 (5.0)	5 (3.4)	31 (1.8)	528 (4.7)	−5 (2.5) ▼
Iceland r	33 (0.3)	518 (2.0)	−3 (0.5) ▼	37 (0.3)	509 (2.3)	−6 (0.5) ▼	30 (0.4)	506 (2.3)	9 (0.5) ▲
Indonesia	12 (2.2)	451 (9.7)	◊ ◊	14 (2.7)	425 (9.6)	◊ ◊	74 (2.9)	393 (4.8)	◊ ◊
Iran, Islamic Rep. of	50 (2.9)	454 (4.1)	1 (4.7)	15 (2.2)	415 (10.3)	6 (3.3) ▲	35 (2.8)	376 (5.7)	−8 (4.2)
Israel	49 (3.9)	534 (6.0)	−3 (5.5)	18 (2.8)	529 (13.1)	−4 (4.3)	33 (3.8)	472 (10.2)	7 (5.0)
Italy	70 (3.6)	554 (2.9)	−6 (4.8)	15 (2.8)	555 (9.2)	1 (3.8)	15 (3.1)	533 (9.9)	5 (3.8)
Kuwait	26 (3.6)	355 (7.6)	◊ ◊	61 (4.0)	321 (5.8)	◊ ◊	14 (3.0)	311 (12.1)	◊ ◊
Latvia	70 (0.7)	548 (2.4)	26 (3.9) ▲	3 (1.5)	528 (6.6)	−15 (4.2) ▼	27 (1.7)	525 (5.9)	−10 (3.4) ▼
Lithuania	72 (2.3)	544 (1.9)	1 (3.6)	3 (1.4)	549 (10.7)	−3 (2.5)	26 (2.1)	516 (3.4)	3 (3.2)
[1] Luxembourg	– –	– –	– –	– –	– –	– –	– –	– –	– –
Macedonia, Rep. of r	51 (3.6)	477 (6.9)	−6 (4.9)	18 (3.5)	443 (13.8)	3 (4.5)	31 (2.6)	401 (9.5)	3 (4.1)
Moldova, Rep. of	29 (2.4)	517 (4.5)	2 (4.1)	6 (2.4)	498 (17.6)	−8 (4.5)	65 (2.5)	492 (3.9)	6 (4.5)
Morocco r	37 (3.3)	363 (7.2)	−3 (5.5)	18 (3.6)	334 (15.8)	−3 (5.7)	45 (3.7)	296 (13.5)	6 (5.5)
Netherlands	26 (4.0)	538 (4.1)	−5 (5.6)	33 (4.7)	553 (3.2)	10 (6.0)	41 (3.5)	547 (2.5)	−5 (5.2)
New Zealand	41 (3.2)	536 (3.3)	3 (5.1)	39 (3.0)	527 (3.6)	−1 (4.7)	21 (2.3)	535 (5.3)	−2 (3.7)
Norway	20 (3.6)	502 (3.8)	1 (5.0)	30 (3.9)	504 (3.9)	4 (5.5)	50 (4.2)	492 (4.2)	−6 (5.5)
Poland	52 (2.1)	528 (2.9)	◊ ◊	5 (1.7)	529 (11.0)	◊ ◊	43 (1.9)	508 (3.8)	◊ ◊
Qatar	65 (0.2)	362 (1.5)	◊ ◊	32 (0.2)	336 (2.0)	◊ ◊	3 (0.0)	318 (8.4)	◊ ◊
Romania	47 (2.2)	515 (6.5)	−3 (3.4)	5 (1.9)	498 (14.8)	1 (2.8)	48 (2.4)	462 (8.0)	2 (3.6)
Russian Federation	63 (2.0)	581 (3.4)	8 (3.2) ▲	6 (1.3)	563 (8.8)	4 (2.3)	31 (2.2)	532 (6.1)	−12 (3.3) ▼
Scotland r	32 (3.5)	517 (4.8)	−2 (5.6)	36 (4.3)	539 (5.3)	−3 (6.9)	32 (3.9)	528 (6.8)	6 (5.9)
Singapore	100 (0.0)	558 (2.9)	0 (0.0)	0 (0.0)	~ ~	0 (0.0)	0 (0.0)	~ ~	0 (0.0)
Slovak Republic	52 (3.0)	544 (2.9)	2 (4.8)	8 (2.5)	537 (8.0)	−2 (3.5)	40 (3.3)	512 (5.9)	0 (4.5)
Slovenia	36 (4.2)	529 (3.6)	−4 (5.3)	37 (4.0)	520 (3.1)	10 (5.5)	27 (3.7)	512 (3.6)	−6 (4.7)
South Africa	17 (1.8)	350 (19.5)	◊ ◊	21 (2.2)	381 (14.9)	◊ ◊	62 (2.0)	261 (3.8)	◊ ◊
Spain	58 (4.3)	524 (3.4)	◊ ◊	20 (3.3)	497 (6.8)	◊ ◊	21 (3.4)	498 (7.1)	◊ ◊
Sweden	27 (4.1)	549 (3.8)	12 (5.1) ▲	55 (4.1)	549 (3.4)	−12 (5.6) ▼	18 (2.8)	550 (4.7)	0 (4.5)
Trinidad and Tobago	19 (2.5)	470 (13.0)	◊ ◊	50 (3.7)	441 (7.0)	◊ ◊	32 (3.0)	408 (8.9)	◊ ◊
United States	28 (3.5)	524 (4.4)	−5 (4.9)	47 (3.9)	550 (3.2)	13 (6.1) ▲	25 (2.7)	539 (9.1)	−8 (4.2)
International Avg.	43 (0.5)	508 (1.0)		24 (0.5)	501 (1.4)		33 (0.5)	483 (1.1)	

▲ Percent in 2006 significantly higher
▼ Percent in 2006 significantly lower

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

We then proceed to merge the school data with the student data using the variables IDCNTRY and IDSCHOOL, and then use the macro JACKPV to obtain the percentages of students and their mean achievement scores within each category of the variable ACBGCOMM for each country. For this analysis, we will

use the data for all available countries, making use of an aggregated school file ACGALLR2 and an aggregated student file ASGALLR2.

The SPSS program that implements this fourth example is presented in Exhibit 4.17 and is included on the DVD under the name EXAMPLE4.SPS. The results of this program are displayed in Exhibit 4.18, edited to show only the first five and last five countries, alphabetically, for the sake of brevity. Note that one of the steps in this program is to select only those students who have non-missing data in our variable of interest ACBGCOMM.

In general, to perform school-level analyses using the school background data files, you should do the following:

1) Identify the variables of interest in the school background data files and note any specific national adaptations to the variables
2) Retrieve the relevant variables from the school background data files, including analysis and classification variables, identification variables (IDCNTRY and IDSCHOOL), and any other variables used in the selection of cases
3) Retrieve the relevant variables from the student background data files, including plausible values for achievement, classification variables, identification variables (IDCNTRY and IDSCHOOL), sampling (JKZONE and JKREP) and weighting (TOTWGT) variables, and any other variables used in the selection of cases
4) Merge the school background data files with the student background data files using the variables IDCNTRY and IDSCHOOL
5) Perform any necessary variable transformations or recodes
6) Use the macros JACKGEN and JACKREG, or JACKPV and JACKREGP if plausible values are involved, with the appropriate arguments and parameters
7) Specify the location of the data files (<datpath>) and the macros (<macpath>)
8) Print out the results file

Exhibit 4.17 First Example SPSS Program to Perform School Variable Analysis (EXAMPLE4.SPS)

```
GET FILE = "<datpath>ACGALLR2.SAV" .

SORT CASES BY IDCNTRY IDSCHOOL .

SAVE OUTFILE = ACGALLR2 .

GET FILE = "<datpath>ASGALLR2.SAV" .

SORT CASES BY IDCNTRY IDSCHOOL .

SAVE OUTFILE= ASGALLR2 .

MATCH FILES
  / FILE  = ASGALLR2
  / TABLE = ACGALLR2
  / BY IDCNTRY IDSCHOOL .

SELECT IF NOT(MISSING(ACBGCOMM)) .

VALUE LABELS
 IDCNTRY < list country formats > .

VALUE LABELS
 ACBGCOMM
     1 'URBAN'
     2 'SUBURBAN'
     3 'RURAL' .

SAVE OUTFILE = MERGED .

INCLUDE "<macpath>JACKPV.SPS" .

JACKPV INFILE = MERGED           /
       CVAR   = IDCNTRY ACBGCOMM /
       ROOTPV = ASRREA0          /
       NPV    = 5                /
       JKZ    = JKZONE           /
       JKR    = JKREP            /
       NJKZ   = 75               /
       WGT    = TOTWGT           .

PRINT FORMATS IDCNTRY ACBGCOMM N (F6.0)
              TOTWGT (F10.0) MNPV MNPV_SE PCT PCT_SE (F6.2) .

REPORT FORMAT = LIST AUTOMATIC
  / VAR = ACBGCOMM (LABEL) N TOTWGT MNPV MNPV_SE PCT PCT_SE
  / BREAK = IDCNTRY (LABEL) .
```

Exhibit 4.18 Output for First Example School Variable Analysis (EXAMPLE 4)

COUNTRY ID	GEN/CHARACTER OF SCH LOCATION AREA	N	TOTWGT	MNPV	MNPV_SE	PCT	PCT_SE
AUSTRIA	URBAN	1848	25951	528.62	4.43	31.40	3.39
AUSTRIA	SUBURBAN	1235	16743	541.75	5.46	20.26	3.24
AUSTRIA	RURAL	1949	39956	543.12	2.51	48.34	3.69
BELGIUM FLEMISH	URBAN	909	13199	540.79	5.39	20.80	3.61
BELGIUM FLEMISH	SUBURBAN	1592	23749	545.58	3.67	37.43	4.72
BELGIUM FLEMISH	RURAL	1777	26498	550.93	2.45	41.77	4.81
BELGIUM FRENCH	URBAN	2055	18745	493.55	5.47	46.72	3.98
BELGIUM FRENCH	SUBURBAN	785	8258	496.47	5.71	20.58	3.93
BELGIUM FRENCH	RURAL	1022	13121	511.77	4.16	32.70	3.85
BULGARIA	URBAN	2976	42641	557.48	5.21	70.41	3.03
BULGARIA	SUBURBAN	221	3158	549.62	10.83	5.22	1.75
BULGARIA	RURAL	467	14760	516.24	10.84	24.37	2.46
CANADA ALBERTA	URBAN	1830	16556	559.40	3.84	46.07	4.51
CANADA ALBERTA	SUBURBAN	1325	9429	571.86	4.36	26.23	3.46
CANADA ALBERTA	RURAL	1018	9955	550.29	4.09	27.70	3.49
		.	.	.	.	.	
SOUTH AFRICA	URBAN	2146	139387	350.20	19.47	17.29	1.77
SOUTH AFRICA	SUBURBAN	2754	166761	380.80	14.93	20.68	2.20
SOUTH AFRICA	RURAL	9469	500254	261.11	3.75	62.04	2.04
SPAIN	URBAN	2349	208013	524.21	3.40	58.36	4.26
SPAIN	SUBURBAN	732	72090	496.57	6.76	20.23	3.33
SPAIN	RURAL	638	76330	498.21	7.07	21.42	3.37
SWEDEN	URBAN	1368	27171	548.74	3.76	26.99	4.11
SWEDEN	SUBURBAN	2507	55264	548.99	3.35	54.89	4.08
SWEDEN	RURAL	479	18242	550.01	4.71	18.12	2.77
TRINIDAD AND TOBAGO	URBAN	907	3162	469.93	13.03	18.74	2.55
TRINIDAD AND TOBAGO	SUBURBAN	1954	8380	441.49	7.04	49.66	3.68
TRINIDAD AND TOBAGO	RURAL	992	5333	407.74	8.87	31.60	3.01
UNITED STATES	URBAN	1461	908132	524.07	4.45	27.50	3.50
UNITED STATES	SUBURBAN	2290	1563272	549.59	3.22	47.34	3.94
UNITED STATES	RURAL	1349	830812	538.95	9.13	25.16	2.72

In this example, each country's results are presented on three lines, one for each value of the ACBGCOMM variable, with the country name in the first column and label of categories of ACBGCOMM being reported in the second column. All other columns are identical in nature to those in the previous two examples.

From the first three lines in Exhibit 4.18, we see that 31.40 percent of Austrian students come from schools in urban areas, 20.26 percent come from schools in suburban areas, and 48.34 percent from rural areas. We also see that the estimated average reading achievement of students in urban areas is 528.62 (with a standard error of 4.43), whereas the estimated average achievement of students in suburban and rural areas are, respectively, 541.75 (standard error of 5.46) and 543.12 (standard error of 2.51).

4.9 Performing Analyses with Home Background Variables

In PIRLS 2006, the students' parents or primary caregivers were asked to complete a questionnaire. The responses to the *Learning to Read Survey* are included in the home background data files. Although home background variables are located in separate files, they are in essence attributes of students and must be analyzed in the same manner as student background variables. This will require us to merge the home background data files with the student background data files.

Our example of a home background analysis will investigate the relationship between the frequency of parents' reading for enjoyment and their children's reading achievement. These results are presented in Exhibit 3.9 of the *PIRLS 2006 International Report* and are duplicated here in Exhibit 4.19. We will use the macro JACKPV in order to estimate the percentages and mean achievement.

The first step in our analysis is to identify the variables of interest in the appropriate files and review the documentation on specific national adaptations to the questions of interest (Supplements 1 and 2). We observe that the variable ASBHRRE in the home background data files contains categorical information on the frequency of parents' reading for enjoyment. Our next step is to review the documentation of national adaptations to the questionnaires to ensure that there were no deviations listed for this variable (see Supplement 2). If no national adaptations were made, we can continue with our analysis without any modifications.

Since we are interested in using a home background variable, as well as student achievement in our analysis, we need to link the home background data files and the student background data files to get the achievement scores. From the ASH data files, we will need our variable of interest (ASBHRRE) and the identification variables IDCNTRY and IDSTUD that will allow us to link the home data to the student data. From the ASG data files, we will need the country and student identification variables (IDCNTRY and IDSTUD) necessary to merge the home data to the student data. We also need the jackknife replication variables (JKZONE and JKREP), the student weighting variable (TOTWGT), and the reading achievement plausible values (ASRREA01 through ASRREA05).

We then proceed to merge the home and student data files by IDCNTRY and IDSTUD. The macro JACKPV is used to obtain the percentages of students and average achievement scores within each category of the variable ASBHRRE for each country, making use of an aggregated home file ASHALLR2 and an aggregated student file ASGALLR2.

Exhibit 4.19 Example Home Background Variable Analysis Taken from the PIRLS 2006 International Report (Exhibit 3.9)

Exhibit 3.9 **Parents Reading for Enjoyment with Trends** — **PIRLS 2006** 4th Grade

Countries		Every Day or Almost Every Day			Once or Twice a Week			Twice a Month or Less		
		2006 Percent of Students	Average Achievement	Difference in Percent from 2001	2006 Percent of Students	Average Achievement	Difference in Percent from 2001	2006 Percent of Students	Average Achievement	Difference in Percent from 2001
Norway		66 (1.0)	505 (2.5)	7 (1.6) ▲	23 (0.9)	493 (4.2)	−4 (1.4) ▼	11 (0.7)	489 (6.5)	−3 (0.9) ▼
Sweden		64 (0.9)	556 (2.4)	−3 (1.3) ▼	21 (0.8)	546 (2.9)	−1 (1.1)	15 (0.7)	540 (4.6)	5 (0.9) ▲
Scotland	s	63 (1.3)	550 (4.0)	1 (1.8)	23 (0.9)	531 (5.3)	−3 (1.5)	14 (1.1)	525 (7.0)	2 (1.4)
Iceland	r	61 (0.9)	523 (1.9)	−1 (1.2)	21 (0.8)	513 (2.9)	−4 (1.1) ▼	18 (0.6)	498 (3.3)	5 (0.8) ▲
New Zealand	s	60 (1.1)	557 (2.4)	4 (1.6) ▲	25 (0.9)	539 (3.6)	−4 (1.4) ▼	15 (0.7)	528 (4.1)	0 (1.1)
Netherlands	s	59 (1.2)	562 (1.8)	−1 (1.8)	25 (0.9)	548 (2.9)	−3 (1.3) ▼	16 (0.9)	543 (2.6)	4 (1.2) ▲
Trinidad and Tobago		59 (0.9)	441 (5.1)	◊ ◊	31 (1.0)	436 (6.1)	◊ ◊	10 (0.7)	434 (9.2)	◊ ◊
Germany	r	58 (0.9)	561 (2.4)	7 (1.3) ▲	28 (0.7)	544 (2.9)	−3 (1.0) ▼	14 (0.6)	535 (4.5)	−3 (0.9) ▼
Denmark		57 (1.0)	554 (2.6)	◊ ◊	27 (0.9)	544 (3.0)	◊ ◊	16 (0.7)	534 (5.0)	◊ ◊
Latvia		57 (1.3)	547 (2.7)	3 (2.0)	32 (1.1)	537 (3.4)	−3 (1.5)	11 (0.6)	532 (4.2)	0 (1.2)
Canada, British Columbia	r	57 (1.1)	569 (3.2)	◊ ◊	29 (1.1)	560 (3.5)	◊ ◊	15 (0.7)	547 (4.7)	◊ ◊
Canada, Alberta	r	55 (1.2)	571 (2.8)	◊ ◊	29 (1.0)	561 (3.5)	◊ ◊	17 (0.7)	554 (3.7)	◊ ◊
Luxembourg		54 (0.7)	573 (1.3)	◊ ◊	29 (0.7)	544 (2.4)	◊ ◊	18 (0.6)	543 (2.4)	◊ ◊
Canada, Nova Scotia		53 (0.9)	556 (2.1)	◊ ◊	27 (0.8)	539 (2.8)	◊ ◊	19 (0.7)	528 (3.2)	◊ ◊
Austria		53 (1.1)	549 (2.3)	◊ ◊	30 (0.9)	533 (3.2)	◊ ◊	17 (0.7)	524 (2.9)	◊ ◊
Lithuania		51 (1.1)	542 (2.0)	−1 (1.5)	36 (1.0)	532 (2.1)	−1 (1.4)	12 (0.6)	534 (2.6)	2 (0.9)
Belgium (French)		51 (1.3)	512 (2.9)	◊ ◊	32 (0.9)	496 (3.3)	◊ ◊	17 (0.8)	483 (4.4)	◊ ◊
Canada, Ontario	r	51 (1.0)	563 (3.4)	−2 (1.6)	32 (1.1)	552 (3.0)	1 (1.6)	18 (0.9)	547 (4.5)	1 (1.3)
France		50 (1.0)	535 (2.5)	−1 (1.4)	32 (0.8)	518 (2.6)	−2 (1.2)	19 (0.8)	505 (3.0)	3 (1.1) ▲
Spain	s	50 (1.2)	531 (3.0)	◊ ◊	33 (1.0)	512 (3.6)	◊ ◊	18 (0.7)	507 (4.5)	◊ ◊
Singapore		50 (0.7)	569 (3.0)	14 (1.1) ▲	33 (0.7)	552 (3.1)	−3 (0.9) ▼	17 (0.5)	546 (4.3)	−11 (0.9) ▼
Canada, Quebec	r	49 (1.4)	542 (3.2)	1 (1.9)	32 (1.2)	536 (3.5)	−2 (1.7)	18 (0.9)	518 (4.1)	1 (1.4)
Italy		49 (1.0)	564 (3.0)	2 (1.4)	32 (1.0)	547 (3.7)	−4 (1.2) ▼	19 (0.9)	535 (4.2)	2 (1.2)
Hungary		49 (1.0)	561 (3.2)	−2 (1.4)	37 (0.8)	546 (3.7)	1 (1.2)	14 (0.6)	544 (4.4)	1 (0.9)
Macedonia, Rep. of	r	48 (1.1)	459 (4.6)	7 (1.7) ▲	41 (0.9)	441 (4.5)	0 (1.4)	11 (0.9)	423 (8.4)	−7 (1.5) ▼
South Africa	r	48 (0.7)	322 (7.5)	◊ ◊	36 (0.5)	301 (5.2)	◊ ◊	16 (0.5)	276 (6.3)	◊ ◊
Israel		47 (1.4)	538 (5.0)	x x	39 (1.2)	520 (4.4)	x x	15 (0.7)	508 (6.5)	x x
Slovak Republic		46 (1.0)	544 (2.4)	−2 (1.5)	39 (1.0)	529 (3.2)	0 (1.4)	14 (0.8)	504 (7.7)	3 (1.0) ▲
Slovenia		45 (0.8)	530 (2.4)	4 (1.3) ▲	36 (0.8)	520 (2.6)	−6 (1.2) ▼	18 (0.6)	510 (3.3)	2 (0.9) ▲
Poland		45 (0.9)	529 (3.0)	◊ ◊	41 (0.8)	517 (2.8)	◊ ◊	14 (0.6)	504 (4.5)	◊ ◊
Qatar	r	44 (0.7)	361 (2.3)	◊ ◊	39 (0.7)	357 (2.3)	◊ ◊	17 (0.5)	348 (3.8)	◊ ◊
Kuwait	r	44 (1.2)	339 (5.4)	◊ ◊	36 (0.9)	337 (5.4)	◊ ◊	20 (0.9)	335 (5.9)	◊ ◊
Russian Federation		42 (1.0)	573 (3.7)	−7 (1.6) ▼	41 (0.8)	560 (3.9)	6 (1.4) ▲	17 (0.6)	557 (3.6)	1 (1.1)
Belgium (Flemish)		40 (0.9)	558 (2.2)	◊ ◊	33 (0.8)	545 (2.4)	◊ ◊	27 (1.0)	538 (2.4)	◊ ◊
Georgia		39 (1.3)	486 (3.4)	◊ ◊	44 (1.0)	467 (3.2)	◊ ◊	17 (1.2)	450 (7.6)	◊ ◊
Bulgaria		39 (1.4)	565 (4.2)	−14 (2.1) ▼	34 (1.1)	547 (4.7)	6 (1.6) ▲	27 (2.0)	532 (7.6)	8 (2.6) ▲
Hong Kong SAR		36 (0.8)	574 (2.3)	6 (1.4) ▲	39 (0.7)	561 (2.6)	10 (1.0) ▲	25 (0.6)	558 (3.4)	−15 (1.3) ▼
Chinese Taipei		35 (0.9)	547 (2.5)	◊ ◊	42 (0.7)	536 (2.5)	◊ ◊	23 (0.7)	520 (2.4)	◊ ◊
Moldova, Rep. of		35 (1.1)	509 (3.3)	−1 (1.9)	46 (1.1)	498 (3.5)	6 (1.7) ▲	20 (1.5)	491 (5.9)	−5 (1.9) ▼
Indonesia		31 (1.2)	413 (4.6)	◊ ◊	44 (1.3)	405 (4.5)	◊ ◊	25 (1.3)	395 (4.9)	◊ ◊
Romania		27 (1.1)	514 (4.7)	11 (1.6) ▲	43 (1.0)	497 (4.8)	11 (1.7) ▲	30 (1.4)	459 (8.0)	−22 (2.4) ▼
Morocco		24 (1.1)	346 (6.4)	– –	34 (1.4)	329 (5.3)	– –	42 (1.9)	308 (10.9)	– –
Iran, Islamic Rep. of		24 (1.1)	441 (4.9)	2 (1.5)	41 (1.1)	426 (3.4)	6 (1.8) ▲	35 (1.6)	402 (4.3)	−8 (2.2) ▼
England		x x	x x	x x	x x	x x	x x	x x	x x	x x
United States		– –	– –	– –	– –	– –	– –	– –	– –	– –
International Avg.		47 (0.2)	512 (0.6)		34 (0.2)	498 (0.6)		18 (0.2)	487 (0.9)	

▲ Percent in 2006 significantly higher
▼ Percent in 2006 significantly lower

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

The SPSS program that implements this example is presented in Exhibit 4.20 and is included on the DVD under the name EXAMPLE5.SPS. The results obtained from this program are displayed in Exhibit 4.21, showing the first five and last five countries, as sorted alphabetically. Note that one of the steps in this program is to select only those students who have non-missing data in our variable of interest

ASBHRRE. A second step consists of combining category 4 of the variable ASBHRRE with category 3 to create a category we will label "Twice a Month or Less".

In general, to perform analyses with home background variables using the home background data files, you should do the following:

1) Identify the variables of interest in the home background data files and note any specific national adaptations to the variables
2) Retrieve the relevant variables from the home background data files, including analysis and classification variables, identification variables (IDCNTRY and IDSTUD), and any other variables used in the selection of cases
3) Retrieve the relevant variables from the student background data files, including plausible values for achievement, classification variables, identification variables (IDCNTRY and IDSTUD), sampling (JKZONE and JKREP) and weighting (TOTWGT) variables, and any other variables used in the selection of cases
4) Merge the home background data files with the student background data files using the variables IDCNTRY and IDSTUD
5) Perform any necessary variable transformations or recodes
6) Use the macros JACKGEN and JACKREG, or JACKPV and JACKREGP if plausible values are involved, with the appropriate arguments and parameters
7) Specify the location of the data files (<datpath>) and the macros (<macpath>)
8) Print out the results file

Exhibit 4.20 Example SPSS Program to Perform Home Background Variable Analysis (EXAMPLE5.SPS)

```
GET FILE = "<datpath>ASHALLR2.SAV" .

SORT CASES BY IDCNTRY IDSTUD .

SAVE OUTFILE = ASHALLR2 .

GET FILE = "<datpath>ASGALLR2.SAV" .

SORT CASES BY IDCNTRY IDSTUD .

SAVE OUTFILE= ASGALLR2 .

MATCH FILES
  / FILE  = ASGALLR2
  / TABLE = ASHALLR2
  / BY IDCNTRY IDSTUD .

SELECT IF NOT(MISSING(ASBHRRE)) .

RECODE ASBHRRE (1=1) (2=2) (3,4=3) (ELSE=SYSMIS) INTO ASBHRRE .

VALUE LABELS
 IDCNTRY < list country formats > .

 VALUE LABELS
  ASBHRRE
     1 'EVERY DAY OR ALMOST EVERY DAY'
     2 'ONCE OR TWICE A WEEK'
     3 'TWICE A MONTH OR LESS' .

SAVE OUTFILE = MERGED .

INCLUDE "<macpath>JACKPV.SPS" .

JACKPV INFILE = MERGED          /
       CVAR   = IDCNTRY ASBHRRE /
       ROOTPV = ASRREA0         /
       NPV    = 5               /
       JKZ    = JKZONE          /
       JKR    = JKREP           /
       NJKZ   = 75              /
       WGT    = TOTWGT          .

PRINT FORMATS IDCNTRY ASBHRRE N (F6.0)
              TOTWGT (F10.0) MNPV MNPV_SE PCT PCT_SE (F6.2) .

REPORT FORMAT = LIST AUTOMATIC
  / VAR = ASBHRRE (LABEL) N TOTWGT MNPV MNPV_SE PCT PCT_SE
  / BREAK = IDCNTRY (LABEL) .
```

Exhibit 4.21 Output for Example Home Background Variable Analysis (EXAMPLE 6)

COUNTRY ID	GEN/HW OFTEN YOU READ FOR OWN ENJOYMENT	N	TOTWGT	MNPV	MNPV_SE	PCT	PCT_SE
AUSTRIA	EVERY DAY OR ALMOST EVERY DAY	2573	41727	548.84	2.30	52.91	1.06
AUSTRIA	ONCE OR TWICE A WEEK	1440	23805	532.70	3.16	30.18	.86
AUSTRIA	TWICE A MONTH OR LESS	787	13337	523.95	2.95	16.91	.66
BELGIUM FLEMISH	EVERY DAY OR ALMOST EVERY DAY	1753	25651	557.56	2.20	40.34	.91
BELGIUM FLEMISH	ONCE OR TWICE A WEEK	1393	20847	545.17	2.41	32.78	.83
BELGIUM FLEMISH	TWICE A MONTH OR LESS	1174	17090	538.20	2.45	26.88	.96
BELGIUM FRENCH	EVERY DAY OR ALMOST EVERY DAY	2087	21692	511.68	2.90	50.95	1.29
BELGIUM FRENCH	ONCE OR TWICE A WEEK	1302	13474	496.11	3.25	31.65	.93
BELGIUM FRENCH	TWICE A MONTH OR LESS	693	7410	483.28	4.43	17.40	.83
BULGARIA	EVERY DAY OR ALMOST EVERY DAY	1540	23568	564.81	4.24	38.84	1.44
BULGARIA	ONCE OR TWICE A WEEK	1294	20638	547.33	4.69	34.01	1.14
BULGARIA	TWICE A MONTH OR LESS	871	16478	532.46	7.62	27.15	1.98
CANADA ALBERTA	EVERY DAY OR ALMOST EVERY DAY	1847	16063	570.51	2.81	54.76	1.20
CANADA ALBERTA	ONCE OR TWICE A WEEK	990	8394	561.38	3.52	28.61	.98
CANADA ALBERTA	TWICE A MONTH OR LESS	579	4879	553.60	3.71	16.63	.70
. . . . .							
SLOVENIA	EVERY DAY OR ALMOST EVERY DAY	2319	7558	530.38	2.42	45.42	.81
SLOVENIA	ONCE OR TWICE A WEEK	1812	6028	520.47	2.57	36.22	.84
SLOVENIA	TWICE A MONTH OR LESS	911	3056	510.26	3.30	18.37	.62
SOUTH AFRICA	EVERY DAY OR ALMOST EVERY DAY	5697	323919	321.78	7.52	48.03	.67
SOUTH AFRICA	ONCE OR TWICE A WEEK	4478	242429	300.65	5.17	35.95	.54
SOUTH AFRICA	TWICE A MONTH OR LESS	1879	108056	275.65	6.32	16.02	.52
SPAIN	EVERY DAY OR ALMOST EVERY DAY	1287	119919	530.99	3.00	49.55	1.23
SPAIN	ONCE OR TWICE A WEEK	825	79606	512.14	3.65	32.89	1.03
SPAIN	TWICE A MONTH OR LESS	443	42493	507.20	4.51	17.56	.74
SWEDEN	EVERY DAY OR ALMOST EVERY DAY	2679	61207	555.81	2.44	64.22	.88
SWEDEN	ONCE OR TWICE A WEEK	862	20167	546.35	2.95	21.16	.77
SWEDEN	TWICE A MONTH OR LESS	563	13929	539.87	4.56	14.62	.74
TRINIDAD AND TOBAGO	EVERY DAY OR ALMOST EVERY DAY	2029	8986	441.41	5.08	58.81	.92
TRINIDAD AND TOBAGO	ONCE OR TWICE A WEEK	1126	4755	436.29	6.08	31.12	.97
TRINIDAD AND TOBAGO	TWICE A MONTH OR LESS	363	1539	433.75	9.22	10.07	.66

Each country's results are presented on three lines, one for each value of the ASBHRRE variable. The countries are identified in the first column and the second column describes the reporting categories for ASBHRRE. All other columns are identical to those in the previous examples.

From the first three lines in Exhibit 4.21, we see that 52.91 percent of students in Austria have parents who reported reading for enjoyment every day or almost every day, 30.18 percent have parents who read for enjoyment once or twice a week, and 16.91 percent twice a month or less. We also see that the estimated average reading achievement of students whose parents read for enjoyment every day or almost every day is 548.84 (with a standard error of 2.30), of those whose parents read once or twice a week is 532.70 (standard error of 3.16), and of those whose parents read twice a month or less is 523.95 (standard error of 2.95).

References

Mullis, I.V.S., Martin, M.O., Kennedy, A.M., & Foy, P. (2007). *PIRLS 2006 international report: IEA's progress in international reading literacy study in primary schools in 40 countries*. Chestnut Hill, MA: Boston College.

SPSS Inc. (2005). *SPSS for Windows* (version 14.0). Chicago, IL: SPSS Inc.

Chapter 5

Performing Analyses with the PIRLS Data Using SAS

5.1 Overview

This chapter presents some basic examples of analyses that can be performed with the PIRLS 2006 international database using the SAS statistical analysis system (SAS, 2002). The analyses presented here are simple in nature, and are designed primarily to familiarize you with the various data files and their structure, as well as the variables to be used in most analyses. In general, the programs compute percentages of students in specified subgroups, average reading achievement in those subgroups, and the proper standard errors for these statistics. Additionally, some programs compute regression coefficients and their standard errors.

The example analyses, using student, home, teacher and school data, replicate some of the analyses that are included in the *PIRLS 2006 International Report* (Mullis, Martin, Kennedy, & Foy, 2007). You are encouraged to practice analyzing the PIRLS data by replicating some of the exhibits presented in the international report.

All SAS programs presented in this chapter also are available in the PIRLS database. They can be adapted to perform a variety of analyses, provided you have some basic knowledge of the SAS language. With a little experience and some practice with these programs, you will be able to make the necessary modifications to obtain the desired results. The example SAS programs invoke SAS macros that will be described in this chapter. Although you will be expected to modify the example programs, there is no need to make any changes within the SAS macros.

5.2 SAS Programs and Macros

The PROGRAMS subfolder on the DVD that accompanies this User Guide includes a number of SAS programs needed to process the SAS data files, compute survey results, and carry out example analyses. This chapter gives detailed instructions on how to adapt and make use of them. The following programs are available:

CONVERT.SAS

This SAS program converts the SAS Export files found on the DVD into SAS data files. All programs and macros described in this chapter require that the SAS Export files be converted into SAS data files.

ASASCRR2.SAS

This SAS program that can be used to convert the response codes to the test items to their corresponding score levels.

JOIN.SAS

This SAS program combines files of the same type from more than one country.

JACKGEN.SAS (and SAMPLEJACKGEN.SAS)

This SAS macro program is used to compute weighted percentages of students within defined subgroups, along with their means on a specified continuous variable. This macro generates replicate weights and computes standard errors using the jackknife repeated replication (JRR) methodology. The analysis variable can be any continuous variable. When computing mean achievement scores based on plausible values, you will need to use the macro JACKPV.SAS.

JACKPV.SAS (and SAMPLEJACKGPV.SAS)

This SAS macro program is used to compute weighted percentages of students within defined subgroups, along with their mean achievement using the available plausible values. This macro generates replicate weights and computes standard errors using the jackknife repeated replication (JRR) and multiple imputation methodologies. This macro should be used when multiple plausible values are used in an analysis.

JACKREG.SAS (and SAMPLEJACKREG.SAS)

This SAS macro program is used to compute weighted regression coefficients and their standard errors within defined subgroups. This macro can be used with any analysis variable, but is not appropriate for analyzing achievement with plausible values.

JACKREGP.SAS (and SAMPLEJACKREGP.SAS)

This SAS macro program is used to compute weighted regression coefficients and their standard errors within defined subgroups when using plausible values as achievement scores.

Each of the four SAS macros above has a corresponding sample program that calls its respective macro and prints out the results. These sample programs are discussed later in this chapter.

EXAMPLE1.SAS, EXAMPLE2.SAS, EXAMPLE3.SAS, EXAMPLE4.SAS, EXAMPLE5.SAS

These are the programs used in the example analyses presented later in this chapter.

5.3 Converting the Files

The DVD provides a program called CONVERT.SAS that converts the SAS Export files provided on the DVD into SAS data files. This conversion is necessary since all the SAS macros and SAS programs presented in this chapter require the use of SAS data files.

To convert SAS Export files into SAS data files, you should to do the following:

1) Open the SAS program file CONVERT.SAS
2) At the beginning of the program, specify the data file type in the parameter "TYPE"
3) Specify the path where the SAS Export files are located in the parameter "EXPPATH"
4) Specify the path where the converted SAS data files will be located in the parameter "DATPATH"

5) List all the countries of interest in the parameter "COUNTRY"
   By default, all PIRLS 2006 countries are listed.

6) Submit the edited code for processing

An example of the CONVERT program is presented in Exhibit 5.1. This example converts the SAS Export files of type ASG for all countries. For this example, all SAS Export files are located in the folder "D:\PIRLS2006\Data\SAS_Data" and the converted SAS data files also will be located in this folder. The CONVERT program should be run for all data file types (ACG, ASA, ASG, ASH, AST and ATG) before moving on to any data analyses, more specifically the data analysis examples in this User Guide.

Exhibit 5.1 Example of CONVERT Program Used to Convert SAS Export Files into SAS Data Files

```
%LET TYPE = ASG ;

%LET EXPPATH = D:\PIRLS2006\Data\SAS_Data\ ;

%LET DATPATH = D:\PIRLS2006\Data\SAS_Data\ ;

%LET COUNTRY = < List of PIRLS 2006 countries >;

%MACRO DOIT ;

 %LET I = 1 ;
 %DO %WHILE(%LENGTH(%SCAN(&COUNTRY,&I))) ;
 %LET CTRY = %SCAN(&COUNTRY,&i) ;

 PROC CIMPORT FILE = "&EXPPATH&TYPE&CTRY.R2.EXP"
              DATA = "&DATPATH&TYPE&CTRY.R2" ;

 %LET I = %EVAL(&I + 1) ;
 %END ;

%MEND DOIT ;

%DOIT ;
```

5.4 Scoring the Items

Two types of items were administered as part of the PIRLS assessment. There were multiple-choice items, where students were asked to select one of four options as the correct response. Numbers 1 through 4 are used to represent response options A through D, respectively, in the achievement data files. There also were constructed-response items, where students were asked to construct a written response to a question, rather than choosing an answer from a list of options. Constructed-response items were worth a total of one, two, or three points. Scorers from the national centers were trained to use the scoring guides

described in Chapter 2 to score the answers to these questions. The numbers 0 through 3 are used to represent the scored responses to these items and also represent their point values. For both types of items, special codes are set aside to represent missing data either as "Not Administered", "Omitted", or "Not Reached".

Responses to multiple-choice items may need to be converted to their appropriate score levels ("1" for correct and "0" for incorrect), as well as responses coded to the special missing codes, in order to carry out specific item-level analyses. For this purpose, the DVD includes a SAS program (ASASCRR2.SAS) which allows you to recode the items from the achievement data files to their score level. The program consists of a macro called SCOREIT and the necessary call to this macro so that all the items in the specified data files are scored. This macro will convert the response option codes for multiple-choice items to dichotomous score levels (0 or 1) based on each item's scoring key. It will also convert the special missing codes as either incorrect (0) or missing. By default, the not administered response code is left as missing and the omitted and not reached response codes as incorrect. These default settings can be modified within the SCOREIT macro, depending on the requirements of your item-level analyses. For example, not reached responses were treated as missing for the purpose of calibrating the PIRLS items, whereas they were treated as incorrect when scoring the results of individual countries and deriving achievement scores for students.

To use the SCOREIT macro, you will need to adapt the program code in the ASASCRR2 program. You should do the following:

1) Open the SAS program file ASASCRR2.SAS
2) Specify the path where the SAS data files are located in the "LIBNAME" statement
3) List all the countries of interest in the parameter "COUNTRY" By default, all PIRLS 2006 countries are listed.
4) Submit the edited code for processing

The program recodes the items and saves the results in SAS data files that have "ASC" instead of "ASA" as the first three characters. If you would like not reached responses to be treated as missing rather than incorrect, you should replace the following statement (which appears twice in the program):

```
IF &ITEM = &NR THEN SCORE = 0 ;
```

with this statement:

```
IF &ITEM = &NR THEN SCORE = . ;
```

Exhibit 5.2 shows a condensed version of the SAS program that scores the items.

Exhibit 5.2 Example of ASASCRR2 Program For Converting Item Response Codes to Their Score Level

```
LIBNAME LIBDAT "D:\PIRLS2006\Data\SAS_Data\" ;

%LET COUNTRY = < List of PIRLS 2006 countries > ;

%LET ARIGHT = < List of multiple-choice items where A is correct > ;
%LET BRIGHT = < List of multiple-choice items where B is correct > ;
%LET CRIGHT = < List of multiple-choice items where C is correct > ;
%LET DRIGHT = < List of multiple-choice items where D is correct > ;
%LET CONSTR = < List of constructed-response items > ;

%MACRO SCOREIT (ITEM, TYPE, RIGHT, NR, NA, OM, OTHER) ;

   .  .  .

%MEND SCOREIT ;

%MACRO DOIT ;

   .  .  .

  DO OVER ARIGHT ; %SCOREIT (ARIGHT, "MC", 1, .R, .A, ., .I) ; END ;
  DO OVER BRIGHT ; %SCOREIT (BRIGHT, "MC", 2, .R, .A, ., .I) ; END ;
  DO OVER CRIGHT ; %SCOREIT (CRIGHT, "MC", 3, .R, .A, ., .I) ; END ;
  DO OVER DRIGHT ; %SCOREIT (DRIGHT, "MC", 4, .R, .A, ., .I) ; END ;
  DO OVER CONSTR ; %SCOREIT (CONSTR, "CR",  , .R, .A, ., .I) ; END ;

   .  .  .

%MEND DOIT ;

%DOIT ;
```

5.5 Joining the Files

The PIRLS database DVD contains separate data files for each country. The DVD provides a SAS program called JOIN.SAS that joins individual country data files of a particular type into a single aggregated data file, facilitating joint analyses involving more than one country. This program however, can only join SAS data files of the same type, (e.g., all ASG data files). The JOIN program can be used for the following data file types: ACG, ASA, ASC, ASG, ASH, AST, and ATG.

To create a SAS data file with more than one country's data, you should do the following:

1) Open the SAS program file JOIN.SAS
2) At the beginning of the program, specify the data file type in the parameter "TYPE"
3) Specify the sorting variables in the parameter "SORTVARS"
   By default the variables IDCNTRY and IDSTUD are specified for the ASA, ASC, ASG and ASH data files. Consult Exhibit 2.10 to determine which sorting variables need to be specified for each data file type
4) Specify the path where the SAS data files are located in the "PATH" statement
5) List all the countries of interest in the parameter "COUNTRY"
   By default, all PIRLS 2006 countries are listed.
6) Submit the edited code for processing

An example of the JOIN program is displayed in Exhibit 5.3. It joins the student background data files of all countries. All country data files are located in the folder "D:\PIRLS2006\Data\SAS_Data" for the sake of this example. The resulting data file, ASGALLR2, will be saved in this folder as well. The joined data file will be sorted by the identification variables IDCNTRY and IDSTUD.

Exhibit 5.3 Example of JOIN Program Used to Join SAS Data Files for More Than One Country

```
%LET TYPE = ASG ;

%LET SORTVARS = IDCNTRY IDSTUD ;

LIBNAME LIBDAT "D:\PIRLS2006\Data\SAS_Data\" ;

%LET COUNTRY = < List of PIRLS 2006 countries >;

%MACRO DOIT ;

 DATA &TYPE.ALLR2 ;
      SET  %LET I = 1 ;
           %DO %WHILE(%LENGTH(%SCAN(&COUNTRY,&I))) ;
             %LET CTRY = %SCAN(&COUNTRY,&I) ;
                LIBDAT.&TYPE&CTRY.R2
             %LET I = %EVAL(&I + 1) ;
           %END ; ;

 PROC SORT DATA = &TYPE.ALLR2 OUT = LIBDAT.&TYPE.ALLR2 ;
      BY &SORTVARS ;

%MEND DOIT ;

%DOIT ;
```

5.6 SAS Macros to Compute Percentages, Means, Regression Coefficients, and their Standard Errors

This section describes the four SAS macros needed to compute specific statistics with their correct standard errors, along with sample SAS programs to demonstrate their use. Each SAS macro serves a specific analytical purpose. Making use of these macros for analyzing the PIRLS 2006 data is the best way to ensure the analyses will be done properly. Sampling weights must be used to analyze the PIRLS data, and standard errors must be computed using the jackknife repeated replication (JRR) method. Furthermore, all achievement scores are based on sets of plausible values that take into account the measurement error arising from the test design and the IRT scaling methodology. These plausible values will be processed by the macros to produce accurate results.

The sample SAS programs presented in this section all use a SAS data file called ASGALLR2 as input, which contains the student background data files of all participating countries. In all sample programs, <datpath> must be edited to specify the folder where the ASGALLR2 file is located.

Computing Means and Their Standard Errors (JACKGEN)

The JACKGEN macro is used to compute percentages and means of continuous variables with their JRR standard errors. We will demonstrate its use with a sample SAS program that calls the macro JACKGEN to compute the percentages of students within specified subgroups and their mean on a variable of choice. The macro also computes the appropriate standard errors for the percentages and means. However, this macro is not appropriate for analyzing achievement means based on plausible values; the JACKPV macro should be used for this purpose.

The JACKGEN macro is a self-contained program, located in the program file JACKGEN.SAS, and should not be modified. It essentially computes sets of replicate weights using the sampling and weighting variables, aggregates the data by subgroups using the replicate weights, and then computes and stores the desired statistics in a SAS working file called FINAL.

The macro JACKGEN is included in a SAS program file by issuing the following command:

```
%INCLUDE "<macpath>JACKGEN.SAS" ;
```

where <macpath> points to the specific FOLDER where the SAS macro program JACKGEN.SAS is located. The macro requires that several parameters be specified as input when it is invoked. These parameters are:

WGT	The sampling weight to be used in the analysis. Generally, TOTWGT should be used. TCHWGT should be used when analyzing teacher data from the AST and ATG files.
JKZ	The variable that captures the assignment of cases to sampling zones. The name of this variable in all PIRLS data files is JKZONE.
JKR	The variable that captures whether the case is to be dropped or have its weight doubled for each set of replicate weights. The name of this variable in all PIRLS data files is JKREP.
NJKZ	The number of replicate weights to be generated when computing the JRR standard errors. The value of NJKZ should be set to 75, the maximum possible value across all participating countries.
CVAR	The list of variables that are to be used to define the subgroups. The list can consist of one or more variables. We recommend that you always include IDCNTRY as the first classification variable.
DVAR	The variable for which means are to be computed. Only one variable can be listed and it should be a continuous variable. Plausible values of achievement scores should not be specified here.
INFILE	The name of the data file that contains the data being analyzed. If the path location is included as part of the file name, the name of the file must be enclosed in quotes. It is important to emphasize that this data file must include only those cases that are of interest in the analysis. If you want to have specific cases excluded from the analysis, for example students with missing data, this should be done prior to invoking the macro.

The JACKGEN macro is invoked by a SAS program using the conventional SAS notation for invoking macros. This involves listing the macro name followed by the list of parameters in parenthesis, each separated by a comma. For example, the macro invoked using the following statement:

```
%JACKGEN (TOTWGT, JKZONE, JKREP, 75, IDCNTRY ITSEX, ASDAGE, ASGALLR2) ;
```

will compute the mean age (ASDAGE) for boys and girls (ITSEX) and their standard errors within each country (IDCNTRY), using the weighting variable TOTWGT. It will also compute the percentages of boys and girls and their standard errors within each country. The data will be read from the data file ASGALLR2 and the standard errors will be computed based on 75 replicate weights.

The results of the JACKGEN macro are stored in a SAS working file called FINAL which is stored in the default folder being used by SAS. The following variables are contained in this results file:

Classification Variables

All classification variables are kept in the results file. In our example, there are two classification variables: IDCNTRY and ITSEX. There is one record in the results file for each subgroup defined by the categories of the classification variables.

N

This variable contains the number of valid cases for each subgroup defined by the classification variables. In our example, it is the number of boys and girls with valid data in each country's sample.

Weight Variable

The weight variable contains the sum of weights within each subgroup defined by the classification variables. In our example, this variable is called TOTWGT since we specified TOTWGT as the weighting variable. When TOTWGT is used (or TCHWGT when analyzing teacher data), this variable will be an estimate of the total population within each subgroup.

MNX

This variable contains the estimated means of the specified analysis variable by subgroup.

MNX_SE

This variable contains the JRR standard errors of the estimated means by subgroup.

PCT

This variable contains the estimated percentages of students in each subgroup for the last classification variable listed. In our example it is the percentage of boys and girls within each country.

PCT_SE

This variable contains the JRR standard errors of the estimated percentages.

The contents of the FINAL file can be printed using the SAS PRINT procedure. The sample SAS program that invokes the JACKGEN macro and a printout of the results are presented in Exhibit 5.4. This program is available on the database DVD in the file called SAMPLEJACKGEN.SAS. It produces the mean ages for boys and girls in all countries, although the exhibit shows the results only for the first five countries.

Exhibit 5.4 Sample SAS Program Invoking the SAS Macro JACKGEN and Output File

```
LIBNAME AR2 "<datpath>" ;

%INCLUDE "<macpath>JACKGEN.SAS" ;

DATA ASGALLR2 ;
     SET AR2.ASGALLR2 ;

     WHERE NMISS (ITSEX, ASDAGE) = 0 ;

PROC FORMAT LIBRARY = WORK ;

 VALUE COUNTRY
   < list country formats >

 VALUE SEX
     1 = 'GIRL'
     2 = 'BOY' ;

%JACKGEN (TOTWGT, JKZONE, JKREP, 75, IDCNTRY ITSEX, ASDAGE, ASGALLR2) ;

PROC PRINT DATA = FINAL NOOBS ;
     VAR IDCNTRY ITSEX N TOTWGT MNX MNX_SE PCT PCT_SE ;
     FORMAT IDCNTRY COUNTRY. ITSEX SEX. N 6.0 TOTWGT 10.0
            MNX MNX_SE PCT PCT_SE 6.2 ;
```

IDCNTRY	ITSEX	N	TOTWGT	MNX	MNX_SE	PCT	PCT_SE
AUSTRIA	GIRL	2501	41134	10.30	0.01	49.46	0.70
AUSTRIA	BOY	2566	42035	10.37	0.01	50.54	0.70
BULGARIA	GIRL	1906	31226	10.87	0.02	49.42	1.01
BULGARIA	BOY	1951	31957	10.88	0.02	50.58	1.01
CHINESE TAIPEI	GIRL	2187	145288	10.09	0.01	47.72	0.51
CHINESE TAIPEI	BOY	2402	159201	10.10	0.01	52.28	0.51
DENMARK	GIRL	2054	32633	10.85	0.01	51.61	0.91
DENMARK	BOY	1947	30598	10.97	0.01	48.39	0.91
FRANCE	GIRL	2121	358440	10.00	0.02	48.47	0.73
FRANCE	BOY	2281	381144	10.04	0.02	51.53	0.73

From the first two lines of the results shown in Exhibit 5.4, we see that in Austria there are 2,501 girls in the sample representing 41,134 girls in the whole population. The mean age for girls at the fourth grade in Austria is estimated to be 10.30 with a standard error of 0.01. Girls made up 49.46 percent of Austria's fourth grade student population. Conversely, Austria sampled 2,566 boys representing 42,035 boys in the whole population. The estimated mean age for boys at the fourth grade in Austria is 10.37 with a standard error of 0.01. Boys made up 50.54 percent of Austria's fourth grade student population.

Computing Achievement Means and Their Standard Errors (JACKPV)

The JACKPV macro computes percentages and mean achievement scores based on plausible values with their JRR standard errors. A second sample program demonstrates the use of the JACKPV macro, which computes the percentages of students within specified subgroups and their mean achievement scores. The SAS macro also computes the appropriate standard errors for those percentages and achievement means.

The JACKPV macro is a self-contained program, located in the program file JACKPV.SAS, and should not be modified. It essentially computes sets of replicate weights using the sampling and weighting variables, aggregates the data by subgroups using the replicate weights, and then computes and stores the desired statistics in a SAS working file called FINAL. The macro aggregates data across all plausible values to obtain the correct results.

The SAS macro JACKPV is included in a SAS program by issuing the following command:

```
%INCLUDE "<macpath>JACKPV.SAS" ;
```

where <macpath> points to the specific FOLDER where the SAS macro program JACKPV.SAS is located. The macro requires that several parameters be specified as input when it is invoked. These parameters are:

WGT	The sampling weight to be used in the analysis. Generally, TOTWGT should be used. TCHWGT should be used when analyzing teacher data from the AST and ATG files.
JKZ	The variable that captures the assignment of cases to sampling zones. The name of this variable in all PIRLS data files is JKZONE.
JKR	The variable that captures whether the case is to be dropped or have its weight doubled for each set of replicate weights. The name of this variable in all PIRLS data files is JKREP.
NJKZ	The number of replicate weights to be generated when computing the JRR standard errors. The value of NJKZ should be set to 75, the maximum possible value across all participating countries.

CVAR — The list of variables that are to be used to define the subgroups. The list can consist of one or more variables. We recommend that you always include IDCNTRY as the first classification variable.

ROOTPV — The variable root used to identify the set of plausible values for the achievement score of interest. It corresponds to the first 7 characters of the plausible values variable name. For example, the root of the overall reading plausible values is ASRREA0, the root of the literary purpose plausible values is ASRLIT0.

NPV — The number of plausible values that will be used for the analysis. Generally, you will want to use all five plausible values for analysis.

INFILE — The name of the data file that contains the data being analyzed. If the path location is included as part of the file name, the name of the file must be enclosed in quotes. It is important to emphasize that this data file must include only those cases that are of interest in the analysis. If you want to have specific cases excluded from the analysis, for example students with missing data, this should be done prior to invoking the macro.

The JACKPV macro is invoked by a SAS program using the conventional SAS notation for invoking macros. This involves listing the macro name followed by the list of parameters in parenthesis, each separated by a comma. For example, the macro invoked using the following statement:

```
%JACKPV (TOTWGT, JKZONE, JKREP, 75, IDCNTRY ITSEX, ASRREA0, 5, ASGALLR2) ;
```

will compute the mean reading achievement (ASRREA01 through ASRREA05) for boys and girls (ITSEX) within each country (IDCNTRY) and their standard errors, using the weighting variable TOTWGT. The macro uses all five plausible values to compute these statistics. It will also compute the percentages of boys and girls within each country, and their standard errors. The data will be read from the data file ASGALLR2 and the standard errors will be computed based on 75 replicate weights.

The results of the JACKPV macro are stored in a SAS working file called FINAL which is stored in the default folder being used by SAS. The following variables are contained in this results file:

Classification Variables

All classification variables are kept in the results file. In our example, there are two classification variables: IDCNTRY and ITSEX. There is one record in the results file for each subgroup defined by the categories of the classification variables.

N

This variable contains the number of valid cases for each subgroup defined by the classification variables. In our example, it is the number of boys and girls with valid data in each country's sample.

Weight Variable

The weight variable contains the sum of weights within each subgroup defined by the classification variables. In our example, this variable is called TOTWGT since we specified TOTWGT as the weighting variable. When TOTWGT is used (also TCHWGT when analyzing teacher data), this variable will be an estimate of the total population within each subgroup.

MNPV

This variable contains the estimated mean achievement by subgroup, based on the plausible values.

MNPV_SE

This variable contains the JRR standard errors of the estimated mean achievement by subgroup, based on the plausible values.

PCT

This variable contains the estimated percentages of students in each subgroup for the last classification variable listed. In our example it is the percentage of boys and girls within each country.

PCT_SE

This variable contains the JRR standard errors of the estimated percentages.

The contents of the FINAL file can be printed using the SAS PRINT procedure. The sample SAS program that invokes the JACKPV macro and a printout of the results are presented in Exhibit 5.5. This program is available on the database DVD in the file called SAMPLEJACKPV.SAS. It produces the mean reading achievement for boys and girls in all countries, although Exhibit 5.5 gives the results only for the first five countries.

Exhibit 5.5 Sample SAS Program Invoking the SAS Macro JACKPV and Output File

```
LIBNAME AR2 "<datpath>" ;

%INCLUDE "<macpath>JACKPV.SAS" ;

DATA ASGALLR2 ;
     SET AR2.ASGALLR2 ;

     WHERE NMISS (ITSEX) = 0 ;

PROC FORMAT LIBRARY = WORK ;

 VALUE COUNTRY
   < list country formats >

 VALUE SEX
     1 = 'GIRL'
     2 = 'BOY' ;

%JACKPV (TOTWGT, JKZONE, JKREP, 75, IDCNTRY ITSEX, ASRREA0, 5, ASGALLR2) ;

PROC PRINT DATA = FINAL NOOBS ;
     VAR IDCNTRY ITSEX N TOTWGT MNPV MNPV_SE PCT PCT_SE ;
     FORMAT IDCNTRY COUNTRY. ITSEX SEX. N 6.0 TOTWGT 10.0
            MNPV MNPV_SE PCT PCT_SE 6.2 ;
```

IDCNTRY	ITSEX	N	TOTWGT	MNPV	MNPV_SE	PCT	PCT_SE
AUSTRIA	GIRL	2501	41134	543.26	2.32	49.46	0.70
AUSTRIA	BOY	2566	42035	533.44	2.64	50.54	0.70
BULGARIA	GIRL	1909	31329	557.52	4.41	49.44	1.01
BULGARIA	BOY	1954	32042	536.78	5.00	50.56	1.01
CHINESE TAIPEI	GIRL	2187	145288	542.08	2.22	47.72	0.51
CHINESE TAIPEI	BOY	2402	159201	529.25	2.27	52.28	0.51
DENMARK	GIRL	2054	32633	552.94	2.84	51.61	0.91
DENMARK	BOY	1947	30598	539.31	2.72	48.39	0.91
FRANCE	GIRL	2121	358440	527.31	2.42	48.46	0.73
FRANCE	BOY	2282	381280	516.22	2.41	51.54	0.73

From the first two lines of the results presented in Exhibit 5.5, we see that in Austria the mean reading achievement of girls is estimated to be 543.26 with a standard error of 2.32. The mean reading achievement of boys in Austria is estimated to be 533.44 with a standard error of 2.64.

Computing Regression Coefficients and Their Standard Errors (JACKREG)

The JACKREG macro is used to perform a multiple linear regression between a dependent variable and a set of independent variables. A third sample program demonstrates the use of the JACKREG macro, which computes the regression coefficients and their JRR standard errors. This macro is not appropriate for regression analyses using achievement scores and their plausible values as dependent variables. The JACKREGP macro should be used for this purpose.

The JACKREG macro is a self-contained program, located in the program file JACKREG.SAS, and should not be modified. It essentially computes sets of replicate weights using the sampling and weighting variables, performs a multiple linear regression by subgroups and aggregates the data using the replicate weights, and then computes and stores the desired statistics in a SAS working file called REG.

The SAS macro JACKREG is included in a SAS program file by issuing the following command:

```
%INCLUDE "<macpath>JACKREG.SAS" ;
```

where <macpath> points to the specific FOLDER where the SAS macro program JACKREG.SAS is located. The macro requires that several parameters be specified as input when it is invoked. These parameters are:

WGT	The sampling weight to be used in the analysis. Generally, TOTWGT should be used. TCHWGT should be used when analyzing teacher data from the AST and ATG files.
JKZ	The variable that captures the assignment of cases to sampling zones. The name of this variable in all PIRLS data files is JKZONE.
JKR	The variable that captures whether the case is to be dropped or have its weight doubled for each set of replicate weights. The name of this variable in all PIRLS data files is JKREP.

NJKZ — The number of replicate weights to be generated when computing the JRR standard errors. The value of NJKZ should be set to 75, the maximum possible value across all participating countries.

CVAR — The list of variables that are to be used to define the subgroups. The list can consist of one or more variables. We recommend that you always include IDCNTRY as the first classification variable.

XVAR — The list of independent variables used as predictors in the regression model. The independent variables can be either continuous or categorical, such as ITSEX for example.

DVAR — The dependent variable to be predicted by the list of independent variables specified in XVAR. Only one variable can be listed and plausible values of achievement scores should not be specified here.

INFILE — The name of the data file that contains the data being analyzed. If the path location is included as part of the file name, the name of the file must be enclosed in quotes. It is important to emphasize that this data file must include only those cases that are of interest in the analysis. If you want to have specific cases excluded from the analysis, for example students with missing data, this should be done prior to invoking the macro.

The JACKREG macro is invoked by a SAS program using the conventional SAS notation for invoking macros. This involves listing the macro name followed by the list of parameters in parenthesis, each separated by a comma. For example, the macro invoked using the following statement:

```
%JACKREG (TOTWGT, JKZONE, JKREP, 75, IDCNTRY, REGSEX, BOOK, ASGALLR2) ;
```

will perform a linear regression with the variable REGSEX as a predictor of the number of books in the home (BOOK), using the weighting variable TOTWGT. It will compute the regression coefficients and their standard errors. The data will be read from the data file ASGALLR2 and the standard errors will be computed based on 75 replicate weights.

The results of the JACKREG macro are stored in a SAS working file called REG which is stored in the default folder being used by SAS. The following variables are contained in this results file:

Classification Variables

All classification variables are kept in the results file. In our example, there is a single classification variable IDCNTRY. There is one record in the results file for each subgroup defined by the categories of the classification variables.

N

This variable contains the number of valid cases for each subgroup defined by the classification variables. In our example, it is the number of students with valid data in each country's sample.

MULT_RSQ

The squared multiple correlation (R^2) coefficient for the regression model applied in each subgroup.

SS_RES, SS_REG, SS_TOTAL

The residual, regression, and total weighted sums of squares for the regression model applied in each subgroup.

Regression Coefficients and Standard Errors (B## and B##.SE)

The regression coefficients for the predictor variables and the intercept with their respective standard errors. The regression coefficients are numbered sequentially, starting with zero (B00) for the intercept, and based on the order of the predictor variables are specified in the parameter XVAR.

The contents of the REG file can be printed using the SAS PRINT procedure. The sample SAS program that invokes the JACKREG macro and a printout of the results are presented in Exhibit 5.6. This program is available on the database DVD in the file called SAMPLEJACKREG.SAS. It performs a linear regression in each country, with the variable REGSEX as a predictor of the number of books in the home (BOOK). The exhibit displays the results for the first five countries.

The regression performed by the sample program uses variables that are transformed within its code. The dependent variable BOOK is derived from the categorical variable ASBGBOOK, where BOOK is given the midpoint value of the intervals specified by each category of ASBGBOOK. The independent variable REGSEX is a dummy-coded version of ITSEX, such that the value zero represents the girls and the value one represents the boys. By performing this dummy

coding, the intercept (B00) will be the estimated number of books in the homes of girls, whereas the regression coefficient B01 will be the estimated number of additional books in the homes of boys. This will also allow us to perform a t-test to determine if the number of books in the home is significantly different between girls and boys.

Exhibit 5.6 Sample SAS Program Invoking the SAS Macro JACKREG and Output File

```
LIBNAME AR2 "<datpath>" ;

%INCLUDE "<macpath>JACKREG.SAS" ;

DATA ASGALLR2 ;
     SET AR2.ASGALLR2 ;

     WHERE NMISS (ITSEX, ASBGBOOK) = 0 ;

     SELECT (ITSEX) ;
       WHEN (1)  REGSEX = 0 ;  * GIRLS ;
       WHEN (2)  REGSEX = 1 ;  * BOYS ;
       OTHERWISE REGSEX = . ;
     END ;

     SELECT (ASBGBOOK) ;
       WHEN (1)  BOOK =   5 ;  *   5 Books at home ;
       WHEN (2)  BOOK =  18 ;  *  18 Books at home ;
       WHEN (3)  BOOK =  63 ;  *  63 Books at home ;
       WHEN (4)  BOOK = 151 ;  * 151 Books at home ;
       WHEN (5)  BOOK = 251 ;  * 251 Books at home ;
       OTHERWISE BOOK =   . ;
     END ;

PROC FORMAT LIBRARY = WORK ;

 VALUE COUNTRY
   < list country formats >

%JACKREG (TOTWGT, JKZONE, JKREP, 75, IDCNTRY, REGSEX, BOOK, ASGALLR2) ;

PROC PRINT DATA = REG NOOBS ;
     VAR IDCNTRY N MULT_RSQ SS_TOTAL SS_REG B00 B00_SE B01 B01_SE ;
     FORMAT IDCNTRY COUNTRY. N 6.0 MULT_RSQ 5.3 SS_TOTAL SS_REG 10.0
            B00 B00_SE B01 B01_SE 6.2 ;
```

IDCNTRY	N	MULT_RSQ	SS_TOTAL	SS_REG	B00	B00_SE	B01	B01_SE
AUSTRIA	4923	0.001	482333720	363379	82.81	2.46	-4.24	3.04
BULGARIA	3723	0.000	483282536	871	85.55	3.97	-0.24	3.99
CHINESE TAIPEI	4465	0.000	2068367933	453550	83.65	1.96	-2.48	2.69
DENMARK	3917	0.000	418862800	8445	102.15	2.61	0.74	3.44
FRANCE	4142	0.000	5407360619	399107	106.40	2.99	1.52	2.70

From the first line of the results displayed in Exhibit 5.6, we see that in Austria the estimated average number of books in the homes of fourth grade girls (B00) is 82.81, with a standard error of 2.46. The Austrian fourth grade boys have an estimated 4.24 fewer books in the home (B01) than girls. With an estimated standard error of 3.04, this difference is not statistically significant at a 95% confidence level.

Computing Regression Coefficients and Their Standard Errors with Achievement Scores (JACKREGP)

The JACKREGP macro is used to perform a multiple linear regression between a set of plausible values as the dependent variable and a set of independent variables. It computes the regression coefficients and their JRR standard errors, taking into account the plausible values. We present a fourth sample program to demonstrate the use of the JACKREGP macro.

The JACKREGP macro is a self-contained program, located in the program file JACKREGP.SAS, and should not be modified. It essentially computes sets of replicate weights using the sampling and weighting variables, performs a multiple linear regression by subgroups and aggregates the data using the replicate weights, and then computes and stores the desired statistics in a SAS working file called REG.

The SAS macro JACKREGP is included in a SAS program file by issuing the following command:

```
%INCLUDE "<macpath>JACKREGP.SAS" ;
```

where <macpath> points to the specific FOLDER where the SAS macro program JACKREGP.SAS is located. The macro requires that several parameters be specified as input when it is invoked. These parameters are:

WGT	The sampling weight to be used in the analysis. Generally, TOTWGT should be used. TCHWGT should be used when analyzing teacher data from the AST and ATG files.
JKZ	The variable that captures the assignment of cases to sampling zones. The name of this variable in all PIRLS data files is JKZONE.
JKR	The variable that captures whether the case is to be dropped or have its weight doubled for each set of replicate weights. The name of this variable in all PIRLS data files is JKREP.
NJKZ	The number of replicate weights to be generated when computing the JRR standard errors. The value of NJKZ should be set to 75, the maximum possible value across all participating countries.
CVAR	The list of variables that are to be used to define the subgroups. The list can consist of one or more variables. We recommend that you always include IDCNTRY as the first classification variable.

XVAR — The list of independent variables used as predictors in the regression model. The independent variables can be either continuous or categorical, such as ITSEX for example.

ROOTPV — The variable root used to identify the set of plausible values for the achievement score of interest. It corresponds to the first 7 characters of the plausible values variable name. For example, the root of the overall reading plausible values is ASRREA0, the root of the literary purpose plausible values is "ASRLIT0."

NPV — The number of plausible values that will be used for the analysis. Generally, you will want to use all five plausible values for analysis.

INFILE — The name of the data file that contains the data being analyzed. If the path location is included as part of the file name, the name of the file must be enclosed in quotes. It is important to emphasize that this data file must include only those cases that are of interest in the analysis. If you want to have specific cases excluded from the analysis (e.g., students with missing data), this should be done prior to invoking the macro.

The JACKREGP macro is invoked by a SAS program using the conventional SAS notation for invoking macros. This involves listing the macro name followed by the list of parameters in parenthesis, each separated by a comma. For example, the macro invoked using the following statement:

```
%JACKREGP (TOTWGT, JKZONE, JKREP, 75, IDCNTRY, REGSEX, ASRREA0, 5, ASGALLR2) ;
```

will perform a linear regression with the variable REGSEX as a predictor of reading achievement based on its five plausible values (ASRREA01 through ASRREA05), using the weighting variable TOTWGT. It will compute the regression coefficients and their standard errors. The data will be read from the data file ASGALLR2 and the standard errors will be computed based on 75 replicate weights.

The results of the JACKREGP macro are stored in a SAS working file called REG which is stored in the default folder being used by SAS. The following variables are contained in this results file:

Classification Variables

All classification variables are kept in the results file. In our example, there is a single classification variable IDCNTRY. There is one record in the results file for each subgroup defined by the categories of the classification variables.

N

This variable contains the number of valid cases for each subgroup defined by the classification variables. In our example, it is the number of students with valid data in each country's sample.

MULT_RSQ

The squared multiple correlation (R^2) coefficient for the regression model applied in each subgroup.

SS_RES, SS_REG, SS_TOTAL

The residual, regression, and total weighted sums of squares for the regression model applied in each subgroup.

Regression Coefficients and Standard Errors (B## and B##.SE)

The regression coefficients for the predictor variables and the intercept with their respective standard errors. The regression coefficients are numbered sequentially, starting with zero (B00) for the intercept, and based on the order of the predictor variables are specified in the parameter XVAR.

The contents of the REG file can be printed using the SAS PRINT procedure. The sample SAS program invoking the JACKREGP macro and a printout of the results are presented in Exhibit 5.7. This program is available on the database DVD in the file called SAMPLEJACKREGP.SAS. It performs a linear regression in each country, with the variable REGSEX as a predictor of reading achievement. The exhibit displays the results for the first five countries.

The regression performed by our sample program uses the variable REGSEX that was defined in our previous example. By using REGSEX, the intercept (B00) will be the estimated mean reading achievement of girls, whereas the regression coefficient B01 will be the estimated difference in the mean reading achievement of boys. This will allow us to perform a t-test to determine if reading achievement is significantly different between girls and boys.

Exhibit 5.7 Sample SAS Program Invoking the SAS Macro JACKREGP and Output File

```
LIBNAME AR2 "<datpath>" ;

%INCLUDE "<macpath>JACKREGP.SAS" ;

DATA ASGALLR2 ;
     SET AR2.ASGALLR2 ;

     WHERE NMISS (ITSEX) = 0 ;

     SELECT (ITSEX) ;
       WHEN (1)  REGSEX = 0 ;  * GIRLS ;
       WHEN (2)  REGSEX = 1 ;  * BOYS ;
       OTHERWISE REGSEX = . ;
     END ;

PROC FORMAT LIBRARY = WORK ;

 VALUE COUNTRY
   < list country formats >

%JACKREGP (TOTWGT, JKZONE, JKREP, 75, IDCNTRY, REGSEX, ASRREA0, 5, ASGALLR2) ;

PROC PRINT DATA = REG NOOBS ;
     VAR IDCNTRY N MULT_RSQ SS_TOTAL SS_REG B00 B00_SE B01 B01_SE ;
     FORMAT IDCNTRY COUNTRY. N 6.0 MULT_RSQ 5.3 SS_TOTAL SS_REG 10.0
            B00 B00_SE B01 B01_SE 6.2 ;


IDCNTRY             N   MULT_RSQ    SS_TOTAL     SS_REG      B00   B00_SE     B01   B01_SE

AUSTRIA          5067    0.006    337056418    2026407    543.26    2.32    -9.82    2.35
BULGARIA         3863    0.016    433266259    6824206    557.52    4.41   -20.74    3.83
CHINESE TAIPEI   4589    0.010   1252797359   12579053    542.08    2.22   -12.83    1.86
DENMARK          4001    0.010    307300169    2951463    552.94    2.84   -13.63    3.25
FRANCE           4403    0.007   3279772121   22867710    527.31    2.42   -11.09    2.54
```

From the first line of the results shown in Exhibit 5.7, we see that in Austria the estimated average reading achievement of fourth grade girls (B00) is 543.26, with a standard error of 2.32. Note that these are the same results obtained from the JACKPV sample program. The Austrian fourth grade boys have an estimated average reading achievement 9.82 points (B01) lower than girls. With an estimated standard error of 2.35, this difference is indeed statistically significant at a 95% confidence level.

5.7 Performing Analyses with Student-level Variables

Many analyses of the PIRLS data can be undertaken using student-level data. Examples in the previous sections illustrate the functioning of the SAS macros. This section presents examples of actual analyses used to produce the exhibits in the *PIRLS 2006 International Report*, using SAS programs provided on the DVD.

Student-Level Analysis Without Achievement Scores

In our first example, we wish to replicate the analysis of students' reports on the number of hours spent reading stories or articles in books or magazines outside of school. The results, presented in Exhibit 4.5 of the *PIRLS 2006 International Report*, are reproduced here in Exhibit 5.8. This example will focus on the results presented in the fourth data column—the average number of hours spent reading stories or articles in books or magazines outside of school, overall. Since we want to report the average number of hours, which does not require any plausible values, we will use the macro JACKGEN.

We need to undertake a number of steps to replicate the results in this exhibit. After reviewing the *Student Questionnaire* and codebook for the ASG data files, we observe that the variable ASBGTSP4 contains categorical information on the number of hours spent reading stories or articles in books or magazines outside of school (see Supplement 1 for a copy of the questionnaire), and this variable is found in the student background data files. Our next step is to review the documentation of national adaptations to the questionnaires to ensure that there were no deviations listed for this variable (see Supplement 2). If no national adaptations were made, we can continue with our analysis without any modifications.

We then proceed to read from the student background data files our variable of interest (ASBGTSP4), the student sampling weight (TOTWGT), the variables that contain the jackknife replication information (JKZONE and JKREP), and the variable containing the country identification code (IDCNTRY). In this analysis, we will use the data for all available countries. We used the JOIN program, described earlier in this chapter, to join the ASG files for all countries into a single file called ASGALLR2.

Exhibit 5.8 Example Student-Level Analysis Without Achievement Scores Taken from the PIRLS 2006 International Report (Exhibit 4.5)

Exhibit 4.5 **Students Read Stories or Articles Outside of School** — **PIRLS 2006** 4th Grade

Countries	Average Number of Hours on a Typical Day Spent Reading											
	Stories or Articles on the Internet						Stories or Articles in Books or Magazines					
	Overall		Girls		Boys		Overall		Girls		Boys	
Austria	0.8	(0.02)	0.8	(0.03)	0.9	(0.03) ▲	1.6	(0.03)	1.8	(0.04) ▲	1.3	(0.04)
Belgium (Flemish)	0.6	(0.03)	0.5	(0.03)	0.6	(0.04) ▲	0.6	(0.02)	0.7	(0.03) ▲	0.5	(0.02)
Belgium (French)	1.2	(0.04)	1.2	(0.05)	1.3	(0.05)	1.2	(0.03)	1.3	(0.04) ▲	1.1	(0.04)
Bulgaria	1.0	(0.05)	0.9	(0.05)	1.1	(0.06) ▲	1.5	(0.05)	1.6	(0.07) ▲	1.3	(0.05)
Canada, Alberta	0.8	(0.03)	0.8	(0.04)	0.9	(0.04)	1.3	(0.03)	1.4	(0.04) ▲	1.2	(0.04)
Canada, British Columbia	0.8	(0.03)	0.7	(0.04)	0.8	(0.04) ▲	1.3	(0.04)	1.4	(0.05) ▲	1.2	(0.05)
Canada, Nova Scotia	0.9	(0.03)	0.9	(0.04)	1.0	(0.04) ▲	1.3	(0.03)	1.5	(0.04) ▲	1.2	(0.04)
Canada, Ontario	0.9	(0.04)	0.8	(0.05)	0.9	(0.04)	1.4	(0.05)	1.5	(0.06) ▲	1.3	(0.06)
Canada, Quebec	1.0	(0.04)	1.0	(0.04)	1.1	(0.06)	1.3	(0.04)	1.5	(0.06) ▲	1.2	(0.04)
Chinese Taipei	1.0	(0.03)	1.1	(0.04) ▲	0.9	(0.03)	1.2	(0.03)	1.4	(0.04) ▲	1.1	(0.04)
Denmark	0.6	(0.03)	0.5	(0.03)	0.7	(0.04) ▲	1.0	(0.03)	1.1	(0.04) ▲	0.9	(0.04)
England	0.9	(0.03)	0.8	(0.04)	0.9	(0.04)	1.2	(0.03)	1.4	(0.05) ▲	1.1	(0.04)
France	0.9	(0.03)	0.9	(0.04)	0.9	(0.05)	1.2	(0.03)	1.2	(0.04) ▲	1.1	(0.04)
Georgia	0.9	(0.05)	0.8	(0.06)	0.9	(0.06)	1.5	(0.05)	1.6	(0.07) ▲	1.4	(0.06)
Germany	0.6	(0.02)	0.6	(0.03)	0.6	(0.03) ▲	1.5	(0.03)	1.7	(0.04) ▲	1.4	(0.05)
Hong Kong SAR	1.1	(0.03)	1.1	(0.04)	1.0	(0.04)	1.0	(0.03)	1.1	(0.04) ▲	1.0	(0.04)
Hungary	0.7	(0.03)	0.7	(0.04)	0.7	(0.03)	1.3	(0.04)	1.4	(0.06) ▲	1.1	(0.04)
Iceland	0.6	(0.02)	0.6	(0.02)	0.7	(0.03) ▲	0.8	(0.02)	0.9	(0.03) ▲	0.7	(0.03)
Indonesia	1.3	(0.05)	1.2	(0.05)	1.3	(0.06)	1.6	(0.04)	1.7	(0.06)	1.6	(0.05)
Iran, Islamic Rep. of	0.3	(0.03)	0.3	(0.05)	0.4	(0.05)	1.5	(0.05)	1.5	(0.06)	1.4	(0.08)
Israel	1.5	(0.04)	1.4	(0.05)	1.5	(0.05) ▲	1.4	(0.04)	1.6	(0.05) ▲	1.3	(0.04)
Italy	0.7	(0.03)	0.6	(0.04)	0.8	(0.04) ▲	1.3	(0.04)	1.4	(0.05) ▲	1.1	(0.05)
Kuwait	2.1	(0.06)	2.0	(0.07)	2.1	(0.08)	2.1	(0.05)	2.2	(0.07)	2.1	(0.07)
Latvia	1.0	(0.04)	0.9	(0.04)	1.1	(0.05) ▲	1.2	(0.03)	1.4	(0.05) ▲	1.0	(0.04)
Lithuania	0.9	(0.03)	0.9	(0.04)	1.0	(0.03)	1.4	(0.03)	1.7	(0.04) ▲	1.2	(0.04)
Luxembourg	0.5	(0.01)	0.5	(0.02)	0.6	(0.02) ▲	0.9	(0.02)	1.0	(0.02) ▲	0.8	(0.03)
Macedonia, Rep. of	1.8	(0.08)	1.7	(0.09)	1.8	(0.08) ▲	2.6	(0.07)	2.7	(0.08) ▲	2.4	(0.07)
Moldova, Rep. of	1.0	(0.06)	0.9	(0.08)	1.0	(0.06)	1.8	(0.05)	1.9	(0.07) ▲	1.7	(0.06)
Morocco	1.3	(0.08)	1.3	(0.09)	1.4	(0.08)	1.3	(0.07)	1.4	(0.09)	1.3	(0.08)
Netherlands	0.5	(0.02)	0.5	(0.02)	0.5	(0.03)	0.8	(0.02)	0.9	(0.04) ▲	0.6	(0.03)
New Zealand	0.9	(0.03)	0.9	(0.04)	1.0	(0.04) ▲	1.4	(0.04)	1.6	(0.05) ▲	1.3	(0.04)
Norway	0.6	(0.03)	0.5	(0.04)	0.6	(0.03)	0.9	(0.04)	0.9	(0.05) ▲	0.8	(0.05)
Poland	0.9	(0.03)	0.8	(0.03)	1.0	(0.04) ▲	1.5	(0.03)	1.7	(0.04) ▲	1.3	(0.05)
Qatar	2.3	(0.03)	2.2	(0.04)	2.4	(0.04) ▲	2.2	(0.03)	2.3	(0.04) ▲	2.1	(0.04)
Romania	0.9	(0.06)	0.8	(0.06)	1.0	(0.07) ▲	1.6	(0.05)	1.8	(0.07) ▲	1.5	(0.06)
Russian Federation	0.5	(0.03)	0.4	(0.02)	0.6	(0.04) ▲	1.5	(0.04)	1.6	(0.05) ▲	1.3	(0.04)
Scotland	0.9	(0.04)	0.9	(0.05)	0.9	(0.05)	1.2	(0.03)	1.4	(0.05) ▲	1.1	(0.05)
Singapore	1.1	(0.03)	1.1	(0.03)	1.0	(0.04)	1.4	(0.02)	1.6	(0.04) ▲	1.2	(0.03)
Slovak Republic	0.7	(0.03)	0.7	(0.04)	0.8	(0.04) ▲	1.5	(0.04)	1.7	(0.05) ▲	1.3	(0.05)
Slovenia	0.7	(0.03)	0.6	(0.03)	0.8	(0.04) ▲	1.0	(0.02)	1.1	(0.03) ▲	0.9	(0.03)
South Africa	2.1	(0.07)	2.1	(0.07)	2.1	(0.07)	2.7	(0.06)	2.7	(0.06) ▲	2.6	(0.07)
Spain	0.9	(0.03)	0.8	(0.04)	1.0	(0.05) ▲	1.2	(0.03)	1.2	(0.05)	1.2	(0.05)
Sweden	0.5	(0.02)	0.4	(0.02)	0.6	(0.04) ▲	0.7	(0.02)	0.7	(0.03) ▲	0.6	(0.03)
Trinidad and Tobago	1.5	(0.07)	1.5	(0.09)	1.5	(0.07)	1.7	(0.06)	1.9	(0.08) ▲	1.6	(0.07)
United States	1.0	(0.05)	1.1	(0.06)	1.0	(0.05)	1.4	(0.04)	1.6	(0.04) ▲	1.2	(0.06)
International Avg.	1.0	(0.01)	0.9	(0.01)	1.0	(0.01) ▲	1.4	(0.01)	1.5	(0.01) ▲	1.3	(0.01)

▲ Average significantly higher than other gender

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

The SAS program used to perform this first example is presented in Exhibit 5.9 and is included on the DVD under the name EXAMPLE1.SAS. The results obtained from this program are displayed in Exhibit 5.10. Note that one of the steps in this program is to select only those students who have non-missing data

in our variable of interest ASBGTSP4. Furthermore, the variable ASBGTSP4 is recoded to have the mid-points of the categorical intervals as its values.

In general, to perform student-level analyses of this type using the student background data files, you should do the following:

1) Identify the variables of interest in the student background data files and note any specific national adaptations to the variables
2) Retrieve the relevant variables from the student background data files, including analysis and classification variables, identification variables, sampling and weighting variables, and any other variables used in the selection of cases
3) Perform any necessary variable transformations or recodes
4) Use the macros JACKGEN and JACKREG with the appropriate parameters
5) Specify the location of the data files (<datpath>) and the macros (<macpath>)
6) Print out the results file

Exhibit 5.9 Example SAS Program to Perform Student-Level Analysis without Achievement Scores (EXAMPLE1.SAS)

```
LIBNAME AR2 "<datpath>" ;

%INCLUDE "<macpath>JACKGEN.SAS" ;

DATA ASGALLR2 ;
     SET AR2.ASGALLR2 ;

     WHERE NMISS (ASBGTSP4) = 0 ;

     SELECT (ASBGTSP4) ;
       WHEN (1)  ASBGTSP4 = 5.5 ;  * 5 or more hours        ;
       WHEN (2)  ASBGTSP4 = 4   ;  * More than 3 but less than 5 hours ;
       WHEN (3)  ASBGTSP4 = 2   ;  * 1 to 3 hours;
       WHEN (4)  ASBGTSP4 = 0.5 ;  * Less than 1 hour ;
       WHEN (5)  ASBGTSP4 = 0   ;  * No time ;
       OTHERWISE ASBGTSP4 = .   ;
     END ;

PROC FORMAT LIBRARY = WORK ;

 VALUE COUNTRY
   < list country formats >

%JACKGEN (TOTWGT, JKZONE, JKREP, 75, IDCNTRY, ASBGTSP4, ASGALLR2) ;

PROC PRINT DATA = FINAL NOOBS ;
     VAR IDCNTRY N TOTWGT MNX MNX_SE PCT PCT_SE ;
     FORMAT IDCNTRY COUNTRY. N 6.0 TOTWGT 10.0
            MNX MNX_SE PCT PCT_SE 6.2 ;
```

Exhibit 5.10 Output for Example Student-Level Analysis without Achievement Scores (EXAMPLE 1)

IDCNTRY	N	TOTWGT	MNX	MNX_SE	PCT	PCT_SE
AUSTRIA	4967	81526	1.57	0.03	0.50	0.02
BELGIUM FLEMISH	4446	65646	0.63	0.02	0.40	0.01
BELGIUM FRENCH	479	46885	1.23	0.03	0.29	0.01
BULGARIA	3739	61273	1.46	0.05	0.38	0.01
CANADA ALBERTA	4113	35481	1.30	0.03	0.22	0.01
CANADA BRITISH COLUMBIA	4045	41663	1.28	0.04	0.26	0.00
CANADA NOVA SCOTIA	4297	9380	1.32	0.03	0.06	0.00
CANADA ONTARIO	3888	136297	1.41	0.05	0.84	0.03
CANADA QUEBEC	3626	76016	1.35	0.04	0.47	0.01
CHINESE TAIPEI	4462	295892	1.24	0.03	1.82	0.03
DENMARK	3939	62239	1.03	0.03	0.38	0.01
ENGLAND	3952	540722	1.24	0.04	3.33	0.09
FRANCE	4269	718044	1.15	0.03	4.42	0.11
GEORGIA	4115	41524	1.53	0.05	0.26	0.01
GERMANY	7340	733674	1.53	0.03	4.52	0.09
HONG KONG (SAR)	4572	68340	1.02	0.03	0.42	0.01
HUNGARY	4032	103739	1.27	0.04	0.64	0.02
ICELAND	3588	3978	0.82	0.02	0.02	0.00
INDONESIA	4505	4008052	1.65	0.04	24.69	0.60
IRAN	5323	1138032	1.47	0.05	7.01	0.19
ISRAEL	3689	80600	1.45	0.04	0.50	0.01
ITALY	3520	503243	1.25	0.04	3.10	0.06
KUWAIT	3392	23650	2.12	0.05	0.15	0.00
LATVIA	4116	19567	1.17	0.03	0.12	0.00
LITHUANIA	4666	32496	1.43	0.03	0.20	0.00
LUXEMBOURG	5072	5140	0.88	0.02	0.03	0.00
MACEDONIA	3647	20817	2.58	0.07	0.13	0.00
MOLDOVA	3946	42786	1.77	0.05	0.26	0.01
MOROCCO	3120	543315	1.33	0.07	3.35	0.10
NETHERLANDS	4106	174450	0.76	0.02	1.07	0.03
NEW ZEALAND	5927	53772	1.44	0.04	0.33	0.01
NORWAY	3732	59819	0.88	0.04	0.37	0.01
POLAND	4770	389159	1.48	0.03	2.40	0.06
QATAR	6098	6516	2.25	0.03	0.04	0.00
ROMANIA	4144	192349	1.63	0.05	1.18	0.04
RUSSIAN FEDERATION	4673	1212611	1.46	0.04	7.47	0.17
SCOTLAND	3703	56010	1.24	0.03	0.35	0.01
SINGAPORE	6325	48760	1.40	0.02	0.30	0.00
SLOVAKIA	5302	51588	1.54	0.04	0.32	0.01
SLOVENIA	5276	17423	0.98	0.02	0.11	0.00
SOUTH AFRICA	13092	744365	2.67	0.06	4.59	0.07
SPAIN	4026	384242	1.19	0.03	2.37	0.05
SWEDEN	4306	100097	0.68	0.02	0.62	0.02
TRINIDAD AND TOBAGO	3748	16249	1.74	0.06	0.10	0.00
UNITED STATES	4922	3187176	1.40	0.04	19.63	0.45

In this example, each country's mean value for the recoded ASBGTSP4 variable is reported for all sampled students. The countries are identified in the first column and the second column reports the number of valid cases. The third column reports the sum of weights of the sampled students, followed by the mean for ASBGTSP4 and its standard error. The last two columns report the weighted percentage of students in the population and its standard error. For this example, the weighted percentages are of little use as they are the proportion each country represents among all participating countries.

From the first line shown in Exhibit 5.10, we see that in Austria valid data were available for 4,967 students and these sampled students represent a population of 81,526 students. On average, Austrian students spend 1.57 hours reading stories or articles in books or magazines outside of school, with a standard error of 0.03.

Student-Level Analysis with Achievement Scores

Our second example replicates another set of results presented in the *PIRLS 2006 International Report*. We are interested in investigating whether students' reading achievement increases as students' reports of the frequency with which they read outside of school increases. These results, presented in Exhibit 4.6 of the *PIRLS 2006 International Report*, are repeated here in Exhibit 5.11. Since the results in this Exhibit are based on plausible values, we need to use the macro JACKPV.

After reviewing the questionnaire and codebook, we observe that the variable ASBGTOC5, found in the ASG data file, contains categorical information on how often students reported reading for fun outside of school. Our next step is to review the documentation of national adaptations to the questionnaires to ensure that there were no deviations listed for this variable (see Supplement 2). If no national adaptations were made, we can continue with our analysis without any modifications.

We then proceed to read from the ASG data files our variable of interest (ASBGTOC5), the five plausible values of reading achievement (ASRREA01 through ASRREA05), the student sampling weight (TOTWGT), the variables that contain the jackknife replication information (JKZONE and JKREP), and the variable containing the country identification code (IDCNTRY). Again, we will use the data of all available countries contained in the file ASGALLR2.

The SAS program that implements this second example is presented in Exhibit 5.12 and is included on the DVD under the name EXAMPLE2.SAS. The results obtained from this program are displayed in Exhibit 5.13. For the sake of conciseness, only the results of first five and last five countries, as sorted alphabetically, are shown. Note that one of the steps in this program is to select only those students who have non-missing data in our variable of interest ASBGTOC5. A second step consists of combining category 4 of the variable ASBGTOC5 with category 3 to create a category we will label "Twice a Month or Less".

Exhibit 5.11 Example Student-Level Analysis with Achievement Scores Taken From the PIRLS 2006 International Report (Exhibit 4.6)

Exhibit 4.6 **Students Reading for Fun Outside of School with Trends** — **PIRLS 2006** 4th Grade

Countries	Every Day or Almost Every Day			Once or Twice a Week			Twice a Month or Less		
	2006 Percent of Students	Average Achievement	Difference in Percent from 2001	2006 Percent of Students	Average Achievement	Difference in Percent from 2001	2006 Percent of Students	Average Achievement	Difference in Percent from 2001
Russian Federation	58 (1.1)	570 (3.8)	0 (1.8)	28 (0.8)	559 (3.9)	–1 (1.3)	14 (0.8)	556 (3.9)	2 (1.3)
Canada, Alberta	53 (0.9)	575 (2.5)	◊ ◊	23 (0.8)	555 (3.5)	◊ ◊	23 (0.9)	537 (2.7)	◊ ◊
Canada, British Columbia	53 (1.0)	573 (2.9)	◊ ◊	26 (0.8)	554 (2.5)	◊ ◊	21 (0.9)	531 (4.1)	◊ ◊
Germany	53 (0.9)	563 (2.7)	5 (1.2) ▲	24 (0.6)	545 (3.0)	0 (0.9)	24 (0.8)	525 (2.5)	–5 (1.2) ▼
Lithuania	52 (1.2)	545 (2.1)	–1 (1.9)	30 (1.0)	533 (2.2)	–1 (1.6)	17 (0.8)	520 (2.8)	2 (1.2)
Moldova, Rep. of	52 (1.4)	507 (3.2)	2 (2.6)	34 (1.1)	498 (4.2)	0 (1.9)	14 (1.0)	484 (5.1)	–2 (1.6)
France	51 (1.0)	540 (2.5)	2 (1.6)	24 (0.8)	517 (2.3)	–2 (1.2)	25 (0.9)	491 (2.7)	0 (1.3)
Canada, Ontario	49 (1.4)	567 (3.2)	14 (2.0) ▲	25 (1.1)	552 (3.6)	2 (1.4)	26 (1.1)	534 (4.1)	–16 (1.9) ▼
Belgium (French)	49 (1.1)	517 (3.0)	◊ ◊	26 (0.7)	495 (2.9)	◊ ◊	25 (0.9)	473 (3.1)	◊ ◊
Iceland	49 (0.9)	527 (1.9)	–3 (1.2) ▼	23 (0.7)	511 (2.6)	2 (1.0)	28 (0.7)	485 (2.3)	1 (1.0)
Denmark	49 (1.1)	559 (2.9)	◊ ◊	30 (0.8)	540 (2.7)	◊ ◊	21 (0.9)	528 (3.2)	◊ ◊
Canada, Nova Scotia	48 (0.9)	560 (2.6)	◊ ◊	25 (0.8)	541 (2.7)	◊ ◊	27 (0.8)	515 (3.2)	◊ ◊
Bulgaria	47 (1.6)	561 (4.4)	–4 (2.3)	27 (1.0)	555 (5.0)	–2 (1.5)	26 (1.6)	520 (6.5)	5 (2.2) ▲
Canada, Quebec	47 (1.3)	549 (3.0)	1 (1.9)	26 (1.0)	530 (3.8)	1 (1.4)	27 (1.2)	509 (3.2)	–3 (1.7)
South Africa	45 (0.9)	303 (6.4)	◊ ◊	26 (0.5)	314 (6.4)	◊ ◊	28 (0.8)	307 (6.4)	◊ ◊
Austria	45 (1.1)	555 (3.0)	◊ ◊	25 (0.8)	535 (2.4)	◊ ◊	29 (1.0)	516 (2.9)	◊ ◊
Spain	45 (1.1)	525 (2.9)	◊ ◊	27 (0.7)	515 (2.5)	◊ ◊	28 (1.0)	494 (3.7)	◊ ◊
Hungary	44 (1.2)	565 (3.7)	4 (1.7) ▲	30 (0.9)	547 (3.4)	–2 (1.4)	26 (1.1)	532 (4.2)	–2 (1.5)
Indonesia	44 (1.4)	405 (4.7)	◊ ◊	31 (1.1)	414 (4.6)	◊ ◊	25 (1.2)	403 (4.8)	◊ ◊
Macedonia, Rep. of	43 (1.2)	453 (5.7)	–3 (1.9)	31 (1.0)	451 (4.6)	3 (1.4) ▲	25 (1.1)	435 (5.1)	0 (1.8)
Poland	43 (1.3)	538 (2.5)	◊ ◊	29 (1.0)	518 (3.2)	◊ ◊	27 (1.0)	495 (3.0)	◊ ◊
New Zealand	42 (1.1)	562 (2.4)	–1 (1.8)	24 (0.7)	531 (2.5)	0 (1.1)	34 (1.0)	500 (3.0)	1 (1.6)
Belgium (Flemish)	40 (1.1)	563 (2.1)	◊ ◊	29 (0.8)	545 (2.9)	◊ ◊	31 (1.2)	529 (2.3)	◊ ◊
Slovak Republic	39 (1.0)	545 (2.9)	0 (1.5)	33 (0.9)	535 (3.2)	0 (1.3)	27 (1.1)	507 (5.4)	0 (1.6)
Italy	38 (1.3)	573 (3.3)	7 (1.7) ▲	25 (0.7)	554 (3.2)	1 (1.0)	37 (1.3)	529 (3.8)	–7 (1.7) ▼
Luxembourg	38 (0.6)	581 (1.8)	◊ ◊	27 (0.7)	551 (2.0)	◊ ◊	35 (0.5)	537 (1.5)	◊ ◊
Israel	38 (1.2)	538 (4.2)	–6 (1.6) ▼	28 (0.9)	518 (4.4)	1 (1.2)	35 (1.1)	497 (4.1)	5 (1.5) ▲
Slovenia	37 (0.9)	543 (2.5)	–8 (1.6) ▼	33 (0.7)	519 (3.0)	4 (1.1) ▲	30 (0.9)	500 (2.6)	4 (1.5) ▲
Netherlands	36 (1.1)	566 (2.1)	0 (1.6)	22 (0.7)	550 (1.8)	2 (1.1) ▲	42 (1.1)	530 (1.8)	–2 (1.6)
Sweden	36 (1.0)	569 (2.8)	–8 (1.3) ▼	31 (0.9)	549 (3.2)	–1 (1.2)	33 (1.0)	530 (2.6)	9 (1.3) ▲
United States	35 (1.3)	561 (4.3)	1 (1.8)	22 (0.7)	550 (3.3)	0 (1.3)	43 (1.4)	521 (3.3)	–1 (1.8)
Latvia	35 (1.2)	556 (3.0)	–8 (1.7) ▼	31 (0.8)	543 (2.8)	–2 (1.2)	34 (1.2)	524 (2.6)	10 (1.7) ▲
Hong Kong SAR	35 (1.0)	575 (2.6)	14 (1.3) ▲	33 (0.9)	567 (2.7)	–5 (1.2) ▼	32 (1.0)	549 (2.8)	–8 (1.4) ▼
Iran, Islamic Rep. of	33 (1.2)	428 (4.2)	–1 (1.7)	41 (1.2)	429 (3.9)	2 (1.9)	26 (1.0)	406 (5.2)	0 (1.5)
Norway	33 (1.1)	514 (3.4)	–5 (1.5) ▼	30 (1.0)	505 (3.2)	2 (1.3)	37 (1.2)	481 (3.1)	4 (1.7) ▲
Qatar	33 (0.6)	357 (2.2)	◊ ◊	28 (0.5)	367 (2.5)	◊ ◊	39 (0.6)	352 (2.2)	◊ ◊
Scotland	33 (1.1)	555 (4.4)	2 (1.6)	24 (1.0)	533 (3.1)	0 (1.3)	44 (1.5)	505 (2.7)	–2 (2.2)
England	33 (1.2)	575 (4.0)	0 (1.8)	25 (0.8)	537 (3.5)	–1 (1.2)	42 (1.3)	517 (2.9)	1 (2.0)
Kuwait	32 (1.1)	338 (5.5)	◊ ◊	32 (1.0)	342 (5.5)	◊ ◊	36 (1.2)	332 (5.4)	◊ ◊
Trinidad and Tobago	32 (1.2)	450 (6.7)	◊ ◊	25 (1.0)	442 (5.9)	◊ ◊	43 (1.4)	427 (5.6)	◊ ◊
Georgia	29 (1.4)	479 (4.5)	◊ ◊	29 (1.2)	484 (4.0)	◊ ◊	41 (1.4)	461 (4.3)	◊ ◊
Morocco	29 (1.3)	317 (8.2)	–3 (2.3)	34 (1.4)	326 (6.9)	–3 (2.0)	37 (1.9)	331 (7.1)	6 (3.1)
Singapore	27 (0.9)	587 (3.9)	–3 (1.5) ▼	26 (0.6)	564 (3.1)	3 (0.9) ▲	47 (1.0)	540 (2.7)	0 (1.5)
Romania	25 (1.3)	510 (5.5)	–3 (2.2)	26 (1.1)	502 (5.5)	–4 (1.6) ▼	50 (1.6)	478 (5.5)	7 (2.3) ▲
Chinese Taipei	24 (0.7)	553 (2.6)	◊ ◊	31 (0.8)	539 (2.6)	◊ ◊	45 (1.0)	525 (2.2)	◊ ◊
International Avg.	40 (0.2)	516 (0.6)		28 (0.1)	503 (0.6)		32 (0.2)	484 (0.6)	

▲ Percent in 2006 significantly higher

▼ Percent in 2006 significantly lower

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

In general, to perform student-level analyses of this type using the student background data files, you should do the following:

1) Identify the variables of interest in the student background data files and note any specific national adaptations to the variables
2) Retrieve the relevant variables from the student background data files, including plausible values for achievement, classification variables, identification variables, sampling and weighting variables, and any other variables used in the selection of cases
3) Perform any necessary variable transformations or recodes
4) Use the macros JACKPV and JACKREGP with the appropriate parameters
5) Specify the location of the data files (<datpath>) and the macros (<macpath>)
6) Print out the results file

Exhibit 5.12 Example SAS Program to Perform Student-Level Analysis with Achievement Scores (EXAMPLE2.SAS)

```
LIBNAME AR2 "<datpath>" ;

%INCLUDE "<macpath>JACKPV.SAS" ;

DATA ASGALLR2 ;
     SET AR2.ASGALLR2 ;

     WHERE NMISS (ASBGTOC5) = 0 ;

     SELECT (ASBGTOC5) ;
       WHEN (1)   ASBGTOC5 = 1 ;  * EVERY DAY OR ALMOST EVERY DAY ;
       WHEN (2)   ASBGTOC5 = 2 ;  * ONCE OR TWICE A WEEK ;
       WHEN (3,4) ASBGTOC5 = 3 ;  * TWICE A MONTH OR LESS ;
       OTHERWISE  ASBGTOC5 = . ;
     END ;

PROC FORMAT LIBRARY = WORK ;

 VALUE COUNTRY
   < list country formats >

VALUE TOC
     1 = 'EVERY DAY OR ALMOST EVERY DAY'
     2 = 'ONCE OR TWICE A WEEK'
     3 = 'TWICE A MONTH OR LESS' ;

%JACKPV (TOTWGT, JKZONE, JKREP, 75, IDCNTRY ASBGTOC5, ASRREA0, 5, ASGALLR2) ;

PROC PRINT DATA = FINAL NOOBS ;
     VAR IDCNTRY ASBGTOC5 N TOTWGT MNPV MNPV_SE PCT PCT_SE ;
     FORMAT IDCNTRY COUNTRY. ASBGTOC5 TOC. N 6.0 TOTWGT 10.0
            MNPV MNPV_SE PCT PCT_SE 6.2 ;
```

Exhibit 5.13 Output for Example Student-Level Analysis with Achievement Scores (EXAMPLE 2)

IDCNTRY	ASBGTOC5	N	TOTWGT	MNPV	MNPV_SE	PCT	PCT_SE
AUSTRIA	EVERY DAY OR ALMOST EVERY DAY	2281	36917	555.44	3.00	45.16	1.10
AUSTRIA	ONCE OR TWICE A WEEK	1255	20821	535.13	2.43	25.47	0.80
AUSTRIA	TWICE A MONTH OR LESS	1445	24016	515.74	2.88	29.38	0.98
BELGIUM FLEMISH	EVERY DAY OR ALMOST EVERY DAY	1776	25976	562.65	2.11	39.61	1.06
BELGIUM FLEMISH	ONCE OR TWICE A WEEK	1281	19015	545.24	2.89	28.99	0.80
BELGIUM FLEMISH	TWICE A MONTH OR LESS	1386	20591	529.46	2.33	31.40	1.16
BELGIUM FRENCH	EVERY DAY OR ALMOST EVERY DAY	2207	22935	516.84	3.00	49.09	1.10
BELGIUM FRENCH	ONCE OR TWICE A WEEK	1146	12063	494.61	2.93	25.82	0.66
BELGIUM FRENCH	TWICE A MONTH OR LESS	1106	11726	472.75	3.14	25.10	0.94
BULGARIA	EVERY DAY OR ALMOST EVERY DAY	1796	29070	560.61	4.37	47.24	1.59
BULGARIA	ONCE OR TWICE A WEEK	1033	16313	555.34	5.00	26.51	1.04
BULGARIA	TWICE A MONTH OR LESS	928	16154	520.18	6.50	26.25	1.56
CANADA ALBERTA	EVERY DAY OR ALMOST EVERY DAY	2200	18971	574.88	2.46	53.44	0.95
CANADA ALBERTA	ONCE OR TWICE A WEEK	971	8276	554.91	3.51	23.31	0.79
CANADA ALBERTA	TWICE A MONTH OR LESS	947	8251	536.85	2.75	23.24	0.86
	. . . . .						
SOUTH AFRICA	EVERY DAY OR ALMOST EVERY DAY	6161	345653	303.08	6.41	45.50	0.91
SOUTH AFRICA	ONCE OR TWICE A WEEK	3443	199247	314.45	6.35	26.23	0.53
SOUTH AFRICA	TWICE A MONTH OR LESS	3783	214845	306.94	6.44	28.28	0.75
SPAIN	EVERY DAY OR ALMOST EVERY DAY	1801	170563	525.03	2.88	44.77	1.15
SPAIN	ONCE OR TWICE A WEEK	1078	102245	514.71	2.45	26.84	0.70
SPAIN	TWICE A MONTH OR LESS	1117	108198	493.87	3.69	28.40	1.01
SWEDEN	EVERY DAY OR ALMOST EVERY DAY	1572	36017	568.61	2.81	35.97	1.02
SWEDEN	ONCE OR TWICE A WEEK	1333	30843	549.35	3.22	30.80	0.86
SWEDEN	TWICE A MONTH OR LESS	1401	33281	530.13	2.56	33.23	0.96
TRINIDAD AND TOBAGO	EVERY DAY OR ALMOST EVERY DAY	1215	5218	449.84	6.74	31.70	1.17
TRINIDAD AND TOBAGO	ONCE OR TWICE A WEEK	956	4170	441.71	5.89	25.33	0.95
TRINIDAD AND TOBAGO	TWICE A MONTH OR LESS	1619	7072	426.74	5.58	42.97	1.43
UNITED STATES	EVERY DAY OR ALMOST EVERY DAY	1717	1131176	560.82	4.26	35.28	1.25
UNITED STATES	ONCE OR TWICE A WEEK	1095	705135	549.57	3.35	21.99	0.74
UNITED STATES	TWICE A MONTH OR LESS	2139	1369921	521.22	3.27	42.73	1.43

In this example, each country's results are presented on three lines, one for each value of the recoded ASBGTOC5 variable. The countries are identified in the first column and the second column presents the label of the category of ASBGTOC5 being reported. The third column reports the number of valid cases, and the fourth the sum of weights of the sampled students. The next two columns report the estimated average reading achievement and its standard error. The last two columns report the weighted percentage of students and its standard error.

From the first three lines in Exhibit 5.13, we see that 45.16 percent of students in Austria reported reading for fun every day or almost every day, 25.47 percent reported reading for fun once or twice a week, and 29.38 percent twice a month or less. We also see that the average reading achievement was 555.44 (standard error of 3.00) for those reporting reading every day or almost every day, 535.13 (standard error of 2.43) for those reading once or twice a week, and 515.74 (standard error of 2.88) for those reading twice a month or less.

5.8 Performing Analyses with Teacher-level Variables

When considering analyses with teacher-level variables, it is important to realize that the teachers do not constitute representative samples of teachers. Rather, they are the teachers of nationally representative samples of students. Therefore, analyses with teacher data should be made with students as units of analysis, in terms of how many students are taught by teachers with a particular attribute, and not in terms of how many teachers in a country have a particular attribute.

When analyzing teacher data, it is first necessary to link the students to their respective teachers. The student-teacher linkage data files were created to facilitate this linkage. Each student record in a student background data file links to at least one reading teacher in its corresponding teacher background data file. There is usually only one student-teacher link, but occasionally two, and in rare circumstances three.

The student-teacher linkage data files contain one record for each student-teacher combination. For example, if a student is taught by two reading teachers, the AST data file has two records for that student, one for each of his/her reading teachers. Each record in the data file holds the number of reading teachers linked to that student.

Student achievement scores (plausible values), jackknife replication information, and a weighting variable, TCHWGT, appropriate for conducting analyses with teacher variables, also are found in the AST data files in order to simplify the merging process for analyses that link teacher variables to student achievement. For such analyses, it is only necessary to merge the teacher background data files with the student-teacher linkage data files. For analyses linking teacher variables to student background variables, it is also necessary to merge the student background data files with the teacher background data files after having been combined with the student-teacher linkage data files.

As our example of an analysis with teacher-level variables, we will investigate the age of the PIRLS reading teachers. The results of such an analysis are presented in Exhibit 6.3 of the *PIRLS 2006 International Report* and are reproduced here in Exhibit 5.14. Although the results in this exhibit do not include any achievement, we will use the macro JACKPV to estimate the percentages we want.

Exhibit 5.14 Example Analysis with Teacher Variable Data Taken from the PIRLS 2006 International Report (Exhibit 6.3)

Exhibit 6.3 **Teachers' Gender, Age, and Number of Years Teaching** — **PIRLS 2006** 4th Grade

Countries	Percentage of Students by Teacher Characteristics						Trends in Number of Years Teaching All Grades		
	Gender		Age						
	Female	Male	29 Years or Under	30-39 Years	40-49 Years	50 Years or Older	2006	Difference from 2001	
Austria	88 (2.1)	12 (2.1)	6 (1.7)	17 (2.5)	39 (3.0)	38 (3.2)	22 (0.7)	◊ ◊	
Belgium (Flemish)	75 (3.0)	25 (3.0)	28 (3.2)	29 (3.3)	29 (2.9)	14 (2.2)	16 (0.6)	◊ ◊	
Belgium (French)	80 (3.0)	20 (3.0)	16 (2.2)	33 (3.2)	37 (3.5)	14 (2.6)	17 (0.6)	◊ ◊	
Bulgaria	94 (1.8)	6 (1.8)	3 (1.4)	26 (2.8)	40 (4.0)	31 (3.9)	21 (0.6)	4 (0.9)	▲
Canada, Alberta	80 (3.2)	20 (3.2)	22 (3.7)	20 (3.3)	26 (3.6)	32 (3.3)	15 (0.8)	◊ ◊	
Canada, British Columbia	r 72 (3.6)	28 (3.6)	r 9 (2.2)	21 (3.0)	32 (4.1)	38 (4.2)	17 (0.8)	◊ ◊	
Canada, Nova Scotia	84 (2.9)	16 (2.9)	12 (2.4)	22 (3.4)	25 (3.0)	40 (3.4)	18 (0.7)	◊ ◊	
Canada, Ontario	75 (4.8)	25 (4.8)	18 (3.9)	42 (5.2)	21 (4.3)	20 (4.3)	12 (0.9)	-4 (1.3)	▼
Canada, Quebec	86 (3.2)	14 (3.2)	12 (2.3)	37 (4.0)	19 (3.5)	32 (3.7)	17 (0.8)	-1 (1.2)	
Chinese Taipei	83 (3.2)	17 (3.2)	24 (3.7)	44 (4.1)	27 (3.5)	5 (1.6)	12 (0.6)	◊ ◊	
Denmark	90 (2.1)	10 (2.1)	16 (2.9)	24 (3.4)	19 (2.9)	41 (4.3)	16 (1.1)	◊ ◊	
England	75 (3.5)	25 (3.5)	30 (3.8)	33 (4.1)	14 (2.9)	23 (3.7)	12 (0.9)	-2 (1.3)	
France	71 (3.3)	29 (3.3)	17 (2.9)	31 (3.0)	34 (3.4)	18 (2.6)	15 (0.7)	-3 (1.1)	▼
Georgia	100 (0.3)	0 (0.0)	9 (1.6)	28 (3.4)	24 (3.4)	39 (3.6)	20 (0.9)	◊ ◊	
Germany	89 (2.4)	11 (2.4)	5 (1.7)	23 (2.9)	22 (3.0)	49 (3.8)	20 (0.9)	-3 (1.1)	▼
Hong Kong SAR	78 (3.7)	22 (3.7)	29 (4.2)	33 (4.1)	19 (3.5)	19 (2.6)	13 (0.7)	0 (1.2)	
Hungary	97 (1.6)	3 (1.6)	6 (2.1)	21 (3.1)	48 (3.7)	24 (3.5)	21 (0.7)	3 (1.0)	▲
Iceland	93 (0.2)	7 (0.2)	13 (0.3)	31 (0.4)	35 (0.4)	22 (0.3)	12 (0.1)	-1 (0.1)	▼
Indonesia	56 (3.9)	44 (3.9)	13 (2.2)	31 (4.1)	39 (3.9)	17 (3.0)	16 (0.8)	◊ ◊	
Iran, Islamic Rep. of	50 (2.3)	50 (2.3)	8 (1.6)	44 (3.6)	41 (3.5)	7 (1.9)	17 (0.5)	3 (0.8)	▲
Israel	92 (1.9)	8 (1.9)	8 (1.8)	41 (4.0)	33 (4.1)	18 (3.4)	16 (0.8)	2 (1.1)	
Italy	98 (1.1)	2 (1.1)	2 (1.3)	14 (2.5)	37 (3.5)	47 (3.6)	22 (0.7)	1 (1.0)	
Kuwait	86 (2.3)	14 (2.3)	36 (3.8)	48 (3.8)	15 (3.1)	1 (0.0)	8 (0.6)	◊ ◊	
Latvia	99 (0.6)	1 (0.0)	7 (2.0)	30 (3.6)	34 (3.6)	29 (3.4)	21 (0.8)	2 (1.3)	
Lithuania	99 (0.5)	1 (0.5)	1 (0.7)	35 (3.4)	40 (3.2)	23 (3.2)	21 (0.6)	1 (1.1)	
Luxembourg	55 (0.2)	45 (0.2)	32 (0.2)	24 (0.1)	16 (0.1)	27 (0.1)	15 (0.0)	◊ ◊	
Macedonia, Rep. of	70 (3.5)	30 (3.5)	5 (1.7)	24 (3.4)	40 (4.1)	31 (4.2)	r 20 (1.0)	0 (1.5)	
Moldova, Rep. of	90 (2.8)	10 (2.8)	8 (2.3)	23 (3.7)	36 (4.3)	33 (4.2)	24 (0.9)	5 (1.2)	▲
Morocco	56 (4.0)	44 (4.0)	18 (3.0)	25 (3.5)	40 (3.4)	17 (2.9)	17 (0.6)	1 (1.0)	
Netherlands	68 (3.4)	32 (3.4)	26 (3.5)	19 (3.0)	19 (3.5)	36 (4.0)	17 (1.0)	0 (1.4)	
New Zealand	77 (2.7)	23 (2.7)	22 (2.1)	28 (2.6)	26 (2.4)	25 (2.4)	12 (0.6)	-1 (1.1)	
Norway	91 (1.9)	9 (1.9)	9 (2.7)	28 (3.3)	29 (3.9)	35 (3.9)	16 (1.0)	-1 (1.3)	
Poland	100 (0.3)	0 (0.3)	2 (1.1)	31 (3.4)	58 (4.0)	8 (2.3)	20 (0.5)	◊ ◊	
Qatar	r 90 (0.1)	10 (0.1)	r 27 (0.2)	54 (0.3)	16 (0.2)	3 (0.1)	11 (0.0)	◊ ◊	
Romania	89 (2.4)	11 (2.4)	18 (3.1)	25 (3.3)	27 (3.5)	30 (3.2)	22 (0.9)	2 (1.2)	
Russian Federation	98 (1.1)	2 (1.1)	6 (1.6)	38 (3.7)	35 (3.5)	21 (2.8)	22 (0.6)	2 (1.0)	▲
Scotland	96 (1.6)	4 (1.6)	26 (3.8)	19 (2.8)	18 (3.5)	37 (4.2)	16 (1.1)	-2 (1.5)	
Singapore	75 (2.2)	25 (2.2)	37 (2.6)	42 (3.0)	16 (2.4)	6 (1.0)	9 (0.4)	-2 (1.0)	▼
Slovak Republic	93 (1.7)	7 (1.7)	13 (2.4)	37 (3.5)	25 (2.7)	25 (3.2)	17 (0.8)	0 (1.2)	
Slovenia	98 (1.1)	2 (1.1)	12 (2.2)	24 (2.9)	44 (3.0)	19 (2.4)	19 (0.7)	0 (1.0)	
South Africa	71 (2.5)	29 (2.5)	4 (1.2)	44 (2.8)	32 (2.9)	20 (2.3)	15 (0.4)	◊ ◊	
Spain	78 (3.6)	22 (3.6)	11 (2.4)	16 (2.6)	25 (3.4)	49 (3.8)	22 (0.9)	◊ ◊	
Sweden	84 (2.8)	16 (2.8)	9 (1.7)	29 (3.3)	24 (2.9)	38 (3.6)	17 (1.0)	1 (1.2)	
Trinidad and Tobago	82 (3.0)	18 (3.0)	11 (2.5)	37 (3.6)	28 (3.3)	24 (3.0)	19 (0.7)	◊ ◊	
United States	85 (2.7)	15 (2.7)	21 (2.8)	27 (2.8)	28 (3.7)	25 (3.4)	12 (0.7)	-3 (1.1)	▼
International Avg.	83 (0.4)	17 (0.4)	15 (0.4)	30 (0.5)	30 (0.5)	25 (0.5)	17 (0.1)	0 (0.0)	

▲ Number of years in 2006 significantly higher ▼ Number of years in 2006 significantly lower

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

Conducting analyses with teacher data requires a few extra steps. As before, we first proceed to identify the variables relevant to the analysis in the appropriate files, and review the documentation for any specific national adaptations to the questions of interest (Supplements 1 and 2). Since we are using a teacher-level variable, we need to use the teacher background data files and the student-teacher

linkage data files to find the variables. From the ATG data files, we will need the variable that contains the information on the reading teachers' age (ATBGAGE), the variable that identifies the country (IDCNTRY), and the two identification variables (IDTEACH and IDLINK) that will allow us to link the teacher data to the student data.

We then proceed to read the necessary information from the student-teacher linkage data files. From these files we need the country identification (IDCNTRY) and the two identification variables (IDTEACH and IDLINK) needed to link the student data to the teacher data. We also need the jackknife replication variables (JKZONE and JKREP), the teacher weighting variable (TCHWGT), and the reading achievement plausible values (ASRREA01 through ASRREA05). Although we are only interested in estimating percentages, the reading achievement plausible values are required input for the JACKPV macro. This may be of analytical interest, providing some insight into the relationship between students' reading achievement and the age of their teachers.

The two file types are merged and the merged file is then input to the JACKPV macro. The merging is done using the combination of identification variables IDCNTRY, IDTEACH, and IDLINK. The combination of values for these three variables is unique within the teacher background data files, but is repeated in the student-teacher linkage data files as many times as needed when a specific student is taught by more than one teacher. After the files are merged, the JACKPV macro is invoked and the results can be printed.

For this analysis, we will again use the data for all available countries, making use of an aggregated teacher background data file, ATGALLR2, and an aggregated student-teacher linkage data file, ASTALLR2. These aggregated files were created using the JOIN macro.

The SAS program that executes this third example is presented in Exhibit 5.15 and is included on the DVD under the name EXAMPLE3.SAS. The results obtained from this program are displayed in Exhibit 5.16, edited to show only the first four and last four countries, alphabetically, for the sake of conciseness. Note that one of the steps in this program is to select only those students who have non-missing data in our variable of interest ATBGAGE. A second step consists of combining categories 1 and 2 and categories 5 and 6 of the variable ATBGAGE in order to match the results presented in Exhibit 5.14, where teachers are categorized into four groups: 29 years or under, 30-39 years old, 40-49 years old, and 50 years or older.

In general, to perform analyses using the teacher background data files, you should do the following:

1) Identify the variables of interest in the appropriate teacher background data files and note any specific national adaptations to the variables
2) Retrieve the relevant variables from the teacher background data files, including analysis and classification variables, identification variables (IDCNTRY, IDTEACH, and IDLINK), and any other variables used in the selection of cases
3) Retrieve the relevant variables from the student-teacher linkage data files, including plausible values for achievement, classification variables, identification variables (IDCNTRY, IDSTUD, IDTEACH, and IDLINK), sampling (JKZONE and JKREP) and weighting (TCHWGT) variables, and any other variables used in the selection of cases
4) Merge the teacher background data files with the student-teacher linkage data files using the variables IDCNTRY, IDTEACH and IDLINK
5) Perform any necessary variable transformations or recodes
6) Use the macros JACKGEN and JACKREG, or JACKPV and JACKREGP if plausible values are involved, with the appropriate arguments and parameters
7) Specify the location of the data files (<datpath>) and the macros (<macpath>)
8) Print out the results file

Exhibit 5.15 Example SAS Program to Analyze Teacher Variables (EXAMPLE3.SAS)

```
LIBNAME AR2 "<datpath>" ;

%INCLUDE "<macpath>JACKPV.SAS" ;

PROC SORT DATA = AR2.ATGALLR2 OUT = ATGALLR2 ;
     BY IDCNTRY IDTEACH IDLINK ;

PROC SORT DATA = AR2.ASTALLR2 OUT = ASTALLR2 ;
     BY IDCNTRY IDTEACH IDLINK ;

DATA MERGED ;
     MERGE ATGALLR2 (IN = INATG)
           ASTALLR2 (IN = INAST) ;
     BY IDCNTRY IDTEACH IDLINK ;
     IF INATG AND INAST ;

DATA MERGED ;
     SET MERGED ;

     IF NMISS (ATBGAGE) = 0 ;

     SELECT (ATBGAGE) ;
       WHEN (1,2) ATBGAGE = 1 ;  * 29 YEARS OR UNDER ;
       WHEN (3)   ATBGAGE = 2 ;  * 30-39 YEARS OLD ;
       WHEN (4)   ATBGAGE = 3 ;  * 40-49 YEARS OLD ;
       WHEN (5,6) ATBGAGE = 4 ;  * 50 YEARS OR OLDER ;
       OTHERWISE  ATBGAGE = . ;
     END ;

PROC FORMAT LIBRARY = WORK ;

 VALUE COUNTRY
   < list country formats >

VALUE AGE
     1 = '29 YEARS OR UNDER'
     2 = '30-39 YEARS OLD'
     3 = '40-49 YEARS OLD'
     4 = '50 YEARS OR OLDER' ;

%JACKPV (TCHWGT, JKZONE, JKREP, 75, IDCNTRY ATBGAGE, ASRREA0, 5, MERGED) ;

PROC PRINT DATA = FINAL NOOBS ;
     VAR IDCNTRY ATBGAGE N TCHWGT MNPV MNPV_SE PCT PCT_SE ;
     FORMAT IDCNTRY COUNTRY. ATBGAGE AGE. N 6.0 TCHWGT 10.0
            MNPV MNPV_SE PCT PCT_SE 6.2 ;
```

Exhibit 5.16 Output for Example Teacher Variable Analysis (EXAMPLE 3)

IDCNTRY	ATBGAGE	N	TCHWGT	MNPV	MNPV_SE	PCT	PCT_SE
AUSTRIA	29 YEARS OR UNDER	320	5289	540.32	6.34	6.41	1.65
AUSTRIA	30-39 YEARS OLD	822	13629	541.44	4.68	16.52	2.48
AUSTRIA	40-49 YEARS OLD	2115	32431	536.95	3.38	39.30	3.01
AUSTRIA	50 YEARS OR OLDER	1812	31169	538.79	3.47	37.77	3.17
BELGIUM FLEMISH	29 YEARS OR UNDER	1202	18077	544.18	3.40	27.71	3.22
BELGIUM FLEMISH	30-39 YEARS OLD	1321	18825	545.51	3.41	28.86	3.34
BELGIUM FLEMISH	40-49 YEARS OLD	1336	19196	554.58	2.08	29.43	2.85
BELGIUM FLEMISH	50 YEARS OR OLDER	551	9139	538.74	5.50	14.01	2.20
BELGIUM FRENCH	29 YEARS OR UNDER	718	7158	493.42	8.30	15.75	2.18
BELGIUM FRENCH	30-39 YEARS OLD	1443	15013	495.70	4.07	33.04	3.17
BELGIUM FRENCH	40-49 YEARS OLD	1569	16708	507.54	3.99	36.77	3.55
BELGIUM FRENCH	50 YEARS OR OLDER	614	6557	496.10	7.24	14.43	2.58
BULGARIA	29 YEARS OR UNDER	100	2135	519.46	21.13	3.47	1.39
BULGARIA	30-39 YEARS OLD	948	15916	541.78	12.08	25.90	2.80
BULGARIA	40-49 YEARS OLD	1603	24384	551.45	5.80	39.68	3.98
BULGARIA	50 YEARS OR OLDER	1089	19013	547.65	7.44	30.94	3.87
		.	.	.	.	.	
SPAIN	29 YEARS OR UNDER	448	39572	499.86	12.17	10.78	2.42
SPAIN	30-39 YEARS OLD	630	57765	523.01	5.52	15.73	2.57
SPAIN	40-49 YEARS OLD	825	91095	511.17	5.23	24.81	3.42
SPAIN	50 YEARS OR OLDER	1921	178735	512.97	3.79	48.68	3.81
SWEDEN	29 YEARS OR UNDER	428	8179	557.13	5.18	8.59	1.65
SWEDEN	30-39 YEARS OLD	1307	27345	550.14	4.12	28.71	3.33
SWEDEN	40-49 YEARS OLD	1083	23098	542.76	3.90	24.25	2.88
SWEDEN	50 YEARS OR OLDER	1697	36636	549.33	3.11	38.46	3.61
TRINIDAD AND TOBAGO	29 YEARS OR UNDER	399	1879	427.77	15.00	11.35	2.54
TRINIDAD AND TOBAGO	30-39 YEARS OLD	1411	6125	428.56	7.69	36.99	3.62
TRINIDAD AND TOBAGO	40-49 YEARS OLD	1012	4592	441.24	9.65	27.73	3.34
TRINIDAD AND TOBAGO	50 YEARS OR OLDER	992	3963	444.38	13.68	23.93	3.00
UNITED STATES	29 YEARS OR UNDER	1011	664097	536.49	8.16	20.72	2.75
UNITED STATES	30-39 YEARS OLD	1376	849591	534.22	4.61	26.51	2.84
UNITED STATES	40-49 YEARS OLD	1252	885575	548.52	4.89	27.63	3.72
UNITED STATES	50 YEARS OR OLDER	1316	805499	535.80	6.93	25.13	3.44

In this example, each country's results are presented on four lines, one for each value of the recoded ATBGAGE variable. The results are presented much in the same manner as in the previous example, where the countries are identified in the first column and the second column describes the categories of ATBGAGE being reported. From the first four lines displayed in Exhibit 5.16, we see that in Austria 6.41 percent of students were taught by teachers 29 years or younger, 16.52 percent by teachers 30 to 39 years old, 39.30 percent by teachers 40 to 49 years old, and 37.77 percent by teachers 50 years or older.

5.9 Performing Analyses with School-level Variables

PIRLS 2006 has representative samples of schools and so it is possible to compute weighted numbers of schools with particular characteristics for providing reasonable estimates of percentages and means across primary schools in each country. However, the school samples were designed to optimize the student

samples and the student-level estimates. For this reason, it is preferable to analyze school-level variables as attributes of students, rather than as elements in their own right.

For student-weighted analyses, the school-level data are analyzed to make statements about the students attending schools with one characteristic or another, rather than about the schools with one characteristic or another. When school-level variables are analyzed, we recommend that you merge the school background data files with the student background data files, and then use the sampling and weight information contained in the student-level file to make the desired analytical statements.

Our example of an analysis with school-level data will investigate the percentage of students who attend schools in areas of different population sizes, and their average reading achievement. The results of this analysis are presented in Exhibit 7.1 of the *PIRLS 2006 International Report* and are reproduced here in Exhibit 5.17. We will use the macro JACKPV in order to estimate the percentages and average achievement.

The first step in our analysis is to identify the variables of interest in the appropriate files and review the documentation on specific national adaptations to the questions of interest (Supplements 1 and 2). We observe that the variable ACBGCOMM in the school background data files contains information on the size of the community where the schools are located. Our next step is to review the documentation of national adaptations to the questionnaire to ensure that there were no deviations listed for this variable. If no adaptations were made, we can continue with our analysis without any modifications.

Since we are using a school-level variable, we need to look in the school background data files and link this to the student background data files to get the achievement scores. From the ACG data files, we will need the variable that contains the information on the schools' location (ACBGCOMM) and the identification variables IDCNTRY and IDSCHOOL that will allow us to link the school data to the student data.

We then read the variables of interest from the student background data files. We will need the country and school identification variables (IDCNTRY and IDSCHOOL), necessary to merge the school data to the student data. We also need the jackknife replication variables (JKZONE and JKREP), the student weighting variable (TOTWGT), and the reading achievement plausible values (ASRREA01 through ASRREA05).

Exhibit 5.17 Example School Variable Analysis Taken from the PIRLS 2006 International Report (Exhibit 7.1)

Exhibit 7.1 **Principals' Reports on Their Schools' Locations with Trends** — **PIRLS 2006** 4th Grade

Countries	Urban			Suburban			Rural		
	2006 Percent of Students	Average Achievement	Difference in Percent from 2001	2006 Percent of Students	Average Achievement	Difference in Percent from 2001	2006 Percent of Students	Average Achievement	Difference in Percent from 2001
Austria	31 (3.4)	529 (4.4)	◊ ◊	20 (3.2)	542 (5.5)	◊ ◊	48 (3.7)	543 (2.5)	◊ ◊
Belgium (Flemish)	21 (3.6)	541 (5.4)	◊ ◊	37 (4.7)	546 (3.7)	◊ ◊	42 (4.8)	551 (2.4)	◊ ◊
Belgium (French) r	47 (4.0)	494 (5.5)	◊ ◊	21 (3.9)	496 (5.7)	◊ ◊	33 (3.8)	512 (4.2)	◊ ◊
Bulgaria	70 (3.0)	557 (5.2)	6 (4.1)	5 (1.7)	550 (10.8)	−6 (3.1) ▼	24 (2.5)	516 (10.8)	0 (3.4)
Canada, Alberta	46 (4.5)	559 (3.8)	◊ ◊	26 (3.5)	572 (4.4)	◊ ◊	28 (3.5)	550 (4.1)	◊ ◊
Canada, British Columbia	38 (4.3)	555 (4.3)	◊ ◊	46 (4.7)	565 (3.5)	◊ ◊	17 (3.2)	545 (6.2)	◊ ◊
Canada, Nova Scotia	25 (3.2)	542 (6.5)	◊ ◊	26 (3.2)	551 (3.9)	◊ ◊	50 (3.5)	537 (2.6)	◊ ◊
Canada, Ontario	51 (4.9)	549 (4.1)	8 (6.9)	36 (5.0)	563 (4.3)	1 (7.0)	14 (3.2)	552 (4.9)	−9 (5.1)
Canada, Quebec	51 (4.8)	533 (3.7)	15 (6.7) ▲	28 (4.0)	538 (6.5)	−19 (6.3) ▼	20 (3.6)	528 (5.6)	3 (5.0)
Chinese Taipei	– –	– –	– –	– –	– –	– –	– –	– –	– –
Denmark	33 (4.1)	545 (3.4)	◊ ◊	30 (3.4)	555 (3.9)	◊ ◊	37 (4.0)	542 (4.4)	◊ ◊
England r	45 (4.2)	523 (5.3)	−2 (6.5)	35 (3.9)	553 (5.1)	4 (6.0)	19 (3.7)	564 (5.2)	−1 (5.2)
France	34 (4.0)	522 (4.3)	−2 (5.7)	25 (3.9)	518 (6.3)	−5 (5.7)	41 (3.9)	524 (2.5)	6 (5.3)
Georgia	42 (3.6)	486 (4.5)	◊ ◊	15 (2.7)	465 (8.1)	◊ ◊	43 (2.6)	459 (5.1)	◊ ◊
Germany	37 (3.3)	535 (4.6)	4 (4.4)	19 (3.1)	557 (3.9)	−3 (4.3)	44 (4.0)	555 (2.3)	−1 (5.5)
Hong Kong SAR	58 (4.4)	573 (3.1)	6 (5.3)	37 (4.1)	555 (4.5)	−9 (4.9)	5 (1.9)	540 (11.0)	3 (2.2)
Hungary	28 (2.2)	565 (6.6)	0 (3.3)	40 (2.5)	557 (5.0)	5 (3.4)	31 (1.8)	528 (4.7)	−5 (2.5) ▼
Iceland r	33 (0.3)	518 (2.0)	−3 (0.5) ▼	37 (0.3)	509 (2.3)	−6 (0.5) ▼	30 (0.4)	506 (2.3)	9 (0.5) ▲
Indonesia	12 (2.2)	451 (9.7)	◊ ◊	14 (2.7)	425 (9.6)	◊ ◊	74 (2.9)	393 (4.8)	◊ ◊
Iran, Islamic Rep. of	50 (2.9)	454 (4.1)	1 (4.7)	15 (2.2)	415 (10.3)	6 (3.3) ▲	35 (2.8)	376 (5.7)	−8 (4.2)
Israel	49 (3.9)	534 (6.0)	−3 (5.5)	18 (2.8)	529 (13.1)	−4 (4.3)	33 (3.8)	472 (10.2)	7 (5.0)
Italy	70 (3.6)	554 (2.9)	−6 (4.8)	15 (2.8)	555 (9.2)	1 (3.8)	15 (3.1)	533 (9.9)	5 (3.8)
Kuwait	26 (3.6)	355 (7.6)	◊ ◊	61 (4.0)	321 (5.8)	◊ ◊	14 (3.0)	311 (12.1)	◊ ◊
Latvia	70 (0.7)	548 (2.4)	26 (3.9) ▲	3 (1.5)	528 (6.6)	−15 (4.2) ▼	27 (1.7)	525 (5.9)	−10 (3.4) ▼
Lithuania	72 (2.3)	544 (1.9)	1 (3.6)	3 (1.4)	549 (10.7)	−3 (2.5)	26 (2.1)	516 (3.4)	3 (3.2)
[1] Luxembourg	– –	– –	– –	– –	– –	– –	– –	– –	– –
Macedonia, Rep. of r	51 (3.6)	477 (6.9)	−6 (4.9)	18 (3.5)	443 (13.8)	3 (4.5)	31 (2.6)	401 (9.5)	3 (4.1)
Moldova, Rep. of	29 (2.4)	517 (4.5)	2 (4.1)	6 (2.4)	498 (17.6)	−8 (4.5)	65 (2.5)	492 (3.9)	6 (4.5)
Morocco r	37 (3.3)	363 (7.2)	−3 (5.5)	18 (3.6)	334 (15.8)	−3 (5.7)	45 (3.7)	296 (13.5)	6 (5.5)
Netherlands	26 (4.0)	538 (4.1)	−5 (5.6)	33 (4.7)	553 (3.2)	10 (6.0)	41 (3.5)	547 (2.5)	−5 (5.2)
New Zealand	41 (3.2)	536 (3.3)	3 (5.1)	39 (3.0)	527 (3.6)	−1 (4.7)	21 (2.3)	535 (5.3)	−2 (3.7)
Norway	20 (3.6)	502 (3.8)	1 (5.0)	30 (3.9)	504 (3.9)	4 (5.5)	50 (4.2)	492 (4.2)	−6 (5.5)
Poland	52 (2.1)	528 (2.9)	◊ ◊	5 (1.7)	529 (11.0)	◊ ◊	43 (1.9)	508 (3.8)	◊ ◊
Qatar	65 (0.2)	362 (1.5)	◊ ◊	32 (0.2)	336 (2.0)	◊ ◊	3 (0.0)	318 (8.4)	◊ ◊
Romania	47 (2.2)	515 (6.5)	−3 (3.4)	5 (1.9)	498 (14.8)	1 (2.8)	48 (2.4)	462 (8.0)	2 (3.6)
Russian Federation	63 (2.0)	581 (3.4)	8 (3.2) ▲	6 (1.3)	563 (8.8)	4 (2.3)	31 (2.2)	532 (6.1)	−12 (3.3) ▼
Scotland r	32 (3.5)	517 (4.8)	−2 (5.6)	36 (4.3)	539 (5.3)	−3 (6.9)	32 (3.9)	528 (6.8)	6 (5.9)
Singapore	100 (0.0)	558 (2.9)	0 (0.0)	0 (0.0)	~ ~	0 (0.0)	0 (0.0)	~ ~	0 (0.0)
Slovak Republic	52 (3.0)	544 (2.9)	2 (4.8)	8 (2.5)	537 (8.0)	−2 (3.5)	40 (3.3)	512 (5.9)	0 (4.5)
Slovenia	36 (4.2)	529 (3.6)	−4 (5.3)	37 (4.0)	520 (3.1)	10 (5.5)	27 (3.7)	512 (3.6)	−6 (4.7)
South Africa	17 (1.8)	350 (19.5)	◊ ◊	21 (2.2)	381 (14.9)	◊ ◊	62 (2.0)	261 (3.8)	◊ ◊
Spain	58 (4.3)	524 (3.4)	◊ ◊	20 (3.3)	497 (6.8)	◊ ◊	21 (3.4)	498 (7.1)	◊ ◊
Sweden	27 (4.1)	549 (3.8)	12 (5.1) ▲	55 (4.1)	549 (3.4)	−12 (5.6) ▼	18 (2.8)	550 (4.7)	0 (4.5)
Trinidad and Tobago	19 (2.5)	470 (13.0)	◊ ◊	50 (3.7)	441 (7.0)	◊ ◊	32 (3.0)	408 (8.9)	◊ ◊
United States	28 (3.5)	524 (4.4)	−5 (4.9)	47 (3.9)	550 (3.2)	13 (6.1) ▲	25 (2.7)	539 (9.1)	−8 (4.2)
International Avg.	43 (0.5)	508 (1.0)		24 (0.5)	501 (1.4)		33 (0.5)	483 (1.1)	

▲ Percent in 2006 significantly higher
▼ Percent in 2006 significantly lower

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

We then proceed to merge the school data with the student data using the variables IDCNTRY and IDSCHOOL, and then use the macro JACKPV to obtain the percentages of students and their mean achievement scores within each category of the variable ACBGCOMM for each country. For this analysis, we will

use the data for all available countries, making use of an aggregated school file ACGALLR2 and an aggregated student file ASGALLR2.

The SAS program that implements this fourth example is presented in Exhibit 5.18 and is included on the DVD under the name EXAMPLE4.SAS. The results of this program are displayed in Exhibit 5.19, edited to show only the first five and last five countries, alphabetically, for the sake of brevity. Note that one of the steps in this program is to select only those students who have non-missing data in our variable of interest ACBGCOMM.

In general, to perform analyses using the school background data files, you should do the following:

1) Identify the variables of interest in the school background data files and note any specific national adaptations to the variables
2) Retrieve the relevant variables from the school background data files, including analysis and classification variables, identification variables (IDCNTRY and IDSCHOOL), and any other variables used in the selection of cases
3) Retrieve the relevant variables from the student background data files, including plausible values for achievement, classification variables, identification variables (IDCNTRY and IDSCHOOL), sampling (JKZONE and JKREP) and weighting (TOTWGT) variables, and any other variables used in the selection of cases
4) Merge the school background data files with the student background data files using the variables IDCNTRY and IDSCHOOL
5) Perform any necessary variable transformations or recodes
6) Use the macros JACKGEN and JACKREG, or JACKPV and JACKREGP if plausible values are involved, with the appropriate arguments and parameters
7) Specify the location of the data files (<datpath>) and the macros (<macpath>)
8) Print out the results file

Exhibit 5.18 First Example SAS Program for School Variable Analysis (EXAMPLE4.SAS)

```
LIBNAME AR2 "<datpath>" ;

%INCLUDE "<macpath>JACKPV.SAS" ;

PROC SORT DATA = AR2.ACGALLR2 OUT = ACGALLR2 ;
     BY IDCNTRY IDSCHOOL ;

PROC SORT DATA = AR2.ASGALLR2 OUT = ASGALLR2 ;
     BY IDCNTRY IDSCHOOL ;

DATA MERGED ;
     MERGE ACGALLR2 (IN = INACG)
           ASGALLR2 (IN = INASG) ;
     BY IDCNTRY IDSCHOOL ;
     IF INACG AND INASG ;

DATA MERGED ;
     SET MERGED ;

     IF NMISS (ACBGCOMM) = 0 ;

PROC FORMAT LIBRARY = WORK ;

 VALUE COUNTRY
   < list country formats >

VALUE COMM
     1 = 'URBAN'
     2 = 'SUBURBAN'
     3 = 'RURAL' ;

%JACKPV (TOTWGT, JKZONE, JKREP, 75, IDCNTRY ACBGCOMM, ASRREA0, 5, MERGED) ;

PROC PRINT DATA = FINAL NOOBS ;
     VAR IDCNTRY ACBGCOMM N TOTWGT MNPV MNPV_SE PCT PCT_SE ;
     FORMAT IDCNTRY COUNTRY. ACBGCOMM COMM. N 6.0 TOTWGT 10.0
            MNPV MNPV_SE PCT PCT_SE 6.2 ;
```

Exhibit 5.19 Output for First Example School Variable Analysis (EXAMPLE 4)

IDCNTRY	ACBGCOMM	N	TOTWGT	MNPV	MNPV_SE	PCT	PCT_SE
AUSTRIA	URBAN	1848	25951	528.62	4.43	31.40	3.39
AUSTRIA	SUBURBAN	1235	16743	541.75	5.46	20.26	3.24
AUSTRIA	RURAL	1949	39956	543.12	2.51	48.34	3.69
BELGIUM FLEMISH	URBAN	909	13199	540.79	5.39	20.80	3.61
BELGIUM FLEMISH	SUBURBAN	1592	23749	545.58	3.67	37.43	4.72
BELGIUM FLEMISH	RURAL	1777	26498	550.93	2.45	41.77	4.81
BELGIUM FRENCH	URBAN	2055	18745	493.55	5.47	46.72	3.98
BELGIUM FRENCH	SUBURBAN	785	8258	496.47	5.71	20.58	3.93
BELGIUM FRENCH	RURAL	1022	13121	511.77	4.16	32.70	3.85
BULGARIA	URBAN	2976	42641	557.48	5.21	70.41	3.03
BULGARIA	SUBURBAN	221	3158	549.62	10.83	5.22	1.75
BULGARIA	RURAL	467	14760	516.24	10.84	24.37	2.46
CANADA ALBERTA	URBAN	1830	16556	559.40	3.84	46.07	4.51
CANADA ALBERTA	SUBURBAN	1325	9429	571.86	4.36	26.23	3.46
CANADA ALBERTA	RURAL	1018	9955	550.29	4.09	27.70	3.49
		.	.	.	.	.	
SOUTH AFRICA	URBAN	2146	139387	350.20	19.47	17.29	1.77
SOUTH AFRICA	SUBURBAN	2754	166761	380.80	14.93	20.68	2.20
SOUTH AFRICA	RURAL	9469	500254	261.11	3.75	62.04	2.04
SPAIN	URBAN	2349	208013	524.21	3.40	58.36	4.26
SPAIN	SUBURBAN	732	72090	496.57	6.76	20.23	3.33
SPAIN	RURAL	638	76330	498.21	7.07	21.42	3.37
SWEDEN	URBAN	1368	27171	548.74	3.76	26.99	4.11
SWEDEN	SUBURBAN	2507	55264	548.99	3.35	54.89	4.08
SWEDEN	RURAL	479	18242	550.01	4.71	18.12	2.77
TRINIDAD AND TOBAGO	URBAN	907	3162	469.93	13.03	18.74	2.55
TRINIDAD AND TOBAGO	SUBURBAN	1954	8380	441.49	7.04	49.66	3.68
TRINIDAD AND TOBAGO	RURAL	992	5333	407.74	8.87	31.60	3.01
UNITED STATES	URBAN	1461	908132	524.07	4.45	27.50	3.50
UNITED STATES	SUBURBAN	2290	1563272	549.59	3.22	47.34	3.94
UNITED STATES	RURAL	1349	830812	538.95	9.13	25.16	2.72

In this example, each country's results are presented on three lines, one for each value of the ACBGCOMM variable, with the country name in the first column and label of categories of ACBGCOMM being reported in the second column. All other columns are identical in nature to those in the previous two examples.

From the first three lines in Exhibit 5.19, we see that 31.40 percent of Austrian students come from schools in urban areas, 20.26 percent come from schools in suburban areas, and 48.34 percent from rural areas. We also see that the estimated average reading achievement of students in urban areas is 528.62 (with a standard error of 4.43), whereas the estimated average achievement of students in suburban and rural areas are, respectively, 541.75 (standard error of 5.46) and 543.12 (standard error of 2.51).

5.10 Performing Analyses with Home Background Variables

In PIRLS 2006, the students' parents or primary caregivers were asked to complete a questionnaire. The responses to the *Learning to Read Survey* are included in the home background data files. Although home background variables are located in separate files, they are in essence attributes of students and must be analyzed in the same manner as student background variables. This will require us to merge the home background data files with the student background data files.

Our example of a home background analysis will investigate the relationship between the frequency of parents' reading for enjoyment and their children's reading achievement. These results are presented in Exhibit 3.9 of the *PIRLS 2006 International Report* and are duplicated here in Exhibit 5.20. We will use the macro JACKPV in order to estimate the percentages and mean achievement.

The first step in our analysis is to identify the variables of interest in the appropriate files and review the documentation on specific national adaptations to the questions of interest (Supplements 1 and 2). We observe that the variable ASBHRRE in the home background data files contains categorical information on the frequency of parents' reading for enjoyment. Our next step is to review the documentation of national adaptations to the questionnaires to ensure that there were no deviations listed for this variable (see Supplement 2). If no national adaptations were made, we can continue with our analysis without any modifications.

Since we are interested in using a home background variable, as well as student achievement in our analysis, we need to link the home background data files and the student background data files to get the achievement scores. From the ASH data files, we will need our variable of interest (ASBHRRE) and the identification variables IDCNTRY and IDSTUD that will allow us to link the home data to the student data. From the ASG data files, we will need the country and student identification variables (IDCNTRY and IDSTUD) necessary to merge the home data to the student data. We also need the jackknife replication variables (JKZONE and JKREP), the student weighting variable (TOTWGT), and the reading achievement plausible values (ASRREA01 through ASRREA05).

We then proceed to merge the home and student data files by IDCNTRY and IDSTUD. The macro JACKPV is used to obtain the percentages of students and average achievement scores within each category of the variable ASBHRRE for each country, making use of an aggregated home file ASHALLR2 and an aggregated student file ASGALLR2.

Exhibit 5.20 Example Home Background Variable Analysis Taken From the PIRLS 2006 International Report (Exhibit 3.9)

Exhibit 3.9 **Parents Reading for Enjoyment with Trends** — **PIRLS 2006** 4th Grade

Countries		Every Day or Almost Every Day			Once or Twice a Week			Twice a Month or Less		
		2006 Percent of Students	Average Achievement	Difference in Percent from 2001	2006 Percent of Students	Average Achievement	Difference in Percent from 2001	2006 Percent of Students	Average Achievement	Difference in Percent from 2001
Norway		66 (1.0)	505 (2.5)	7 (1.6) ▲	23 (0.9)	493 (4.2)	−4 (1.4) ▼	11 (0.7)	489 (6.5)	−3 (0.9) ▼
Sweden		64 (0.9)	556 (2.4)	−3 (1.3) ▼	21 (0.8)	546 (2.9)	−1 (1.1)	15 (0.7)	540 (4.6)	5 (0.9) ▲
Scotland	s	63 (1.3)	550 (4.0)	1 (1.8)	23 (0.9)	531 (5.3)	−3 (1.5)	14 (1.1)	525 (7.0)	2 (1.4)
Iceland	r	61 (0.9)	523 (1.9)	−1 (1.2)	21 (0.8)	513 (2.9)	−4 (1.1) ▼	18 (0.6)	498 (3.3)	5 (0.8) ▲
New Zealand	s	60 (1.1)	557 (2.4)	4 (1.6) ▲	25 (0.9)	539 (3.6)	−4 (1.4) ▼	15 (0.7)	528 (4.1)	0 (1.1)
Netherlands	s	59 (1.2)	562 (1.8)	−1 (1.8)	25 (0.9)	548 (2.9)	−3 (1.3) ▼	16 (0.9)	543 (2.6)	4 (1.2) ▲
Trinidad and Tobago		59 (0.9)	441 (5.1)	◊ ◊	31 (1.0)	436 (6.1)	◊ ◊	10 (0.7)	434 (9.2)	◊ ◊
Germany	r	58 (0.9)	561 (2.4)	7 (1.3) ▲	28 (0.7)	544 (2.9)	−3 (1.0) ▼	14 (0.6)	535 (4.5)	−3 (0.9) ▼
Denmark		57 (1.0)	554 (2.6)	◊ ◊	27 (0.9)	544 (3.0)	◊ ◊	16 (0.7)	534 (5.0)	◊ ◊
Latvia		57 (1.3)	547 (2.7)	3 (2.0)	32 (1.1)	537 (3.4)	−3 (1.5)	11 (0.6)	532 (4.2)	0 (1.2)
Canada, British Columbia	r	57 (1.1)	569 (3.2)	◊ ◊	29 (1.1)	560 (3.5)	◊ ◊	15 (0.7)	547 (4.7)	◊ ◊
Canada, Alberta	r	55 (1.2)	571 (2.8)	◊ ◊	29 (1.0)	561 (3.5)	◊ ◊	17 (0.7)	554 (3.7)	◊ ◊
Luxembourg		54 (0.7)	573 (1.3)	◊ ◊	29 (0.7)	544 (2.4)	◊ ◊	18 (0.6)	543 (2.4)	◊ ◊
Canada, Nova Scotia		53 (0.9)	556 (2.1)	◊ ◊	27 (0.8)	539 (2.8)	◊ ◊	19 (0.7)	528 (3.2)	◊ ◊
Austria		53 (1.1)	549 (2.3)	◊ ◊	30 (0.9)	533 (3.2)	◊ ◊	17 (0.7)	524 (2.9)	◊ ◊
Lithuania		51 (1.1)	542 (2.0)	−1 (1.5)	36 (1.0)	532 (2.1)	−1 (1.4)	12 (0.6)	534 (2.6)	2 (0.9)
Belgium (French)		51 (1.3)	512 (2.9)	◊ ◊	32 (0.9)	496 (3.3)	◊ ◊	17 (0.8)	483 (4.4)	◊ ◊
Canada, Ontario	r	51 (1.0)	563 (3.4)	−2 (1.6)	32 (1.1)	552 (3.0)	1 (1.6)	18 (0.9)	547 (4.5)	1 (1.3)
France		50 (1.0)	535 (2.5)	−1 (1.4)	32 (0.8)	518 (2.6)	−2 (1.2)	19 (0.8)	505 (3.0)	3 (1.1) ▲
Spain	s	50 (1.2)	531 (3.0)	◊ ◊	33 (1.0)	512 (3.6)	◊ ◊	18 (0.7)	507 (4.5)	◊ ◊
Singapore		50 (0.7)	569 (3.0)	14 (1.1) ▲	33 (0.7)	552 (3.1)	−3 (0.9) ▼	17 (0.5)	546 (4.3)	−11 (0.9) ▼
Canada, Quebec	r	49 (1.4)	542 (3.2)	1 (1.9)	32 (1.2)	536 (3.5)	−2 (1.7)	18 (0.9)	518 (4.1)	1 (1.4)
Italy		49 (1.0)	564 (3.0)	2 (1.4)	32 (1.0)	547 (3.7)	−4 (1.2) ▼	19 (0.9)	535 (4.2)	2 (1.2)
Hungary		49 (1.0)	561 (3.2)	−2 (1.4)	37 (0.8)	546 (3.7)	1 (1.2)	14 (0.6)	544 (4.4)	1 (0.9)
Macedonia, Rep. of	r	48 (1.1)	459 (4.6)	7 (1.7) ▲	41 (0.9)	441 (4.5)	0 (1.4)	11 (0.9)	423 (8.4)	−7 (1.5) ▼
South Africa	r	48 (0.7)	322 (7.5)	◊ ◊	36 (0.5)	301 (5.2)	◊ ◊	16 (0.5)	276 (6.3)	◊ ◊
Israel		47 (1.4)	538 (5.0)	x x	39 (1.2)	520 (4.4)	x x	15 (0.7)	508 (6.5)	x x
Slovak Republic		46 (1.0)	544 (2.4)	−2 (1.5)	39 (1.0)	529 (3.2)	0 (1.4)	14 (0.8)	504 (7.7)	3 (1.0) ▲
Slovenia		45 (0.8)	530 (2.4)	4 (1.3) ▲	36 (0.8)	520 (2.6)	−6 (1.2) ▼	18 (0.6)	510 (3.3)	2 (0.9) ▲
Poland		45 (0.9)	529 (3.0)	◊ ◊	41 (0.8)	517 (2.8)	◊ ◊	14 (0.6)	504 (4.5)	◊ ◊
Qatar	r	44 (0.7)	361 (2.3)	◊ ◊	39 (0.7)	357 (2.3)	◊ ◊	17 (0.5)	348 (3.8)	◊ ◊
Kuwait	r	44 (1.2)	339 (5.4)	◊ ◊	36 (0.9)	337 (5.4)	◊ ◊	20 (0.9)	335 (5.9)	◊ ◊
Russian Federation		42 (1.0)	573 (3.7)	−7 (1.6) ▼	41 (0.8)	560 (3.9)	6 (1.4) ▲	17 (0.6)	557 (3.6)	1 (1.1)
Belgium (Flemish)		40 (0.9)	558 (2.2)	◊ ◊	33 (0.8)	545 (2.4)	◊ ◊	27 (1.0)	538 (2.4)	◊ ◊
Georgia		39 (1.3)	486 (3.4)	◊ ◊	44 (1.0)	467 (3.2)	◊ ◊	17 (1.2)	450 (7.6)	◊ ◊
Bulgaria		39 (1.4)	565 (4.2)	−14 (2.1) ▼	34 (1.1)	547 (4.7)	6 (1.6) ▲	27 (2.0)	532 (7.6)	8 (2.6) ▲
Hong Kong SAR		36 (0.8)	574 (2.3)	6 (1.4) ▲	39 (0.7)	561 (2.6)	10 (1.0) ▲	25 (0.6)	558 (3.4)	−15 (1.3) ▼
Chinese Taipei		35 (0.9)	547 (2.5)	◊ ◊	42 (0.7)	536 (2.5)	◊ ◊	23 (0.7)	520 (2.4)	◊ ◊
Moldova, Rep. of		35 (1.1)	509 (3.3)	−1 (1.9)	46 (1.1)	498 (3.5)	6 (1.7) ▲	20 (1.5)	491 (5.9)	−5 (1.9) ▼
Indonesia		31 (1.2)	413 (4.6)	◊ ◊	44 (1.3)	405 (4.5)	◊ ◊	25 (1.3)	395 (4.9)	◊ ◊
Romania		27 (1.1)	514 (4.7)	11 (1.6) ▲	43 (1.0)	497 (4.8)	11 (1.7) ▲	30 (1.4)	459 (8.0)	−22 (2.4) ▼
Morocco		24 (1.1)	346 (6.4)	– –	34 (1.4)	329 (5.3)	– –	42 (1.9)	308 (10.9)	– –
Iran, Islamic Rep. of		24 (1.1)	441 (4.9)	2 (1.5)	41 (1.1)	426 (3.4)	6 (1.8) ▲	35 (1.6)	402 (4.3)	−8 (2.2) ▼
England		x x	x x	x x	x x	x x	x x	x x	x x	x x
United States		– –	– –	– –	– –	– –	– –	– –	– –	– –
International Avg.		47 (0.2)	512 (0.6)		34 (0.2)	498 (0.6)		18 (0.2)	487 (0.9)	

▲ Percent in 2006 significantly higher
▼ Percent in 2006 significantly lower

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

The SAS program that implements this example is presented in Exhibit 5.21 and is included on the DVD under the name EXAMPLE5.SAS. The results obtained from this program are displayed in Exhibit 5.22, showing the first five and last five countries, as sorted alphabetically. Note that one of the steps in this program is to select only those students who have non-missing data in our variable of interest

ASBHRRE. A second step consists of combining category 4 of the variable ASBHRRE with category 3 to create a category we will label "Twice a Month or Less".

In general, to perform analyses with home background variables using the home background data files, you should do the following:

1) Identify the variables of interest in the home background data files and note any specific national adaptations to the variables
2) Retrieve the relevant variables from the home background data files, including analysis and classification variables, identification variables (IDCNTRY and IDSTUD), and any other variables used in the selection of cases
3) Retrieve the relevant variables from the student background data files, including plausible values for achievement, classification variables, identification variables (IDCNTRY and IDSTUD), sampling (JKZONE and JKREP) and weighting (TOTWGT) variables, and any other variables used in the selection of cases
4) Merge the home background data files with the student background data files using the variables IDCNTRY and IDSTUD
5) Perform any necessary variable transformations or recodes
6) Use the macros JACKGEN and JACKREG, or JACKPV and JACKREGP if plausible values are involved, with the appropriate arguments and parameters
7) Specify the location of the data files (<datpath>) and the macros (<macpath>)
8) Print out the results file

Exhibit 5.21 Example SAS Program to Perform Home Background Variable Analysis (EXAMPLE5.SAS)

```
LIBNAME AR2 "<datpath>" ;

%INCLUDE "<macpath>JACKPV.SAS" ;

PROC SORT DATA = AR2. ASHALLR2 OUT = ASHALLR2 ;
     BY IDCNTRY IDSTUD ;

PROC SORT DATA = AR2.ASGALLR2 OUT = ASGALLR2 ;
     BY IDCNTRY IDSTUD;

DATA MERGED ;
     MERGE ASHALLR2 (IN = INASH)
           ASGALLR2 (IN = INASG) ;
     BY IDCNTRY IDSTUD;
     IF INASH AND INASG ;

DATA MERGED ;
     SET MERGED ;

     IF NMISS (ASBHRRE) = 0 ;

     SELECT (ASBHRRE) ;
       WHEN (1)   ASBHRRE = 1 ;  * EVERY DAY OR ALMOST EVERY DAY ;
       WHEN (2)   ASBHRRE = 2 ;  * ONCE OR TWICE A WEEK ;
       WHEN (3,4) ASBHRRE = 3 ;  * TWICE A MONTH OR LESS ;
       OTHERWISE  ASBHRRE = . ;
     END ;

PROC FORMAT LIBRARY = WORK ;

 VALUE COUNTRY
   < list country formats >

VALUE RRE
     1 = 'EVERY DAY OR ALMOST EVERY DAY'
     2 = 'ONCE OR TWICE A WEEK'
     3 = 'TWICE A MONTH OR LESS' ;

%JACKPV (TOTWGT, JKZONE, JKREP, 75, IDCNTRY ASBHRRE, ASRREA0, 5, MERGED) ;

PROC PRINT DATA = FINAL NOOBS ;
     VAR IDCNTRY ASBHRRE N TOTWGT MNPV MNPV_SE PCT PCT_SE ;
     FORMAT IDCNTRY COUNTRY. ASBHRRE RRE. N 6.0 TOTWGT 10.0
            MNPV MNPV_SE PCT PCT_SE 6.2 ;
```

Exhibit 5.22 Output for Example Home Background Variable Analysis (EXAMPLE 6)

IDCNTRY	ASBHRRE	N	TOTWGT	MNPV	MNPV_SE	PCT	PCT_SE
AUSTRIA	EVERY DAY OR ALMOST EVERY DAY	2573	41727	548.84	2.30	52.91	1.06
AUSTRIA	ONCE OR TWICE A WEEK	1440	23805	532.70	3.16	30.18	0.86
AUSTRIA	TWICE A MONTH OR LESS	787	13337	523.95	2.95	16.91	0.66
BELGIUM FLEMISH	EVERY DAY OR ALMOST EVERY DAY	1753	25651	557.56	2.20	40.34	0.91
BELGIUM FLEMISH	ONCE OR TWICE A WEEK	1393	20847	545.17	2.41	32.78	0.83
BELGIUM FLEMISH	TWICE A MONTH OR LESS	1174	17090	538.20	2.45	26.88	0.96
BELGIUM FRENCH	EVERY DAY OR ALMOST EVERY DAY	2087	21692	511.68	2.90	50.95	1.29
BELGIUM FRENCH	ONCE OR TWICE A WEEK	1302	13474	496.11	3.25	31.65	0.93
BELGIUM FRENCH	TWICE A MONTH OR LESS	693	7410	483.28	4.43	17.40	0.83
BULGARIA	EVERY DAY OR ALMOST EVERY DAY	1540	23568	564.81	4.24	38.84	1.44
BULGARIA	ONCE OR TWICE A WEEK	1294	20638	547.33	4.69	34.01	1.14
BULGARIA	TWICE A MONTH OR LESS	871	16478	532.46	7.62	27.15	1.98
CANADA ALBERTA	EVERY DAY OR ALMOST EVERY DAY	1847	16063	570.51	2.81	54.76	1.20
CANADA ALBERTA	ONCE OR TWICE A WEEK	990	8394	561.38	3.52	28.61	0.98
CANADA ALBERTA	TWICE A MONTH OR LESS	579	4879	553.60	3.71	16.63	0.70
	. . . . .						
SLOVENIA	EVERY DAY OR ALMOST EVERY DAY	2319	7558	530.38	2.42	45.42	0.81
SLOVENIA	ONCE OR TWICE A WEEK	1812	6028	520.47	2.57	36.22	0.84
SLOVENIA	TWICE A MONTH OR LESS	911	3056	510.26	3.30	18.37	0.62
SOUTH AFRICA	EVERY DAY OR ALMOST EVERY DAY	5697	323919	321.78	7.52	48.03	0.67
SOUTH AFRICA	ONCE OR TWICE A WEEK	4478	242429	300.65	5.17	35.95	0.54
SOUTH AFRICA	TWICE A MONTH OR LESS	1879	108056	275.65	6.32	16.02	0.52
SPAIN	EVERY DAY OR ALMOST EVERY DAY	1287	119919	530.99	3.00	49.55	1.23
SPAIN	ONCE OR TWICE A WEEK	825	79606	512.14	3.65	32.89	1.03
SPAIN	TWICE A MONTH OR LESS	443	42493	507.20	4.51	17.56	0.74
SWEDEN	EVERY DAY OR ALMOST EVERY DAY	2679	61207	555.81	2.44	64.22	0.88
SWEDEN	ONCE OR TWICE A WEEK	862	20167	546.35	2.95	21.16	0.77
SWEDEN	TWICE A MONTH OR LESS	563	13929	539.87	4.56	14.62	0.74
TRINIDAD AND TOBAGO	EVERY DAY OR ALMOST EVERY DAY	2029	8986	441.41	5.08	58.81	0.92
TRINIDAD AND TOBAGO	ONCE OR TWICE A WEEK	1126	4755	436.29	6.08	31.12	0.97
TRINIDAD AND TOBAGO	TWICE A MONTH OR LESS	363	1539	433.75	9.22	10.07	0.66

Each country's results are presented on three lines, one for each value of the ASBHRRE variable. The countries are identified in the first column and the second column describes the reporting categories for ASBHRRE. All other columns are identical to those in the previous examples.

From the first three lines in Exhibit 5.22, we see that 52.91 percent of students in Austria have parents who reported reading for enjoyment every day or almost every day, 30.18 percent have parents who read for enjoyment once or twice a week, and 16.91 percent twice a month or less. We also see that the estimated average reading achievement of students whose parents read for enjoyment every day or almost every day is 548.84 (with a standard error of 2.30), of those whose parents read once or twice a week is 532.70 (standard error of 3.16), and of those whose parents read twice a month or less is 523.95 (standard error of 2.95).

References

Mullis, I.V.S., Martin, M.O., Kennedy, A.M., & Foy, P. (2007). *PIRLS 2006 international report: IEA's progress in international reading literacy study in primary schools in 40 countries*. Chestnut Hill, MA: Boston College.

SAS Institute (2002). *SAS system for Windows* (version 9.1). Cary, NC: SAS Institute.

Appendix A

Acknowledgements

Introduction

PIRLS 2006 was a collaborative effort involving hundreds of individuals around the world. This appendix recognizes the individuals and organizations for their contributions. Given that work on PIRLS 2006 has spanned approximately 5 years and has involved so many people and organizations, this list may not include all who contributed. Any omission is inadvertent.

Of the first importance, PIRLS 2006 is deeply indebted to the students, parents, teachers, and school principals who contributed their time and effort to the study.

Management and Coordination

PIRLS is a major undertaking of IEA, and together with the Trends in International Mathematics and Science Study (TIMSS), comprises the core of IEA's regular cycles of studies. The PIRLS assessment at the fourth grade complements TIMSS, which regularly assesses mathematics and science achievement at fourth and eighth grades.

The TIMSS & PIRLS International Study Center at Boston College has responsibility for the overall direction and management of the TIMSS and PIRLS projects. Headed by Drs. Ina V.S. Mullis and Michael O. Martin, the study center is located in the Lynch School of Education. Dr. Ann M. Kennedy is the PIRLS Project Coordinator. In carrying out the project, the TIMSS & PIRLS International Study Center worked closely with the IEA Secretariat in Amsterdam, which provided guidance overall and was responsible for verification of all translations produced by the participating countries. The IEA Data Processing and Research Center in Hamburg was responsible for processing and verifying the data submitted by the participants; Statistics Canada in Ottawa was responsible for school and student

sampling activities; and Educational Testing Service (ETS) in Princeton, New Jersey consulted on psychometric methodology and provided software for scaling the achievement data.

The Project Management Team, comprised of study directors and representatives from the TIMSS & PIRLS International Study Center, the IEA Secretariat, the IEA Data Processing and Research Center, Statistics Canada, and ETS met twice a year throughout the study to discuss the study's progress, procedures, and schedule. In addition, the study directors met with members of IEA's Technical Executive Group twice yearly to review technical issues.

Dr. Marian Sainsbury from the National Foundation for Educational Research in England (NFER) was the PIRLS 2006 Reading Coordinator and Dr. Patricia Donahue from ETS was a special reading assessment consultant. Together with the Reading Development Group, a panel of internationally recognized experts in reading research, instruction, and assessment, they provided excellent guidance throughout PIRLS 2006.

To work with the international team and coordinate within-country activities, each participating country designated an individual to be the PIRLS National Research Coordinator (NRC). The NRCs have the complicated and challenging task of implementing the PIRLS study in their countries in accordance with the PIRLS guidelines and procedures. The quality of the PIRLS assessment and data depends on the work of the NRCs and their colleagues in carrying out the very complex sampling, data collection, and scoring tasks involved. In addition, the Questionnaire Development Group, comprised of NRCs, provided advice on questionnaire development.

Continuing the tradition of truly exemplary work established in PIRLS 2001, the PIRLS 2006 NRCs (often the same NRCs as in 2001), performed their many tasks with dedication, competence, energy, and goodwill, and have been commended by the IEA Secretariat, the TIMSS & PIRLS International Study Center, the IEA Data Processing and Research Center, and Statistics Canada for their commitment to the project and the high quality of their work.

Funding

A project of this magnitude requires considerable financial support. IEA's major funding partners for PIRLS included the World Bank, the U.S. Department of Education through the National Center for Education Statistics, and those countries that contributed by way of fees. The financial support provided by Boston College and NFER also is gratefully acknowledged.

IEA Secretariat

Seamus Hegarty, IEA Chair
Hans Wagemaker, Executive Director
Barbara Malak, Manager, Membership Relations
Juriaan Hartenburg, Financial Manager
Suzanne Morony, Senior Manager Assistant

TIMSS & PIRLS International Study Center at Boston College

Ina V.S. Mullis, Co-Director
Michael O. Martin, Co-Director
Pierre Foy, Director of Sampling and Data Analysis
Ann M. Kennedy, Coordinator of Project Development and Operations, PIRLS Coordinator
Alka Arora, TIMSS Advanced 2008 Project Coordinator
Debra Berger, Production Editor
Marcie Bligh, Manager of Office Administration
Joann Cusack, Administrative Coordinator
Ebru Erberber, TIMSS Science Research Associate
Susan Farrell, Co-Manager of Publications
Joseph Galia, Senior Statistician/Programmer
Christine Hoage, Manager of Finance
Ieva Johansone, Survey Operations Coordinator
Isaac Li, Statistician/Programmer
Dana Milne, TIMSS Graduate Assistant
Jennifer Moher, Data Graphics Specialist
Mario Pita, Co-Manager of Publications
Corinna Preuschoff, TIMSS Mathematics Research Associate
Ruthanne Ryan, Data Graphics Specialist
Gabrielle Stanco, TIMSS Graduate Assistant
Feng Tian, TIMSS Graduate Assistant
Kathleen L. Trong, PIRLS Research Associate

IEA Data Processing and Research Center

Dirk Hastedt, Co-Director
Juliane Barth, Co-Manager, TIMSS & PIRLS Data Processing
Oliver Neuschmidt, Co-Manager, TIMSS & PIRLS Data Processing
Yasin Afana, Researcher
Milena Taneva, Researcher
Marta Kostek-Drosihn, Researcher
Sabine Meinck, Researcher
Olaf Zuehlke, Researcher
Christine Busch, Researcher
Alena Becker, Researcher
Simone Uecker, Researcher
Michael Jung, Researcher
Tim Daniel, Researcher
Dirk Oehler, Researcher
Stephan Petzchen, Senior Programmer
Ralph Carstens, Programmer
Hauke Heyen, Programmer
Harpreet Singh Choudry, Programmer

Statistics Canada

Marc Joncas, Senior Methodologist

Educational Testing Service

Mathias von Davier, Senior Research Scientist

Sampling Referee

Keith Rust, Vice President and Associate Director of the Statistical Group, Westat, Inc.

PIRLS 2006 Assessment Development

Reading Coordinator

Marian Sainsbury, NFER

Reading Assessment Consultant

Patricia Donahue, ETS

PIRLS 2006 Reading Development Group (RDG)

Dominique Lafontaine, *Service de Pedagogie Experimentale, Belgium*
Jan Mejding, *Danish University of Education, Denmark*
Sue Horner, *Qualifications and Curriculum Authority, England*
Renate Valtin, *Abteilung Grundschulpädagogik, Humboldt Universität, Germany*

Galina Zuckerman, *Psychological Institute, Russian Academy of Education, Russian Federation*
Elizabeth Pang and Selene Tan, *Ministry of Education HQ, Singapore*
Karen Wixson, *University of Michigan, United States*

PIRLS Questionnaire Development Group (QDG)

Meng Hong Wei, *The China National Institute of Education, China*
Marc Colmant, *Ministere de l'Education Nationale, France*
Knut Schwippert, *University of Hamburg, Institute for Comparative & Multicultural Education, Germany*
Gabriella Pavan de Gregorio, *Instituto Nazionale per la Valuatazione del Sistema Dell'Istruzione, Italy*
Bojana Naceva, *Bureau for Development of Education, Republic of Macedonia*
Mieke van Diepen, *Expertisecentrum Nederlands, Netherlands*
Ragnar Gees Solheim, *National Center for Reading Education and Research, Norway*
Larry Ogle, *National Center for Education Statistics, United States*

PIRLS 2006 National Research Coordinators (NRCs)

Austria

Günter Haider
Birgit Suchań
Austrian IEA Research Centre, Universität Salzburg

Belgium

Flemish

Jan Van Damme
Katholieke Universiteit Leuven

French

Annette Lafontaine
Université de Liège

Bulgaria

Tatyana Angelova
Felianka Kaftandjieva (through 2004)
University of Sofia

Canada

Alberta

Ping Yang
Learner Assessment Branch, Alberta Education

British Columbia

Diane Lalancette
Exams & Assessment Policy

Nova Scotia

Marthe Craig
Evaluation Coordinator, Evaluation Services

Ontario

Michael Kozlow
Francine Jaques (through 2004)
Education Quality and Accountability Office

Québec

Serge Baillargeon
Ministère de l'Éducation

Chinese Taipei

Hwawei Ko
Graduate Institute of Learning and Instruction
National Central University

Denmark

Jan Mejding
The Danish University of Education

England

Liz Twist
National Foundation for Educational Research in England and Wales

France

Marc Colmant
Ministère de l'Éducation Nationale

Georgia

Maia Miminoshvili
National Assessment and Examinations Center

Germany

Wilfried Bos
Sabine Hornberg
Institut fuer Schulentwicklungsfurschung
University of Dortmund

Hong Kong

Tse Shek-Kam
The University of Hong Kong

Hungary

Ildiko Balazsi
Péter Balkányi
Annamária Szazó Rábai (through 2004)
National Institute of Public Education
Centre for Evaluation Studies

Iceland

Brynhildur Scheving Thorsteinsson
Institute for Educational Research

Indonesia

Burhanuddin Tola
Center for Educational Assessment
Bahrul Hayat (through 2004)
Ministry of National Education

Iran, Islamic Republic of

Abdol'azim Karimi
Institute for Educational Research

Israel

Elite Olshtain
Hebrew University
Ruth Zuzovsky
Tel Aviv University

Italy

Silvana Serra
Gabriella Pavan de Gregorio (through 2005)
Instituto Nazionale per la Valuatazione del Sistema Dell'Istruzione

Kuwait

Abdul Ghani Al-Bazzaz
Ministry of Education

Latvia

Antra Ozola
University of Latvia

Lithuania

Aiste Elijio
Ministry of Education and Science

Luxembourg

Pierre Reding
Martin Frieberg
Ministère de l'Éducation Nationale

Macedonia, Republic of

Tanja Andonova
Pedagogical Institute of Macedonia
Bojana Naceva (through 2006)
Bureau for Development of Education

Moldova, Republic of

Ilie Nasu
Ministry of Education and Science

Morocco

Mohammed Sassi
Departement de l'Évaluation Nationale

Netherlands

Andrea Netten
Mieke van Diepen (through 2004)
Expertisecentrum Nederlands

New Zealand

Megan Chamberlain
Ministry of Education

Norway

Ragnar Gees Solheim
Victor van Daal
Finn-Egil Toennessen (through 2005)
National Centre for Reading,
Education and Reading Research
University of Stavanger

Poland

Krzysztof Konarzewski
Institute of Psychology
Polish Academy of Science

Qatar

Abdessalem Buslama
Marcus Broer (through 2006)
Evaluation Institute
Supreme Education Council
Office of Student Assessment

Romania

Gabriela Noveanu
Institute for Educational Sciences
Evaluation and Forecasting Division

Russian Federation

Galina Kovalyova
The Russian Academy of Education

Scotland

Fiona Fraser
Jo MacDonald (through 2003)
Scottish Office, Education Department

Singapore

Wong Look Kwang
New Yi Cheen (through 2005)
Research and Evaluation
Ministry of Education

Slovak Republic

Eva Obrancova
Zuzana Lukackova (through 2004)
SPU—National Institute for Education

Slovenia

Marjeta Doupona-Horvat
Educational Research Institute

South Africa

Sarah Howie
Elsie Venter
University of Pretoria

Spain

Mar Gonzalez Garcia
Flora Gil Traver (through 2005)
Instituto Nacional de Calidad y Evaluacion
del Sistema Educativo

Sweden

Bo Palaszewski
National Agency for Education
Caroline Liberg
Uppsala University

Trinidad and Tobago

Harrilal Seecharan
Mervyn Sambucharan
Division of Educational Research
and Evaluation

United States

Laurence Ogle
National Center for Education Statistics
U.S. Department of Education

TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

ISBN 1-889938-45-9

BOSTON
COLLEGE

pirls.bc.edu