TIMSS 1999 USER GUIDE for the INTERNATIONAL DATABASE

IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade

Edited By:
Eugenio J. Gonzalez and Julie A. Miles

TIMSS REPEAT 1999

ISC

Boston College
Lynch School
of Education

TIMSS 1999

User Guide for the International Database

Edited by
Eugenio J. Gonzalez
Julie A. Miles

with contributions by
Pierre Foy
Robert Garden
Kelvin D. Gregory
Dirk Hastedt
Edward Kulick
Marc Joncas
Barbara Malak
Michael O. Martin
Ina V.S. Mullis
Oliver Neuschmidt
Kathleen M. O'Connor
Teresa A. Smith
Kentaro Yamamoto

June 2001

TIMSS REPEAT •1999•

ISC
BOSTON COLLEGE

TIMSS 1999 User Guide for the International
Database / Edited by Eugenio J. Gonzalez and
Julie A. Miles

Publisher: International Study Center
Lynch School of Education
Boston College

Library of Congress
Catalog Card Number: 2001091861

ISBN 1-889938-21-1

For more information about TIMSS contact:

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

For information on ordering this report, write to
the above address or call +1-617-552-1600

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

Funding for the international coordination of TIMSS 1999 was provided by the National Center for Education Statistics of the U.S. Department of Education, the U.S. National Science Foundation, the World Bank, and participating countries. Each participating country was responsible for funding national project costs and implementing TIMSS 1999 in accordance with the international procedures.

Printed and bound in the United States.

Table of Contents

1 Overview of TIMSS and the 1999 International Database User Guide

2 TIMSS Test Instruments and Booklet Design

3 TIMSS Background Questionnaires

4 Data Collection, Scoring, Materials Processing, and Database Creation

5 Sampling and Sampling Weights

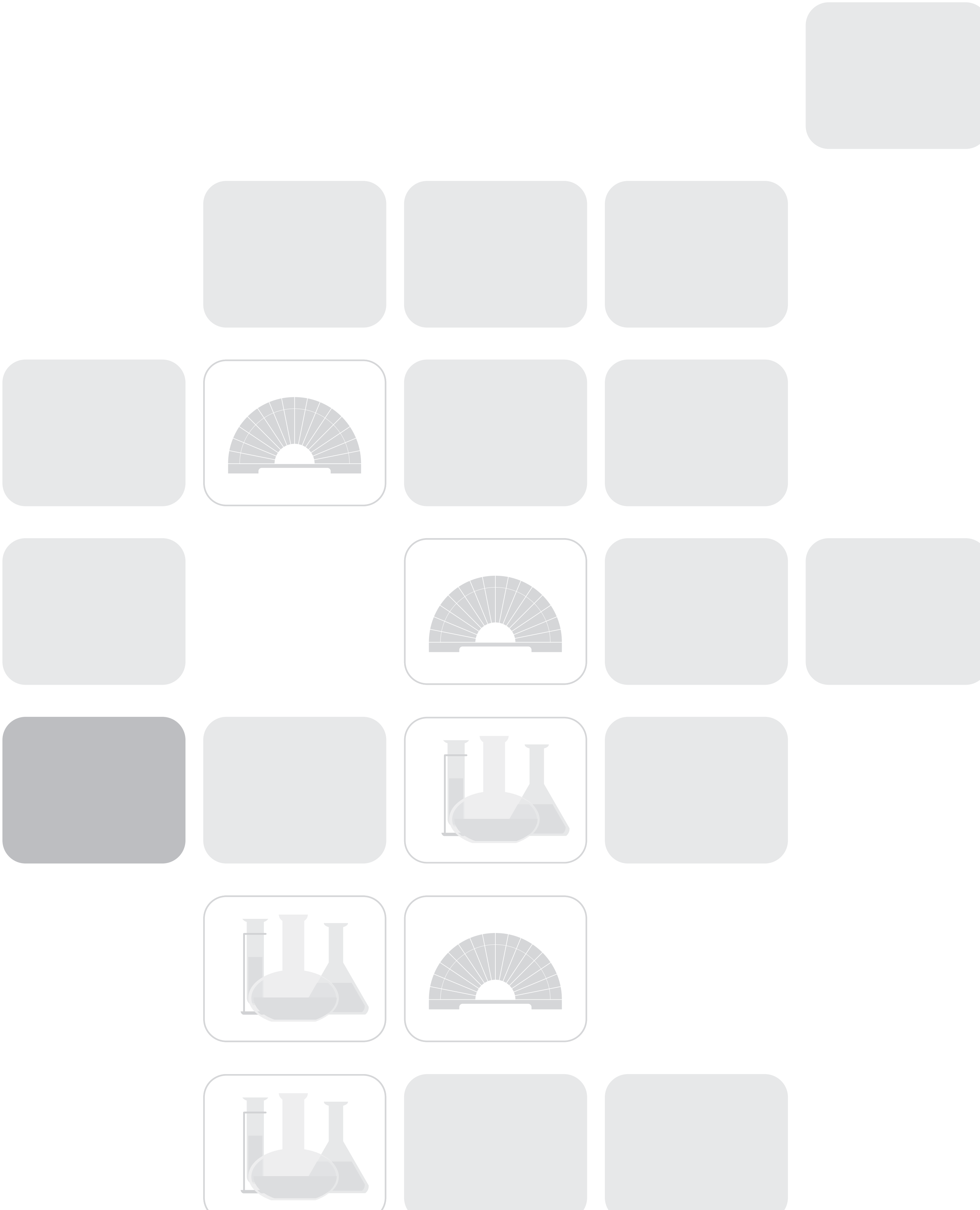
6 Scaling Methodology and Achievement Scores

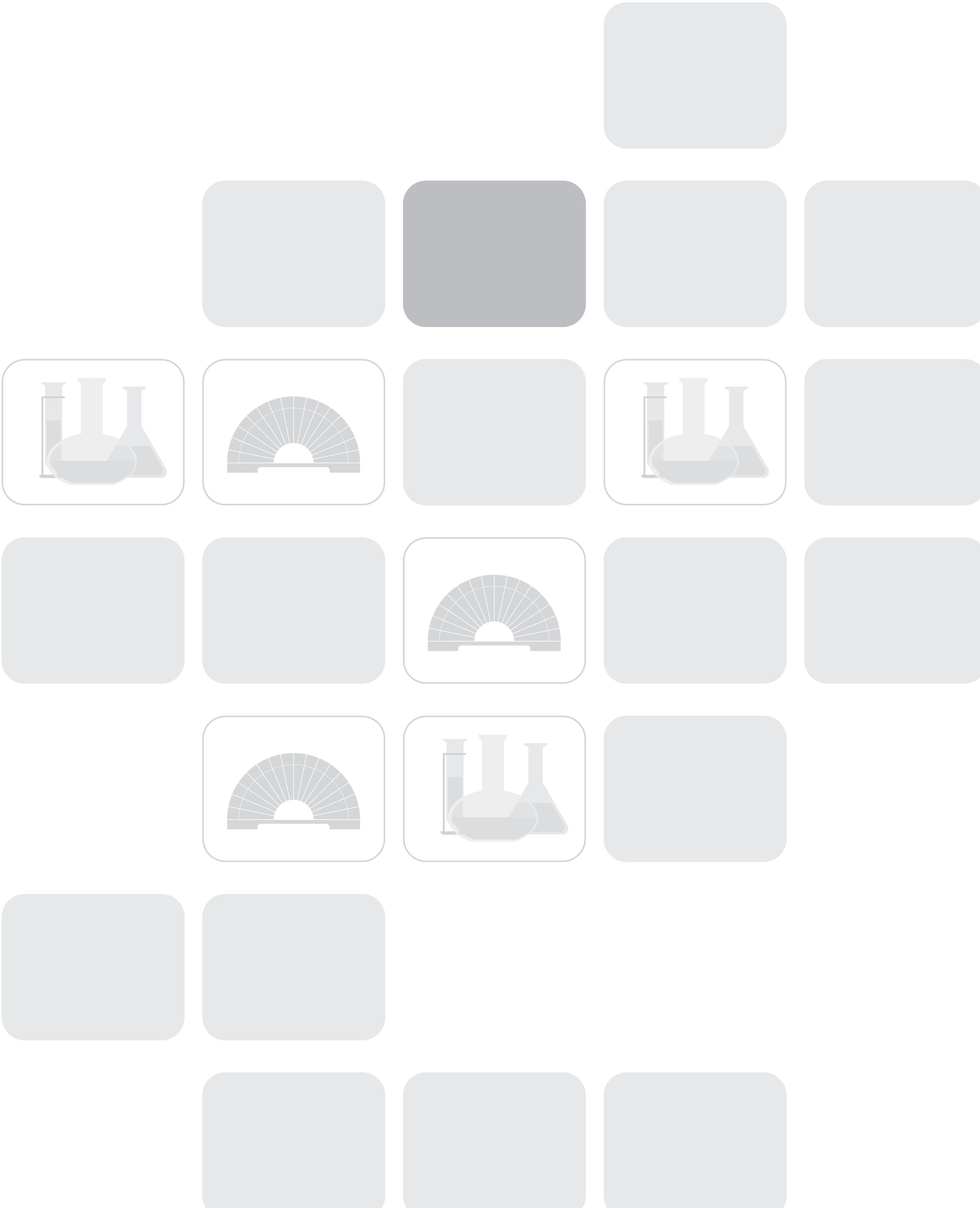
7 Estimating Sampling and Imputation Variance

8 Content and Format of Database Files

9 Performing Analyses with the TIMSS Data Using SPSS

10 Performing Analyses with the TIMSS Data Using SAS





1 Overview of TIMSS and the 1999 International Database User Guide

1.1 Overview

This User Guide accompanies the TIMSS 1999 International Database. The database contains achievement data and student, teacher, and school background data collected in the 38 countries that participated in TIMSS 1999. It also contains the data for the 42 countries that took part in the 1995 Population 2[1] assessment. In total, the TIMSS 1999 database contains responses of nearly 500,000 students, over 50,000 mathematics and science teachers, and over 12,000 school principals collected between 1995 and 1999. Each country gave the IEA permission to release its national data.

The TIMSS 1999 International Database contains the following for each country for which internationally comparable data are available.

- Students' responses to cognitive mathematics and science items
- Students' responses to the background questionnaire
- Teachers' responses to the background questionnaire
- Principals' responses to the background questionnaire
- Test-curriculum matching analysis data
- Data almanacs

Given the size and complexity of TIMSS and the psychometric innovations employed, the TIMSS database is extremely large and complex. Every effort has been made to organize the database and provide adequate documentation so that researchers can access the database for secondary analysis.

This guide describes different aspects of TIMSS, including the data collection instruments, sample design, scaling, and data collection procedures, as this information is needed to use the data. This user guide also documents the content and format of the data files in the International Database and provides example analyses. Appropriate use of the various files and variables, as well as special considerations arising from the complex design, are described. There are three supplements to the User Guide, provided separately, containing copies of the TIMSS international background questionnaires, documentation of national adaptations of the international background questionnaire items, and documentation of derived variables reported in the international reports.

▼ ▼ ▼

[1] This corresponds to 7th and 8th grades in most countries.

This chapter of the User Guide provides an overview of TIMSS, briefly describes the contents of the database, and describes the contents of this User Guide. Further and more detailed information is available from the TIMSS series of technical reports cited at the end of this chapter.

Exhibit 1.1 Countries Participating in TIMSS 1999 and 1995

Country	TIMSS 1999 Grade 8	TIMSS 1995 (Grade 7 or 8)
Australia	✿	✿
Austria		✿
Belgium (Flemish)	✿	✿
Belgium (French)		✿
Bulgaria	✿	✿
Canada	✿	✿
Chile	✿	
Chinese Taipei	✿	
Colombia		✿
Cyprus	✿	✿
Czech Republic	✿	✿
Denmark		✿
England	✿	✿
Finland	✿	
France		✿
Germany		✿
Greece		✿
Hong Kong, SAR	✿	✿
Hungary	✿	✿
Iceland		✿
Indonesia	✿	
Iran, Islamic Rep.	✿	✿
Ireland		✿
Israel	✿	✿
Italy	✿	✿
Japan	✿	✿
Jordan	✿	
Korea, Rep. Of	✿	✿

Country	TIMSS 1999 Grade 8	TIMSS 1995 (Grade 7 or 8)
Kuwait		✿
Latvia	✿	✿
Lithuania	✿	✿
Macedonia, Rep. of	✿	
Malaysia	✿	
Moldova	✿	
Morocco	✿	
Netherlands	✿	✿
New Zealand	✿	✿
Norway		✿
Philippines[1]	✿	
Portugal		✿
Romania	✿	✿
Russian Federation	✿	✿
Scotland		✿
Singapore	✿	✿
Slovak Republic	✿	✿
Slovenia	✿	✿
South Africa	✿	✿
Spain		✿
Sweden		✿
Switzerland		✿
Thailand	✿	✿
Tunisia	✿	
Turkey	✿	
United States	✿	✿

1 In 1995, the Philippines had unapproved sampling procedures at the school level and sampling weights could not be computed. Therefore, data for the Philippines are unweighted. This is effectively implemented by assigning a weight of 1 to all students in the sample for the Philippines.

1.2 Overview of TIMSS

TIMSS 1999 represents the continuation of a long series of studies conducted by the International Association for the Evaluation of Educational Achievement (IEA). Since its inception in 1959, the IEA has conducted more than 15 studies of cross-national achievement in the curricular areas of mathematics, science, language, civics, and reading. The IEA conducted its First International Science Study (FISS) in 1970-71 and the Second International Science Study (SISS) in 1983-84. The First and Second International Mathematics Studies (FIMS and SIMS) took place in 1964 and 1980-82, respectively.

The Third International Mathematics and Science Study (TIMSS), conducted in 1994-1995, was the largest and most complex IEA study to date, and included both mathematics and science at third and fourth grades, seventh and eighth grades, and the final year of secondary school.

Traditionally, IEA studies have worked toward gaining more in-depth understanding of how various factors contribute to the overall outcomes of schooling. Particular emphasis has been given to refining our understanding of students' opportunity to learn, as this opportunity becomes successively defined and implemented by curricular and instructional practices. In an effort to extend what had been learned from previous studies and provide contextual and explanatory information, TIMSS expanded beyond the already substantial task of measuring achievement in two subject areas by including a thorough investigation of curriculum and how it is delivered in classrooms around the world. In addition, extending the work of previous IEA studies, TIMSS included a performance assessment in the 1995 assessment.

In 1998-1999, TIMSS again assessed eighth-grade students in both mathematics and science to measure trends in student achievement since 1995. TIMSS 1999 is also known as TIMSS-Repeat or TIMSS-R. TIMSS plans to assess mathematics and science again in 2003, this time at both 4th and 8th grades.

The results of TIMSS 1999 were published in two companion volumes, *TIMSS 1999 International Mathematics Report* (Mullis, Martin, Gonzalez, Gregory, Garden, O'Connor, Chrostowski, and Smith, 2000) and *TIMSS 1999 International Science Report* (Martin, Mullis, Gonzalez, et al., 2000). An additional volume, the *TIMSS 1999 Technical Report* (Martin, Gregory, and Stemler, 2000) describes the technical aspects of the study and summarizes the main activities involved in the development of the data collection instruments, the data collection itself, and the analysis and reporting of the data.

Participants in TIMSS

Of the 42 countries that participated in TIMSS[2] at the eighth grade in 1995, 26 availed themselves of the opportunity to measure changes in the achievement of their students by also taking part in TIMSS 1999. Twelve additional countries participated in 1999, for a total of 38 countries. Exhibit 1.1 presents the list of countries that participated in TIMSS and whose data is contained in the TIMSS 1999 database. Of the 38 countries taking part in 1999, 19 had also participated in 1995 at the fourth grade. Since fourth-grade students in 1995 were in eighth grade in 1999, these countries can compare their eighth-grade performance with their performance at the fourth grade, as well as with the eighth-grade performance of students in other countries. Fourth grade data are not included in this database.

The Student Population

In 1995 one of the TIMSS target populations was students enrolled in the two adjacent grades that contained the largest proportion of 13-year-old students at the time of testing, corresponding to seventh- and eighth-grade students in most countries. TIMSS in 1999 used the same definition to identify the target grades, but assessed students in the upper of the two grades only, the eighth grade in most countries.

The TIMSS 1999 database contains data for the 42 countries that participated in 1995 at the seventh and eighth-grade level (Population 2) and the 38 countries that participated in 1999 at the eighth-grade level.

Survey Administration Dates

Since school systems in countries in the Northern and Southern Hemispheres do not have the same school year, TIMSS 1999 had to operate on two schedules. The Southern Hemisphere countries administered the survey from September to November 1998, while the Northern Hemisphere countries did so from February to May 1999. In 1995, Southern Hemisphere countries administered the survey from September to November 1994, while the Northern Hemisphere countries did so from February to May 1995.

The TIMSS Assessment Framework

IEA studies have the central aim of measuring student achievement in school subjects, with a view to learning more about the nature and extent of achievement and the context in which it occurs. The goal is to isolate the factors directly relating

▼ ▼ ▼

2 Results of 41 countries are reported in the TIMSS 1995 international reports; Italy also completed the 1995 testing, but too late to be included. It is treated as a 1995 country in this report and included in all trend exhibits in the 1999 international reports.

to student learning that can be manipulated through policy changes in, for example, curricular emphasis, allocation of resources, or instructional practices. Clearly, an adequate understanding of the influences on student learning can come only from careful study of the nature of student achievement and the characteristics of the learners themselves, the curricula they follow, the teaching methods of their teachers, and the resources in their classrooms and their schools.

The designers of TIMSS in 1995 chose to focus on curriculum as a broad explanatory factor underlying student achievement (Robitaille and Garden, 1996). From that perspective, curriculum was considered to be a three-strand model: what society would like to see taught (the intended curriculum), what is actually taught (the implemented curriculum), and what the students learn (the attained curriculum). This view was first conceptualized for the IEA's Second International Mathematics Study (Travers and Westbury, 1989).

The three strands of the curriculum bring together three major influences on student achievement. The intended curriculum states society's goals for teaching and learning. These goals reflect the ideals and traditions of the greater society and are constrained by the resources of the education system. The implemented curriculum is what is taught in the classroom. Although presumably inspired by the intended curriculum, actual classroom events are usually determined in large part by the teacher, whose behavior may be greatly influenced by his or her own education, training, and experience, by the nature and organizational structure of the school, by interaction with teaching colleagues, and by the composition of the student body. The attained curriculum is what the students actually learn. Student achievement depends partly on the implemented curriculum and its social and educational context, and to a large extent on the characteristics of individual students, including ability, attitude, interests, and effort.

Since TIMSS 1999 essentially replicated the eighth-grade part of the 1995 study, much of the conceptual underpinning of the 1999 study was derived from the three-strand model of curriculum. The organization and coverage of the intended curriculum were investigated through curriculum questionnaires that were completed by National Research Coordinators (NRCs) and their curriculum advisors. Although more modest in scope than the extensive curriculum analysis component of the 1995 study (Schmidt et al., 1997a; 1997b), the TIMSS 1999 questionnaires yielded valuable information on the curricular intentions of participating countries. These data are summarized in Chapter 5 and Reference Chapter 2 of the TIMSS 1999 international reports.

Data on the implemented curriculum were collected as part of the TIMSS 1999 survey of student achievement. Questionnaires completed by the mathematics and science teachers of the students in the survey, and by the principals of their schools, provided information about the topics in mathematics and science that were taught, the instructional methods used in the classroom, the organizational structures that supported teaching, and the factors that were seen to facilitate or inhibit teaching and learning.

The student achievement survey provided data for the study of the attained curriculum. The wide-ranging mathematics and science tests that were administered to nationally representative samples of students provided not only a sound basis for international comparisons of student achievement, but a rich resource for the study of the attained curriculum in each country. Information about students' characteristics, attitudes, beliefs, and experiences, was collected from each participating student. This information was used to identify the student characteristics associated with learning and provide a context for the study of the attained curriculum.

1.3 Developing the TIMSS 1999 Achievement Tests

The TIMSS curriculum framework underlying the mathematics and science tests was developed for TIMSS in 1995 by groups of mathematics and science educators with input from the TIMSS NRCs. As shown in Exhibit 1.2, the curriculum framework contains three dimensions or aspects. The *content* aspect represents the subject matter content of school mathematics and science. The *performance expectations* aspect describes, in a non-hierarchical way, the many kinds of performance or behavior that might be expected of students in school mathematics and science. The *perspectives* aspect focuses on the development of students' attitudes, interest, and motivation in the subjects. Because the framework was developed for the entire span of curricula from the beginning of schooling through the completion of secondary school, not all aspects are reflected in the eighth-grade TIMSS assessment.[3] Working within the framework, mathematics and science test specifications for TIMSS in 1995 included items representing a wide range of mathematics and science topics and eliciting a range of skills from the students. The 1995 tests were developed through an international consensus process involving input from experts in mathematics, science, and measurement, ensuring that the tests reflected current thinking and priorities in mathematics and science education.

▼ ▼ ▼

[3] The complete TIMSS curriculum framework can be found in Robitialle et al. (1993).

Exhibit 1.2 The Three Aspects and Major Categories of the TIMSS 1999 Mathematics and Science Test Items

Subject	Content	Performance Expectation	Perspectives
Mathematics	Numbers Measurement Geometry Proportionality Functions, Relations, and Equations Data Representation Probability and Statistics Elementary Analysis, Validation and Structure	Knowing Using Routine Procedures Investigating and Problem Solving Mathematical Reasoning Communicating	Attitudes Careers Participation Increasing Interest Habits of Mind
Science	Earth Science Life Sciences Physical Science History of Science and Technology Environmental and Resource Issues Nature of Science Science and Other Disciplines	Understanding Theorizing, Analyzing, and Solving Problems Using Tools, Routine Procedures and Science Processes Investigating the Natural World Communicating	Attitudes Careers Increasing Interest Safety Habits of Mind

About one-third of the items in the TIMSS 1995 assessment were kept secure to measure trends over time; the remaining items were released for public use. An essential part of the development of the 1999 assessment, therefore, was to replace the released items with items of similar content, format, and difficulty. With the assistance of the Science and Mathematics Item Replacement Committee, a group of internationally prominent mathematics and science educators nominated by participating countries to advise on subject matter issues in the assessment, over 300 mathematics and science items were developed as potential replacements. After an extensive process of review and field testing, 212 items (114 in mathematics and 98 in science) were selected as replacements and were included in the 1999 mathematics assessment.

Exhibit 1.3 presents the five content areas included in the TIMSS 1999 mathematics test and the six content areas in the TIMSS 1999 science test, together with the number of items and score points in each area. About one-fourth of the items were in the free-response format, requiring students to generate and write their own answers. Designed to take about one-third of students' test time, some free-response questions asked for short answers while others required extended responses with students showing their work or providing explanations for their answers. The remaining questions were in the multiple-choice format. Correct answers to most questions were worth one point. Consistent with longer response times for the constructed-response questions, however, responses to some of these questions (particularly those requiring extended responses) were evaluated for partial credit, with a fully correct answer being awarded two points. The number of score points available for analysis thus exceeds the number of items.

Exhibit 1.3 Number of Test Items and Score Points by Reporting Category in TIMSS 1999

Reporting Category	Number of Items	Score Points
Mathematics		
Fractions and Number Sense	61	62
Measurement	24	26
Data Representation, Analysis and Probability	21	22
Geometry	21	21
Algebra	35	38
Total	162	169
Science		
Earth Science	22	23
Life Science	40	42
Physics	39	39
Chemistry	20	22
Environmental and Resource Issues	13	14
Scientific Inquiry and the Nature of Science	12	13
Total	146	153

TIMSS Test Design

Not all of the students in the TIMSS assessment responded to all of the mathematics and science items. To ensure broad subject matter coverage without overburdening students, TIMSS used a rotated design that included both the mathematics and science items (Adams and Gonzalez, 1996). Thus, the same students were tested in both mathematics and science. As in 1995, the 1999 assessment consisted of eight booklets, each requiring 90 minutes of response time. Each participating student was assigned one booklet only. In accordance with the design, the mathematics and science items were assembled into 26 clusters (labeled A through Z). The secure trend items were in clusters A through H, and items replacing the released 1995 items were in clusters I through Z.

Background Questionnaires

In 1999, TIMSS administered a broad array of questionnaires to collect data on the educational contexts for student achievement. NRCs, with the assistance of their curriculum experts, provided detailed information on the organization, content coverage and emphases of the mathematics and science curricula.

The students who were tested answered questions pertaining to their attitude towards mathematics and science, their academic self-concept, classroom activities, home background, and out-of-school activities. A special version of the student questionnaire was prepared for countries where earth science, physics, chemistry, and biology were taught as separate science subjects. Although not strictly related to the question of what students have learned in mathematics or science, general characteristics of pupils can be important correlates for understanding educational processes and attainments. Therefore, students also provided general home and demographic information.

The mathematics and science teachers of sampled students each completed a teacher questionnaire comprised of two sections. The first section covered general background information on preparation, training, and experience, about how teachers spend their time in school, and probed their views on mathematics and science. The second section related to instructional practices in the class or classes selected for the TIMSS 1999 testing. To obtain information about the implemented curriculum, teachers were asked how many periods the class spent on a range of mathematics and science topics, and about the instructional strategies used in the class, including the use of calculators and computers. Teachers also responded to questions about teaching emphasis on the topics in the curriculum frameworks.

The heads of schools responded to questions about school staffing and resources, mathematics and science course offerings, and support for teachers.

A copy of the different background questionnaires is included in Supplement 1 of this User Guide.

Translation and Verification

The TIMSS 1999 instruments were prepared in English and translated into 33 languages (see Exhibit 1.4). Ten of the 38 countries collected data in two languages. In addition, the international versions sometimes needed to be modified for cultural reasons, even in the nine countries that tested in English. This process represented an enormous effort for the national centers, with many checks along the way. The translation effort included developing explicit guidelines for translation and cultural adaptation; translation of the instruments by the national centers in accordance with the guidelines, using two or more independent translators; consultation with subject matter experts on cultural adaptations to ensure that the meaning and difficulty of items did not change; verification of translation quality by professional translators from an independent translation company; corrections by the national centers in accordance with the suggestions made; verification by the International Study Center that corrections were made; and a series of statistical checks after the testing to detect items that did not perform comparably across countries.

Exhibit 1.4 Language of Testing in Each Country – 1999 Countries Only[1]

Country	Language(s) of Test	Country	Language(s) of Test
Australia	English	**Latvia**	Latvian
Belgium (Flemish)	Flemish	**Lithuania**	Lithuanian
Bulgaria	Bulgarian	**Macedonia**	Macedonian and Albanian
Canada	English and French	**Malaysia**	Malay
Chile	Spanish	**Moldova**	Moldavian and Russian
Chinese Taipei	Chinese	**Morocco**	Arabic
Cyprus	Greek	**Netherlands**	Dutch
Czech Republic	Czech	**New Zealand**	English
England	English	**Philippines**	English and Filipino
Finland	Finnish and Swedish	**Romania**	Romanian and Hungarian
Hong Kong	Chinese, English, or Both	**Russian Federation**	Russian
Hungary	Hungarian	**Singapore**	English
Indonesia	Indonesian	**Slovak Republic**	Slovak
Iran, Islamic Rep.	Farsi	**Slovenia**	Slovenian
Israel	Hebrew and Arabic	**South Africa**	English and Afrikaans
Italy*	Italian and German	**Thailand**	Thai
Japan	Japanese	**Tunisia****	Arabic
Jordan	Arabic	**Turkey**	Turkish
Korea, Rep.	Korean	**United States**	English

1 For information regarding the language of testing in each country in the TIMSS 1995 assessment refer to the 1995 User Guide.

* Italy did not have the German version of the items and student questionnaire verified. Less than 1% of the population took the assessment and student questionnaire in German.

** Tunisia also translated the teacher questionnaires into French; student tests and questionnaires were translated only into Arabic.

Data Collection

Each participating country was responsible for carrying out all aspects of the data collection, using standardized procedures developed for the study. Training manuals were created for School Coordinators and Test Administrators that explained procedures for receipt and distribution of materials as well as for the activities related to the testing sessions. These manuals included procedures for test security, standardized scripts to regulate directions and timing, rules for answering students' questions, and steps to ensure that identification on the test booklets and questionnaires corresponded to the information on the forms used to track students.

Each country was responsible for conducting quality control procedures and describing this effort in the NRC's report documenting procedures used in the study. In addition, the International Study Center considered it essential to monitor compliance with the standardized procedures. NRCs were asked to nominate one or more persons unconnected with their national center, such as retired school teachers, to serve as quality control monitors for their countries. The International Study Center developed manuals for the monitors and briefed them in two-day training sessions about TIMSS, the responsibilities of the national centers in conducting the study, and their own roles and responsibilities.

The quality control monitors interviewed the NRCs about data collection plans and procedures. They also visited a sample of 15 schools where they observed testing sessions and interviewed School Coordinators.

Scoring the Free-Response Items

Because about one-third of the test time was devoted to free-response items, TIMSS needed to develop procedures for reliably evaluating student responses within and across countries. Scoring used two-digit codes with rubrics specific to each item. The first digit designates the correctness level of the response. The second digit, combined with the first, represents a diagnostic code identifying specific types of approaches, strategies, or common errors and misconceptions. Analyses of responses based on the second digit should provide insight into ways to help students better understand mathematics and science concepts and problem-solving approaches.

To ensure reliable scoring procedures based on the TIMSS rubrics, detailed guides were prepared containing the rubrics and explanations of how to use them, together with example student responses for each rubric. These guides, along with training packets containing extensive examples of student responses for practice in applying the rubrics, served as a basis for intensive training in scoring the free-response items. The training sessions were designed to help representatives of national centers who would then be responsible for training personnel in their countries to apply the two digit codes reliably. The scoring rubrics are further discussed in Chapter 2 of this User Guide.

Data Processing

To ensure the availability of comparable, high-quality data for analysis, TIMSS took rigorous quality control steps to create the International Database. TIMSS prepared manuals and software for countries to use in entering their data, so that the information would be in a standardized international format before being forwarded to the IEA Data Processing Center in Hamburg for inclusion in the International Database. Upon arrival at the Data Processing Center, the data underwent an exhaustive cleaning process. This involved several iterative steps and procedures designed to identify, document, and correct deviations from the international instruments, file structures, and coding schemes. The process also emphasized consistency of information within national data sets and appropriate linking among the many student, teacher, and school data files.

Throughout the process, the data were checked by the IEA Data Processing Center, the International Study Center, and the national centers. The national centers were contacted regularly and given multiple opportunities to review the data for their countries. In conjunction with the IEA Data Processing Center, the International Study Center reviewed item statistics for each cognitive item in each country to identify poorly performing items. Generally, the items had very good psychometric properties in all countries. Ocassionally an item had poor statistics for a country (negative point-biserials for the key, large item-by-country interactions, and statistics indicating lack of fit with the model) these almost invariably were due to translation, adaptation, or printing deviations. The data included in this database are the result of an exhaustive procedure of data cleaning.

IRT Scaling and Data Analysis

The reporting of the TIMSS achievement data was based primarily on item response theory (IRT) scaling methods. The mathematics and science results were summarized using a family of 2-parameter and 3-parameter IRT models for dichotomously scored items (right or wrong), and 2-parameter generalized partial credit models for items with more than 1 available score point. The IRT scaling method produces a score by averaging the responses of each student to the items in the student's test booklet in a way that takes into account the difficulty and discriminating power of each item. The method used in TIMSS includes refinements that enable reliable scores to be produced even though individual students responded to relatively small subsets of the total item pool.

Achievement scales were produced for five mathematics content areas (fractions and number sense; measurement; data representation, analysis, and probability; geometry; and algebra), as well as for mathematics overall. In science, achievement scales were produced for six content areas (earth science, life science, physics, chemistry, environmental and resource issues, and scientific inquiry and the nature of science), as well as for science overall. Even though IRT scales were created for each of the mathematics and science content areas for the 1999 data, insufficient items were used in both 1995 and in 1999 to establish reliable IRT content area scales for trend purposes.

The IRT method was preferred for developing comparable estimates of performance for all students, since students answered different test items depending upon which of the eight test booklets they received. For a reliable measure of student achievement in both 1999 and 1995, the overall mathematics and science scales were calibrated across years using students from the countries that participated in both years. When all countries participating in 1995 at the eighth grade are treated equally, the TIMSS scale average over those countries is 500 and the standard deviation is 100. The average and standard deviation of the scale scores are arbitrary and do not affect scale interpretation. When the metric of the scale had been established, students from the countries that tested in 1999 but not 1995 were assigned scores based on the new scale.

To allow more accurate estimation of summary statistics for student subpopulations, the TIMSS scaling made use of plausible-value technology, whereby five separate estimates of each student's score were generated on each scale, based on the responses to the items in the student's booklet and the student's background characteristics. The five score estimates are known as "plausible values," and the variability between them encapsulates the uncertainty inherent in the estimation of these scores. Later in this User Guide, examples of how to calculate this uncertainty are presented.

1.4 Management and Operations

Like all previous IEA studies, TIMSS 1999 was a cooperative venture among independent research centers around the world. While country representatives came together to work on instruments and procedures, they were each responsible for conducting TIMSS 1999 in their own country in accordance with the international standards. Each national center provided its own funding and contributed to the support of the international coordination of the study. A study of the scope and magnitude of TIMSS 1999 offered a tremendous operational and logistic challenge. In order to yield comparable data, the achievement survey was replicated in each participating country in a timely and consistent manner. This was the responsibility of the NRC in each country.

The TIMSS International Study Center, located at Boston College in the United States, was responsible for managing all aspects of the design and implementation of the study at the international level.

Several important TIMSS functions, including test and questionnaire development, translation checking, sampling, data processing, and scaling, were conducted by centers around the world, under the direction of the TIMSS International Study Center. In particular, the following centers played important roles in TIMSS 1999:

- The IEA Secretariat, based in Amsterdam, the Netherlands, coordinated the verification of each country's translations and organized the visits of the international quality control monitors.
- The IEA Data Processing Center (DPC), located in Hamburg, Germany, was responsible for checking and processing data and for constructing the International Database. The DPC also worked with Statistics Canada to develop software to facilitate the within-school sampling activities.
- Statistics Canada, located in Ottawa, Canada, was responsible for advising NRCs on their sampling plans, monitoring progress in all aspects of sampling, and computing the sampling weights.
- The Educational Testing Service, located in Princeton, New Jersey, conducted psychometric analyses of the data and was responsible for scaling the achievement data from the main data collections.

The Project Management Team, consisting of the International Study Center Co-Directors and representatives of each of the above organizations, met regularly throughout the study to plan major activities and to monitor progress.

1.5 TIMSS International Reports

The TIMSS 1999 International Database contains the data that were published in 2000, in two reports prepared by the TIMSS International Study Center at Boston College. These reports are:

- Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., Gregory, K.D., Garden, R.A., O'Connor, K.M., Chrostowski, S.J., and Smith, T.A. (2000), *TIMSS 1999 International Science Report, Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*, Chestnut Hill, MA: Boston College.
- Mullis, I.V.S., Martin, M.O. Gonzalez, E.J., Gregory, K.D, Smith, T.A., Chrostowski, S.J., Garden, R.A., and O'Connor, K.M. (2000), *TIMSS 1999 International Mathematics Report, Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*, Chestnut Hill, MA: Boston College.

1.6 Contents of this User Guide

Given the size and complexity of the TIMSS International Database, a description of its contents is also complex. Before trying to use the files contained on the CDs, we recommend that you read through this guide to get an understanding of the study and a sense of the structure and contents of the database. During this first reading, there may be particular sections that you can skim and other sections that you may want to read more carefully. Nonetheless, a preliminary read-through (before actually opening up the files and trying to use them) will help you understand the complexities of the study and the database. When using the files, you will need to follow this guide carefully and refer to the supplements as necessary. The contents of each chapter and the supplements are summarized below.

Chapter 2
TIMSS Test Instruments and Booklet Design

Describes the content and organization of the TIMSS tests, and provides a brief introduction to scoring guides used. The TIMSS item release policy also is described.

Chapter 3
TIMSS Background Questionnaires

Describes the Student, Teacher, and School Background questionnaires administered in TIMSS 1999.

Chapter 4
Data Collection, Materials Processing, Scoring, and Database Creation

Describes the data collection and field administration procedures used in TIMSS, the scoring of the free-response items, data entry procedures, and the creation of the International Database, including the data verification and database restructuring.

Chapter 5
Sampling and Sampling Weights

Describes the sampling design for TIMSS, and the use of sampling weights to obtain proper population estimates, as well as the sampling and weight variables included in the data files.

Chapter 6
Scaling Methodology and Achievement Scores

Provides an overview of the scaling methodology used in TIMSS, including a description of the scaling model and plausible values technology. It also describes the student-level achievement scores that are available in the International Database, including how they were derived and used by TIMSS, and how they can be used.

Chapter 7
Estimating Sampling and Imputation Variance

Describes the procedure for estimating sampling and imputation variance for statistics.

Chapter 8
Content and Format of Database Files

Provides detailed descriptions of the data files included in the database including data files, codebook files, data access programs, and data.

Chapter 9
Performing Analyses with the TIMSS Data: Examples in SPSS

Provides example programs in SPSS for conducting analyses on the TIMSS data, including merging data files and using the SPSS macros to analyze the TIMSS data.

Chapter 10
Performing Analysis with the TIMSS Data: Examples in SAS

Provides example programs in SAS for conducting analyses on the TIMSS data, including merging data files and using the SAS macros to analyze the TIMSS data.

Supplement 1
International Versions of the Background Questionnaires

Contains the international versions of the student, teacher, and school background questionnaires for TIMSS 1999, and tables that map each question to a variable in the TIMSS 1999 database and to analogous variables in the 1995 questionnaires.

Supplement 2
Documentation of National Adaptations of the International Background Questionnaire Items

Contains documentation of national adaptations made to the international versions of the student, teacher, and school questionnaire items. This documentation provides you with a guide to the availability of internationally comparable data for secondary analyses, including information about the availability of 1995 data for trend analyses.

Supplement 3
Documentation of Derived Variables Based on Student, Teacher and School Background Questionnaire Items

Contains variables derived from questions in the student, teacher and school questionnaires that are included in the database. Also contains documentation of additional analyses performed for certain exhibits in the TIMSS 1999 International reports.

1.7 Additional Resources

Although this User Guide is intended to provide you with sufficient information to conduct analyses on the TIMSS data, you may want additional information about TIMSS. Further documentation on the study design, implementation, analysis and results of TIMSS 1995 and TIMSS 1999 can be found in the following publications.

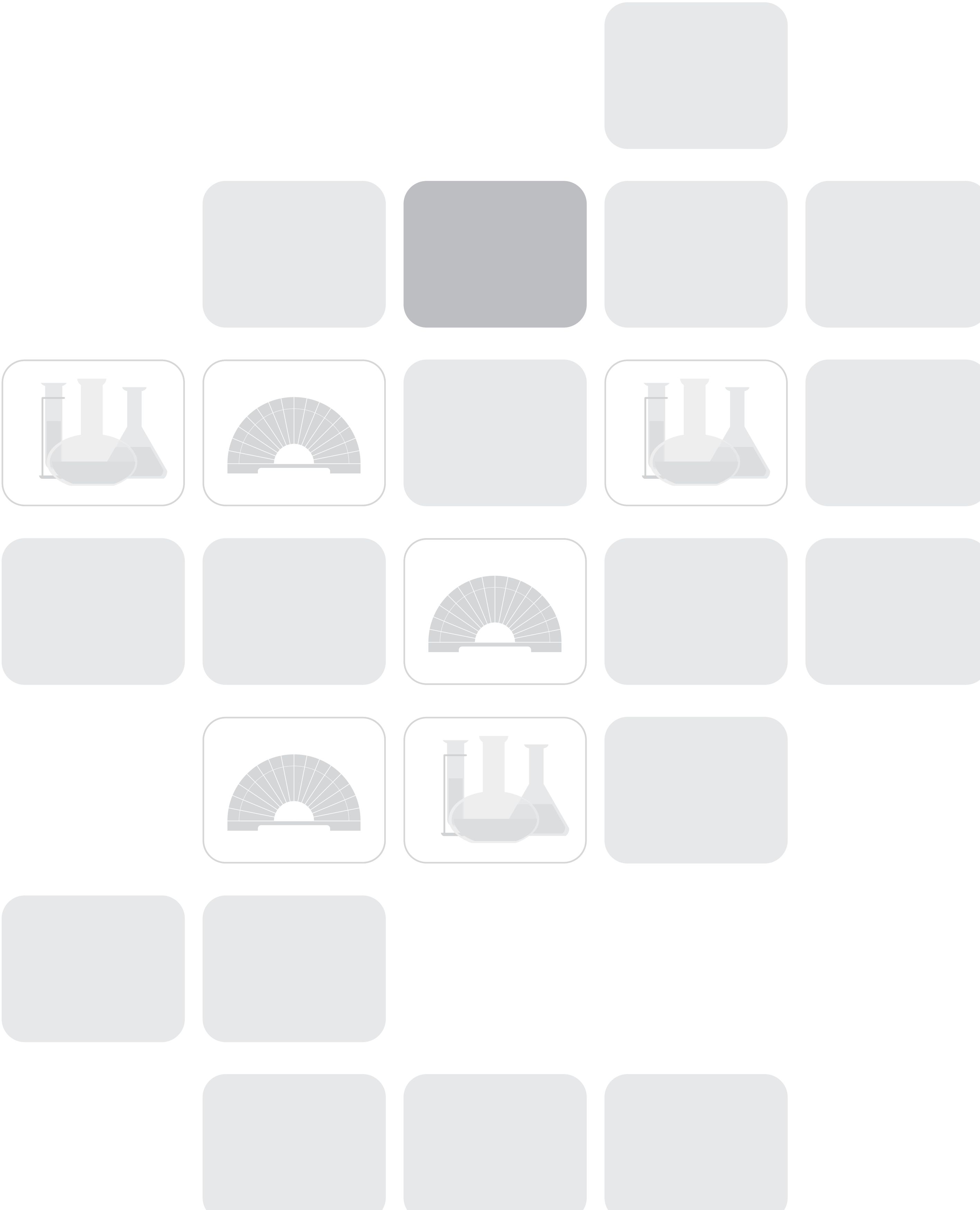
TIMSS 1995 Reports

- Beaton, A.E., Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., Smith, T.A., and Kelly, D.L. (1996), *Science achievement in the middle school years: IEA's Third International Mathematics and Science Study*, Chestnut Hill, MA: Boston College.
- Beaton, A.E., Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Kelly, D.L., and Smith, T.A. (1996), *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study*, Chestnut Hill, MA: Boston College.
- Gonzalez, E.J., and Smith, T.A. (1997), *User guide for the TIMSS international database, Primary and middle school years, 1995 assessment*, Chestnut Hill, MA: Boston College.

- Gonzalez, E.J., Smith, T.A., and Sibberns, H. (1998), *User guide for the TIMSS international database, Final year of secondary school, 1995 assessment*, Chestnut Hill, MA: Boston College.
- Kelly, D.L., Mullis, I.V.S., Martin, M.O. (2000), *Profiles of student achievement in mathematics at the TIMSS international benchmarks: U.S. performance and standards in an international context*, Chestnut Hill, MA: Boston College.
- Martin, M.O., and Kelly, D.L. (1996), *TIMSS technical report, volume I: Design and development*, Chestnut Hill, MA: Boston College.
- Martin, M.O., and Kelly, D.L. (1997), *TIMSS technical report, volume II: Implementation and analysis, Primary and Middle School Years*, Chestnut Hill, MA: Boston College.
- Martin, M.O., and Mullis I.V.S. (1996), *TIMSS: Quality assurance in data collection*, Chestnut Hill, MA: Boston College.
- Martin, M.O., Mullis, I.V.S., Beaton, A.E., Gonzalez, E.J., Smith, T.A., and Kelly, D.L. (1997), *Science achievement in the primary school years: IEA's Third International Mathematics and Science Study*, Chestnut Hill, MA: Boston College.
- Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., Smith, T.A., and Kelly, D.L. (1999), *School contexts for learning and instruction*, Chestnut Hill, MA: Boston College.
- Martin, M.O., Mullis, I.V.S., Gregory, K.D., Hoyle, C. and Shen C. (2000), *Effective schools in science and mathematics*, Chestnut Hill, MA: Boston College.
- Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., and Smith T.A. (1997), *Mathematics achievement in the primary school years: IEA's Third International Mathematics and Science Study*, Chestnut Hill, MA: Boston College.
- Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., and Smith T.A. (1998), *Mathematics and science achievement in the final year of secondary school*, Chestnut Hill, MA: Boston College.
- Mullis, I.V.S., Martin, M.O., Fierros, E.G., Goldberg, A.L., and Stemler, S.E. (2000), *Gender differences in achievement*, Chestnut Hill, MA: Boston College.
- Smith, T.A., Martin, M.O., Mullis, I.V.S., and Kelly, D.L. (2000), *Profiles of student achievement in science at the TIMSS international benchmarks: U.S. performance and standards in an international context*, Chestnut Hill, MA: Boston College.

TIMSS 1999 Reports

- Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., Gregory, K.D., Garden, R.A., O'Connor, K.M., Chrostowski, S.J., and Smith, T.A. (2000), *TIMSS 1999 international science report, Findings from IEA's repeat of the Third International Mathematics and Science Study at the eighth grade*, Chestnut Hill, MA: Boston College.
- Mullis, I.V.S., Martin, M.O. Gonzalez, E.J., Gregory, K.D, Smith, T.A., Chrostowski, S.J., Garden, R.A., and O'Connor, K.M. (2000), *TIMSS 1999 international mathematics report, Findings from IEA's repeat of the Third International Mathematics and Science Study at the eighth grade*, Chestnut Hill, MA: Boston College.
- Martin, M.O., Gregory, K.D., and Stemler, S.E. (2000), *TIMSS 1999 technical report, IEA's repeat of the Third International Mathematics and Science Study at the Eighth Grade*, Chestnut Hill, MA: Boston College.



2 TIMSS Test Instruments and Booklet Design

2.1 Overview

Since the TIMSS 1999 assessment was essentially a replication of the eighth-grade assessment in 1995, it made use of the same assessment and booklet design. However, to provide the public with information about the nature and scope of the TIMSS 1995 achievement tests, almost two-thirds of the items on the tests administered in 1995 were released to the public, with the remaining one-third kept secure as a basis for accurately measuring trends in student achievement from 1995 to 1999. One of the challenges for TIMSS 1999 therefore was to develop tests containing replacement items that were similar in terms of subject matter content and expectations for student performance to those released in 1995, to be used alongside the secure items from 1995. This would provide a reliable and informative assessment of student achievement in mathematics and science in 1999, comparable in scope and coverage to the TIMSS 1995 assessment, while also providing a valid measure of the changes in achievement since 1995.

2.2 The 1995 and 1999 Achievement Tests

In the TIMSS 1995 assessment, mathematics and science items were organized into 26 clusters, labeled A-Z. These clusters were rotated through eight student test booklets, with five or seven clusters in each book, according to the scheme shown in Exhibit 2.1.[1] The same booklet design was used in TIMSS 1999.

In 1995, items in clusters A-H were kept secure for use in the 1999 study, and the remaining 18 clusters (I-Z) were released to the public. The secure clusters A-H were used in TIMSS 1999 exactly as in TIMSS 1995. The 103 mathematics and 87 science items released in 1995 were replaced with similar items. Replacement items assessed the same basic content area and performance expectation and, as nearly as possible, matched the difficulty level of the 1995 items. The same item format was also maintained for the replacement items. Thus, the TIMSS 1999 tests were made to closely resemble those of TIMSS 1995 in structure and content. See Exhibit 2.2 for the number of common and unique items in the 1995 and 1999 assessments.

▼ ▼ ▼

[1] Adams, R.J. & Gonzalez, E.J. (1996), The TIMSS Test Design, in M.O. Martin & D.L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report Volume I: Design and Development*. Chestnut Hill, MA: Boston College.

Exhibit 2.1 Assignment of Item Clusters to Student Test Booklets — TIMSS 1995 and 1999

Cluster Type	Cluster Label	Booklet[1]							
		1	2	3	4	5	6	7	8
Core Cluster (12 minutes) *Mathematics and Science Items-Multiple-Choice*	A	2	2	2	2	2	2	2	2
Focus Clusters (12 minutes)	B	1				5		3	1
Mathematics and Science Items-Multiple-Choice	C	3	1				5		
	D		3	1				5	
	E	5		3	1				
	F		5		3	1			
	G			5		3	1		
	H				5		3	1	
Breadth Clusters (22 minutes)	I	6							
Mathematics and Science Items-Multiple-Choice and Free-Response	J		6						
	K			6					
	L				6				
	M					6			
	N						6		
	O							6	
	P								6
	Q								3
	R								5
Mathematics Free-Response Clusters (10 minutes)	S	4							
	T	7		4					
	U			7		4			
	V					7		4	
Science Free-Response Clusters (10 minutes)	W		4					7	
	X		7		4				
	Y				7		4		
	Z						7		

1 The number within the cell indicates the position of the cluster in the booklet.

Exhibit 2.2 Numbers of Items Common and Unique to TIMSS 1995 and TIMSS 1999

Subject	Items	TIMSS 1995	TIMSS 1999
Mathematics	Unique to TIMSS 1995	111	
	Unique to TIMSS 1999		115
	Common to both TIMSS 1995 and TIMSS 1999	48	
	Total	159	163
	Grand Total for Mathematics	274	
Science	Unique to TIMSS 1995	94	
	Unique to TIMSS 1999		106
	Common to both TIMSS 1995 and TIMSS 1999	48	
	Total	142	154
	Grand Total for Science	248	

2.3 Scoring Guides for Free-Response Items

In TIMSS, both short-answer and extended-response items were scored using two-digit codes with rubrics specific to each item (Lie, Taylor, and Harmon, 1996). The first digit designates the correctness level of the response. The second digit, combined with the first, represents a diagnostic code used to identify specific types of approaches, strategies, or common errors and misconceptions. The general scoring scheme used for a two-point and a one-point item in TIMSS 1995 is shown in Exhibit 2.3.[2]

▼ ▼ ▼

[2] Specific examples of scoring rubrics may be found in the Released Item Sets available on the World Wide Web at http://timss.bc.edu/timss1999i/study.html.

Exhibit 2.3 TIMSS Two-Digit Scoring Scheme for Free-Response Items

Two-Point Item Codes	
Code	Definition
20	fully-correct response; answer category/method #1
21	fully-correct response; answer category/method #2
22	fully-correct response; answer category/method #3
29	fully-correct response; some other method used
10	partially-correct response; answer category/method #1
11	partially-correct response; answer category/method #2
12	partially-correct response; answer category/ method#3
19	partially-correct response; some other method used
70	incorrect response; common misconception/error #1
71	incorrect response; common misconception/error #2
76	incorrect response; information in stem repeated
79	incorrect response; some other error made
90	crossed out/erased, illegible, or impossible to interpret
99	Blank

One-Point Item Codes	
Code	Definition
10	correct response; answer category/method #1
11	correct response; answer category/method #2
12	correct response; answer category/method #3
19	correct response; some other method used
70	incorrect response; common misconception/error #1
71	incorrect response; common misconception/error #2
76	incorrect response; information in stem repeated
79	incorrect response; some other error made
90	crossed out/erased, illegible, or impossible to interpret
99	Blank

In TIMSS 1999, this scoring scheme was retained with minor modifications. The use of code 76 for responses that merely repeated information in the stem of the item was discontinued for TIMSS 1999. Code 90 was also deleted, and responses in this category were coded as 79. For both surveys, the second-digit codes of 7 and 8 were reserved for nationally-defined diagnostic codes used by the national centers to monitor the occurrence of certain common response types that were not already captured with the internationally-defined diagnostic codes. In processing the data for the international database, these country-specific codes were recoded to the "other" response category (second digit 9) at the appropriate score level.

Some free-response items were awarded one point, others two points, some free-response items had more than one part, each worth one or two points. In general, one point was allocated for short-answer items (essentially scored correct or incorrect) that required students to provide a brief response to a question. In mathematics, these questions usually called for a numerical result. In science, the one-point items usually required a short explanation or factual description in one or two sentences. In both subjects, two point items were those judged to demand more than a numerical response or a short written response. In mathematics, students were asked to show their work or explain their methods, and these responses were taken into account in scoring their correctness. In science, the two-point items required an explanation of more than one or two sentences and a demonstration of knowledge of science concepts.

Generalized scoring guides were developed to clarify the types of responses that would merit two points, as compared with those meriting only one point. These generalized scoring guides for mathematics are presented in Exhibit 2.4 and those for science are presented in Exhibit 2.5.

Exhibit 2.4 TIMSS 1999 Mathematics Generalized Scoring Guide

Score Points for Extended-Response Items
2 Points: A two-point response is complete and correct. The response demonstrates a thorough understanding of the mathematical concepts and/or procedures embodied in the task. Indicates that the student has completed the task, showing mathematically sound procedures. Contains clear, complete explanations and/or adequate work when required.
1 Point: A one-point response is only partially correct. The response demonstrates only a partial understanding of the mathematical concepts and/or procedures embodied in the task. Addresses some elements of the task correctly but may be incomplete or contain some procedural or conceptual flaws. May contain a correct solution with incorrect, unrelated, or no work and/or explanation when required. May contain an incorrect solution but applies a mathematically appropriate process.
0 Points: A zero-point response is completely incorrect, irrelevant, or incoherent.
Score Points for Short-Answer Items
1 Point: A one-point response is correct. The response indicates that the student has completed the task correctly.
0 Points: A zero-point response is completely incorrect, irrelevant, or incoherent.

Exhibit 2.5 TIMSS 1999 Science Generalized Scoring Guide

Score Points for Extended-Response Items

2 Points:

A two-point response is complete and correct. The response demonstrates a thorough understanding of the science concepts and/or procedures embodied in the task.

Indicates that the student has completed all aspects of the task, showing the correct application of scientific concepts and/or procedures.

Contains clear, complete explanations and/or adequate work when required.

1 Point:

A one-point response is only partially correct. The response demonstrates only a partial understanding of the scientific concepts and/or procedures embodied in the task.

Addresses some elements of the task correctly but may be incomplete or contain some procedural or conceptual flaws.

May contain a correct answer but with an incomplete explanation.

May contain an incorrect answer but with an explanation indicating a correct understanding of some of the scientific concepts.

0 Points:

A zero-point response is seriously inaccurate or inadequate, irrelevant, or incoherent.

Score Points for Short-Answer Items

1 Point:

A one-point response is correct. The response indicates that the student has completed the task correctly.

0 Points:

A zero-point response is completely incorrect, irrelevant, or incoherent.

2.4 The Achievement Content Areas in 1999

For both mathematics and science, coverage by content area reporting category in TIMSS 1999 was very similar to that in TIMSS 1995. TIMSS 1999 was modified in some respects, however, in order to improve the stability of trend comparisons. In mathematics, TIMSS 1995 had six reporting categories, including *Proportionality*, with only 11 items classified in this content area. For TIMSS 1999 reporting, these items were allocated to other content categories, mainly *Fractions and Number Sense*. In TIMSS 1995, there were five science reporting categories. *Environmental Issues and the Nature of Science* was included as a combined reporting category, with 14 items. For TIMSS 1999, an additional 11 items were developed, permitting the reporting of achievement results separately for the content areas of *Environmental and Resource Issues* and *Scientific Inquiry and the Nature of Science*.

Exhibits 2.6 and 2.7 show for the TIMSS 1999 math and science tests, respectively, the number of items by item type and reporting category with the associated maximum number of score points. Since some of the free-response items were evaluated for partial credit and were awarded a maximum of two points, the number of score points exceeds the number of items.

Exhibit 2.6 Number of TIMSS 1999 Mathematics Test Items and Score Points by Type and Reporting Category

Reporting Category	Item Type				
	Multiple-Choice	Short-Answer	Extended-Response	Number of Items	Score Points
Fractions and Number Sense	47	11	3	61	62
Measurement	15	4	5	24	26
Data Representation, Analysis and Probability	19	1	1	21	22
Geometry	20	1	-	21	21
Algebra	24	4	7	35	38
Total	125	21	16	162	169

Exhibit 2.7 Number of TIMSS 1999 Science Test Items and Score Points by Type and Reporting Category

Reporting Category	Item Type				
	Multiple-Choice	Short-Answer	Extended-Response	Number of Items	Score Points
Earth Science	17	4	1	22	23
Life Science	28	7	5	40	42
Physics	28	11	-	39	39
Chemistry	15	2	3	20	22
Environmental and Resource Issues	7	2	4	13	14
Scientific Inquiry and the Nature of Science	9	2	1	12	13
Total	104	28	14	146	153

The TIMSS 1999 final test contained four more mathematics items and eight more science items than the 1995 test. These extra 12 items were incorporated into the item clusters so that each booklet included one or two of them. Experience with TIMSS 1995 indicated that students would still have ample time to complete the test.

2.5 Item Review

Prior to scaling the achievement items, the item statistics were thoroughly reviewed to identify any items that were functioning differently in different countries. As a result of this review process, a few items were identified as not being internationally comparable in some or all countries and are coded as not administered in the international data files.

2.6 Release Status for TIMSS Test Items and Background Questionnaires

TIMSS releases a large proportion of it items for public use as a matter of course. In 1995, items in clusters I through Z were classified as public release and are available to secondary users (TIMSS, 1996a, 1996b, 1997a, 1997b). This represents two-thirds of the items used in the assessment.

In 1999, one-half of the items were released for public use. The item information sheets contained in the database indicate whether or not the item has been released for public use. The released items are all available on the World Wide Web[3] and electronic copies have been included in the CD accompanying this guide.

▼ ▼ ▼

[3] The IEA TIMSS publications are available on the Internet: http://isc.bc.edu.

3 TIMSS Background Questionnaires

3.1 Overview

TIMSS 1999 was designed to measure trends in student achievement over time by building on the data collected from TIMSS 1995. Consequently, it was important not only to have measures of student achievement that linked the two assessments, but also to have background questionnaires that had much in common. Four background questionnaires were used to gather information at various levels of the educational system: curriculum questionnaires addressed issues of curriculum design and emphasis in mathematics and science; a school questionnaire asked school principals to provide information about school staffing and facilities, as well as curricular and instructional arrangements; teacher questionnaires asked mathematics and science teachers about their backgrounds, attitudes, and teaching activities and approaches; and a student questionnaire asked students about their home backgrounds and attitudes, and their experiences in mathematics and science classes.

The questionnaires developed for TIMSS 1999 retained the parts of the 1995 questionnaires that were found to be most valuable in analysis and reporting and concentrated development efforts on areas needing expansion or refinement. The international version of these questionnaires is included in Supplement 1 of this User Guide.

3.2 Conceptual Framework

The conceptual framework for TIMSS was greatly influenced by IEA's Second International Mathematics Study (SIMS) which focused on the curriculum as a major explanatory factor for international variation in student achievement. In the SIMS model, the curriculum was viewed as having three aspects: the *intended* curriculum, the *implemented* curriculum, and the *attained* curriculum.

- The **intended curriculum** refers to the curricular goals of the education system and the structures established to achieve them.
- The **implemented curriculum** refers to the practices, activities, and institutional arrangements within the school and classroom that are designed to implement the goals of the system.
- The **attained curriculum** refers to the products of schooling – what students actually gained from their educational experience.

Building on this view of the educational process, TIMSS in 1995 and in 1999 sought to assess, through context questionnaires, the factors likely to influence students' learning of mathematics and the sciences at the national (or regional), school, classroom, and student level (Schmidt and Cogan, 1996).

3.3 Research Questions

TIMSS in 1995 posed four general research questions to guide the development of the tests and questionnaires and to provide a focus for the analysis and reporting of results:

- What kinds of mathematics and science are students expected to learn?
- Who provides the instruction?
- How is instruction organized?
- What have students learned?

These questions were also the focus of TIMSS in 1999. The question of what students are expected to learn was addressed using questionnaires that were distributed to mathematics and science curriculum experts in participating countries. The question about the characteristics and preparation of mathematics and science teachers was addressed using questionnaires that were distributed to school principals and teachers. The third question, on instructional approaches to the teaching of mathematics and science, was also addressed through questionnaires to principals, teachers, and students. The fourth question was measured by performance on the TIMSS 1999 achievement tests.

The research questions cast a broad net for exploring associations with achievement in mathematics and science. For example, in attempting to answer the question "Who provides the instruction?" the questionnaires investigated characteristics of the person providing instruction, such as gender, age, years of experience, attitude towards the subject, and time spent preparing lessons. The background questionnaires allow you to examine the most influential characteristics of the people, practices, and policies affecting student achievement.

Curriculum Questionnaires

The TIMSS 1999 study included curriculum questionnaires that were not available for the 1995 survey. These were designed to collect basic information about the organization of the mathematics and science curricula in each country, and about the topics intended to be covered up to the eighth grade. The National Research Coordinator (NRC) in each country was asked to complete one questionnaire on the mathematics curriculum and one on the science curriculum, drawing on the expertise of mathematics and science specialists in the country as necessary.

Each curriculum questionnaire consisted of two parts. The first part sought information about the organization and structure of the curriculum. The second part asked whether a wide range of detailed topics in mathematics and science were in the intended curriculum. In addition, the questionnaires asked what percentage of the eighth-grade student body was taught each of the topics in the intended curriculum.

Because each country submitted only one mathematics curriculum questionnaire and one science curriculum questionnaire, it was possible to conduct follow-up interviews with NRCs to resolve ambiguities and develop a clear understanding of each country's curricula. Several important research questions addressed by the questionnaires were

- Is there a country-level curriculum? If so, how is implementation monitored?
- What is the nature of country-level assessments, if there are any?
- What content is emphasized in the national curriculum?

Because of the complex nature of the data collected with the curriculum questionnaires, these data are not available in the public release of the database. However, the results from these questionnaires are presented in Chapter 5 and Reference Chapter 2 of the TIMSS 1999 international reports.

School Questionnaire

The school questionnaire was completed by the school principal and was designed to elicit information concerning some of the major factors thought to influence student achievement. Several important research questions addressed by the school questionnaire were:

- What staffing and resources are available at each school?
- What are the roles and responsibilities of the teachers and staff?
- How is the mathematics curriculum organized?
- How is the science curriculum organized?
- What is the school social climate?

The TIMSS 1999 version of the school questionnaire was very similar to the 1995 version. Four questions about scheduled time for teachers were removed, since they seemed more appropriate to the teacher questionnaires. Questions on computer availability were revised and extended to include access to the Internet for instructional or educational purposes. Finally, questions dealing with provisions for students of different abilities were extensively revised, since responses to the original questions were not as informative as expected.

The complete contents of the school questionnaire are described further in Exhibit 3.1. A copy of the school questionnaire administered in 1999 is included in Supplement 1 of this User Guide.

Exhibit 3.1 Contents of the TIMSS 1999 School Questionnaire

Question Number	Item Content	Description
1	Community	Situates the school within a community of a specific type.
2-4	Staff	Describes the school's professional full and part-time staff and the percentage of teachers at the school for 5 or more years.
5	Years Students Stay with Teacher	Indicates the number of years students typically stay with the same teacher.
6	Collaboration Policy	Identifies the existence of a school policy promoting teacher cooperation and collaboration.
7	Principal's Time	Indicates the amount of time the school's lead administrator typically spends on particular roles and functions.
8	School Decisions	Identifies who has the responsibility for various decisions for the school.
9	Curriculum Decisions	Identifies the amount of influence various individuals and educational and community groups have on curriculum decisions.
10	Formal Goals Statement	Indicates the existence of school-level curriculum goals for mathematics and science.
11-12	Instructional Resources	Provides a description of the material factors limiting the schools instructional activities.
13	Students in the school	Provides total school enrollment and attendance data.
14	Students in the target grade	Provides target grade enrollment and attendance data, student's enrollment in mathematics and science courses, and typical class sizes.
15	Number of Computers	Provides the number of computers for use by students in the target grade, by teachers, and in total.
16	Internet Access	Identifies whether the school has Internet access and whether the school actively posts any school information on the World Wide Web.
17	Student Behaviors	Provides a description of the frequency with which schools encounter various unacceptable student behaviors.
18	Instructional Time	Indicates the amount of instructional time scheduled for the target grade, according to the school's academic calendar.
19	Instructional Periods	Indicates the existence and length of weekly instructional periods for the target grade.
20	Organization of Mathematics Instruction	Describes the school's provision for students with different ability levels in mathematics (e.g., setting/streaming, tracking, and remedial/enrichment programs).
21	Program Decision Factors in Mathematics	Indicates how important various factors are in assigning students to different educational programs or tracks in mathematics.
22	Organization of Science Instruction	Describes the school's provision for students with different ability levels in science (e.g., setting/streaming, tracking, and remedial/enrichment programs).
23	Program Decision Factors in Science	Indicates how important various factors are in assigning students to different educational programs or tracks in science.
24	Admissions	Describes the basis on which students are admitted to the school.
25	Parental Involvement	Describes the kinds of activities in which parents are expected to participate (e.g., serve as teacher's aids, fundraising).

Teacher Questionnaires

In each participating school, a single mathematics class was sampled for the TIMSS 1999 testing. The mathematics teacher or teachers of the students in that class were asked to complete a questionnaire that sought information on the teacher's background, beliefs, attitudes, educational preparation, and teaching work load, as well as details of the instructional approaches used in teaching mathematics to the sampled class. The science teacher or teachers of the students in that class were asked to complete another questionnaire, which in many respects paralleled that for the mathematics teachers. Although the general background questions were the same for the two versions, questions pertaining to instructional practices, content coverage, classroom organization, teachers' perceptions about teaching, and views of subject matter were tailored to either mathematics or science. Teachers were asked to answer questions related to classroom characteristics, activities, and homework practices, with respect to the specific mathematics and science classes of the sampled TIMSS students.

Like the school questionnaire, the teacher questionnaires were carefully constructed to elicit information on variables thought to be associated with student achievement. Some of the important research questions addressed by the teacher questionnaires were:

- What are the background characteristics of mathematics and science teachers?
- What are teachers' perceptions about mathematics and science?
- How do teachers spend their school-related time?
- How are mathematics and science classes organized?
- What activities do students do in their mathematics and science lessons?
- How are calculators and computers used?
- How much homework are students assigned?
- What assessment and evaluation procedures do teachers use?

Several changes were made in the mathematics and science teacher questionnaires for the 1999 assessment. The first section of the teacher questionnaires dealt with teacher background, experience, attitudes, and teaching load. The 1999 version omitted questions about grades taught, and added several questions on teacher education and preparation for teaching.

The second section of the teacher questionnaires dealt with teaching mathematics or science to the classes sampled for TIMSS 1999 testing. This section was shortened from the 1995 version, mainly by omitting a set of questions on teaching activities in a recent lesson. A lengthy set of questions on the coverage of mathematics and science topics in class was also simplified and shortened considerably. Additions to the teacher questionnaires for 1999 included questions on subject matter emphasis in class, use of computers and the Internet in class, and teacher activities in class. Two sections of the 1995 questionnaires dealing with opportunity to learn and pedagogical approach were judged by NRCs to be too lengthy; these were omitted from the TIMSS 1999 questionnaires.

The complete contents of the mathematics and science teacher questionnaires are described further in Exhibit 3.2. A copy of each of the questionnaires as administered in 1999 is included of Supplement 1 of this User Guide.

Exhibit 3.2 Contents of the TIMSS 1999 Teacher Questionnaires

Section A Question Number	Item Content	Description
1-2	Age and Sex	Identifies teacher's sex and age range.
3	Teaching Experience	Describes the teacher's number of years of teaching experience.
4-5	Instructional Time	Identifies the number of hours per week the teacher devotes to teaching mathematics, science, and other subjects.
6	Administrative Tasks	Identifies the number of hours per week spent on administrative tasks such as student supervision and counseling.
7	Other Teaching-Related Activities	Describes the amount of time teachers are involved in various professional responsibilities outside the formally scheduled school day.
8	Teaching Activities	Describes the total number of hours per week spent on teaching activities.
9	Meet with Other Teachers	Describes the frequency with which teachers collaborate and consult with their colleagues.
10	Teacher's Influence	Describes the amount of influence that teachers perceive they have on various instructional decisions.
11	Being Good at Mathematics /Science	Describes teacher's beliefs about what skills are necessary for students to be good at mathematics/science.
12	Ideas about Mathematics/Science	Describes teacher's beliefs about the nature of mathematics/science and how the subject should be taught.
13	Document Familiarity	Describes teacher's knowledge of curriculum guides, teaching guides, and examination prescriptions (country-specific options).
14	Mathematics/Science Topics Prepared to Teach	Provides an indication of teacher's perceptions of their own preparedness to teach the TIMSS 1999 in-depth topic areas in mathematics or science.
15-18	Formal Education and Teacher Training	Describes the highest level of formal education completed by the teacher, the number of years of teacher training completed, and the teacher's major area of study.
International Options		
19-20	Career Choices	Identifies whether teaching was a first choice and if the teacher would change careers if given the opportunity.
21	Social Appreciation	Describes whether teachers believe society appreciates their work.
22	Student Appreciation	Describes whether teachers believe students appreciates their work.
23	Books in Home	Provides an indicator of teacher's cultural capital.

Exhibit 3.2 Contents of the TIMSS 1999 Teacher Questionnaires (continued)

Section B Question Number	Item Content	Description
1	Target Class	Identifies the number of students in the TIMSS 1999 tested class, by gender.
2	Instructional Emphasis	Identifies the subject matter emphasized most in the target mathematics/science class.
3	Instructional Time	Identifies the number of minutes per week the class is taught.
4	Textbook Use	Identifies whether textbook is used in mathematics/science class as well as the approximate percentage of weekly instructional time that is based on the textbook.
5-7	Calculators	Describes the availability of calculators and how they are used in the target class.
8	Computers	Describes the availability of computers and whether they are used to access the Internet.
9	Planning Lessons	Identifies the extent to which a teacher relies on various sources for planning lessons (e.g., curriculum guides, textbooks, exam specifications).
10	Tasks Students are Asked to Do	Describes the frequency with which teachers ask students various types of questions and ask students to perform various mathematics/science activities during lessons.
11	Student's Work Arrangements	Describes how often students work in various group arrangements.
12	Time Allocation	Describes the percentage of time spent on each of several activities associated with teaching (e.g., homework review, tests).
13	Mathematics/Science Topic Coverage	Indicates the extent of teacher's coverage in target class of mathematics/science topics included in the assessment.
14	Classroom Factors	Identifies the extent to which teachers perceive that various factors limit classroom instructional activities.
15-16	Amount of Homework Assigned	Describes the frequency and amount of homework assigned to the target class.
17-18	Type and Use of Homework	Describes the homework assignments and how the homework is used by the teacher.
19-20	Assessment	Describes the kind and use of various forms of student assessment in the target class.

Student Questionnaire

Each student in the sampled class was asked to complete a student questionnaire. This questionnaire sought information about the student's home background, attitudes and beliefs about mathematics and science, and experiences in mathematics and science classes. As in 1995, two versions of the questionnaire were used:

- General science version: intended for systems where sciences are taught as a single integrated subject
- Separate science subject version: intended for systems where sciences are taught as separate subjects (e.g., biology, chemistry, earth science, and physics)

Countries administered the version of the student questionnaire that was consistent with the way in which science instruction was organized at the target grade. Although the two versions differed with respect to the science questions, the general background and mathematics-related questions were identical across the two forms. In the general science version, science-related questions pertaining to students' attitudes and classroom activities were based on single questions asking about "science," to which students were to respond in terms of the "general or integrated science" course they were taking. In the separate science subject version, several questions were asked about each science subject area, and students were to respond with respect to each science course they were taking. This structure accommodated the diverse systems that participated in TIMSS.

Consistent with the other questionnaires, the student questionnaires were designed to elicit information on some of the major factors thought to influence student achievement. Several important research questions addressed by the student background questionnaires were:

- What educational resources do students have in their homes?
- What are the academic expectations of students, their families, and their friends?
- How do students spend their out-of-school time during the school week?
- How do students perceive success in mathematics and science?
- What are students' attitudes towards mathematics and science?

Five questions from the TIMSS 1995 student questionnaire which were considered to be of lesser importance were moved from the body of the questionnaire to the "international option" section at the end. Questions added to the body of the TIMSS 1999 questionnaire dealt with the following topics:

- Student self-concept in mathematics and science
- Internet access and use for mathematics and science activities
- Instructional activities in mathematics and science class.

The complete contents of the student questionnaires are described further in Exhibit 3.3. A copy of each of the versions of the questionnaires as administered in 1999 is included in Supplement 1 of this User Guide.

Exhibit 3.3 Contents of the TIMSS 1999 Student Questionnaires

Question Number		Item Content	Description
General Version	Separate Science Version		
1-4	1-4	Student Demographics	Provides basic demographic information such as age, sex, language of the home, whether born in country and if not how long he/she has lived in country.
5	5	Academic Activities Outside of School	Provides information on student activities that can affect their academic achievement (e.g., extra lessons, science club).
6	6	Time Spent Outside of School	Provides information about the amount of time student spends on homework and leisure activities on a normal school day.
7	7	Parents' Education	Provides information about the educational level of the student's mother and father. Used as an indicator of the home environment and socioeconomic status.
8	8	Student's Future Educational Plan	Identifies the student's plans for further education.
9	9	Parents' Country of Birth	Provides information regarding immigrant status.
10	10	Books in the Home	Provides information about the number of books in the home. Used as an indicator of the home environment and socioeconomic status.
11	11	Possessions in the Home	Provides information about possessions found in the home (e.g., calculator, computer, study desk, country-specific items). Used as an indicator of academic support in the home environment as well as an indicator of socioeconomic status.
12	12	Mother's Values	Provides information about the student's perception of the degree of importance his/her mother places on academics and other activities. Used as an indicator of the home environment and general academic pressure.
13	13	Student's Behavior in Mathematics Class	Provides a description of typical student behavior during mathematics lessons.
14	14	Peers' Values	Provides information about the student's perception of the degree of importance his/her peers place on academics and other activities. Used as an indicator of peers' values and student's social environment.
15	15	Student's Values	Provides information about the degree of importance the student places on academics and other activities. Used as an indicator of students' values.
16	16	Competence in Mathematics/Science	Provides an indication of student's self-description of academic competence in mathematics and science (specialized version asks about biology, earth science, chemistry, and physics separately).
17	17	Difficulty of Mathematics	Provides a description of student's perception of the difficulty level of mathematics.
18	18	Doing Well in Mathematics	Identifies student's attributions for doing well in mathematics.
19	19-22	Difficulty of Science	Provides a description of student's perception of the difficulty level of science (specialized version asks about biology, earth science, chemistry, and physics separately).
20	23	Doing Well in Science	Identifies student's attributions for doing well in science.

Exhibit 3.3 Contents of the TIMSS 1999 Student Questionnaires (continued)

Question Number		Item Content	Description
General Version	Separate Science Version		
21	24	Liking Mathematics/Science	Identifies how much students like mathematics and science; a key component of student motivation (specialized version asks about biology, earth science, chemistry, and physics separately).
22	25	Liking Computers for Mathematics/Science	Identifies how much students like using computers to learn mathematics and science.
23	26	Internet Access	Identifies whether students are accessing the Internet and for what purposes they are using it.
24	27	Interest, Importance, & Value of Mathematics	Provides a description of student's interest, importance rating, and value attributed to mathematics.
25	28	Reasons to Do Well in Mathematics	Provides the extent to which students endorse certain reasons they need to do well in mathematics.
26	29	Classroom Practices in Mathematics	Provides a description of student's perceptions of classroom practices in mathematics instruction.
27	30	Beginning a New Mathematics Topic	Describes the frequency with which specific strategies are used in the classroom to introduce a new mathematics topic.
28	31	Taking Science Class(es)	Identifies whether or not the student is enrolled in science classes this year (specialized version asks about biology, earth science, chemistry, and physics separately).
29	32, 36, 40, 44	Interest, Importance, & Value of Science	Provides a description of student's interest, importance rating, and value attributed to science (specialized version asks about biology, earth science, chemistry, and physics separately).
30	33, 37, 41, 45	Reasons to Do Well in Science	Provides the extent to which students endorse certain reasons they need to do well in science (specialized version asks about biology, earth science, chemistry, and physics separately).
31	34, 38, 42, 46	Classroom Practices in Science	Provides a description of student's perceptions of classroom practices in science instruction (specialized version asks about biology, earth science, chemistry, and physics separately).
32	35, 39, 43, 47	Beginning a New Science Topic	Describes the frequency with which specific strategies are used in the classroom to introduce a new science topic (specialized version asks about biology, earth science, chemistry, and physics separately).
International Options			
33-34	48-49	People Living in the Home	Provides information about the home environment as an indicator of academic support and economic capital.
35-36	50-51	Cultural Activities	Provides a description of student's involvement in cultural events or programming such as plays or concerts.
37	52	Report on Student Behaviors	Provides an indication of the student's perspective on the existence of specific problematic student behaviors at school.
38	53	Environmental Issues	Provides and indication of student's beliefs about how much the application of science can help in addressing environmental issues.
39	54	Science Use in a Career	Identifies preference for sciences in careers.

3.4 Reporting Questionnaire Data

Reporting Student Background Data

Reporting the data from the student questionnaire is fairly straightforward. Most of the exhibits in the international reports present weighted percentages of students in each country for each response category, together with the mean achievement (mathematics or science) of those students. International averages are also displayed for each category.

Reporting student attitudes, self-perceptions, and activities related to science is complicated, however, by the fact that in some countries, science is taught as a general, integrated subject, while in others the fields of science - earth science, physics, chemistry, and biology - are taught as separate subjects. Countries could choose the appropriate version of the student questionnaire to administer: the general science version or the version for countries with separate science subjects. As a consequence results for questions that differed between the two versions need to be treated differently.

As a further complication, separate science countries do not always treat the science subjects in the same way. In some countries, earth science or chemistry was not applicable for the eighth grade. Also, in some countries combined courses such as physical science (physics/chemistry) or natural science (biology/earth science) were taught. In these cases, separate questions were still asked about separate science subjects (earth science, biology, physics, and chemistry). An exception was the Netherlands, where students were asked about earth science, biology, and physics/chemistry. When using the science background data for analysis, you will need to address these issues.

In TIMSS 1999, 23 countries administered the general version of the student questionnaire, and 15 countries administered the separate science subject version. Exhibit 3.4 lists the countries administering the general and separate science versions and indicates which science subjects were taught in each of the latter. In two countries, Chinese Taipei and Indonesia, the sciences were taught as separate subjects but students received a single science course grade, and so the general version of the student questionnaire was administered.

Exhibit 3.4 Countries that Administered the TIMSS 1999 General Science and Separate Science Subject Versions of the Student Questionnaire

Country	General Version	Separate Science Version			
	General/ Integrated Science	Earth Science	Biology	Physics	Chemistry
Australia	●				
Belgium (Flemish)		●	●	●	
Bulgaria		●	●	●	●
Canada	●				
Chile	●				
Chinese Taipei [a]	●				
Cyprus	●				
Czech Republic		●	●	●	●
England	●				
Finland		●	●	●	●
Hong Kong, SAR	●				
Hungary		●	●	●	●
Indonesia [b]	●				
Iran, Islamic Republic	●				
Israel	●				
Italy	●				
Japan	●				
Jordan	●				
Korea, Republic of	●				
Latvia			●	●	●
Lithuania			●	●	●
Macedonia, Republic of		●	●	●	●
Malaysia	●				
Moldova		●	●	●	●
Morocco		●	●	●	●
Netherlands [c]		●	●	●	
New Zealand	●				
Philippines	●				
Romania		●	●	●	●
Russian Federation		●	●	●	●
Singapore	●				
Slovak Republic		●	●	●	●
Slovenia			●	●	●
South Africa	●				
Thailand	●				
Tunisia	●				
Turkey	●				
United States	●				

a Chinese Taipei: separate sciences are taught starting in grade 7, with biology in grade 7 and physics/chemistry in grade 8. Since the students in the target grade take only one science course (physics/chemistry), the general version of the questionnaire was administered and students were asked about 'natural science', which would pertain to the physics/chemistry course in grade 8.

b Indonesia: students are taught 'IPA science' by separate biology and physics teachers, but students receive a single composite grade. The general version of the questionnaire was used, and students were asked about 'IPA science'.

c Netherlands: students were asked questions about integrated physics/chemistry; data for questions pertaining to physics/chemistry were reported in the physics panel in exhibits in the international reports.

Reporting Teacher Background Data

At the eighth grade level, mathematics and science are not generally taught by the same teachers. Accordingly, there was a questionnaire for mathematics teachers and another for science teachers, the two questionnaires having some general questions in common but using different subject matter related questions. The procedure was to sample a mathematics class from each participating school, administer the test to those students, and ask the mathematics and science teachers of those students to complete a teacher questionnaire. In countries with different teachers for each of the science subjects, this included all science teachers of the students in the sampled classes.[1]

The teacher questionnaires were divided into two sections: Section A asked about teachers' general background and Section B asked class-specific questions about instructional practices. Where teachers taught more than one mathematics or science class to the sampled students, they were to complete only one Section A but a separate Section B for each class taught. Thus, the information about instruction was tied directly to the students tested and the specific mathematics and science classes in which they were taught.

Because the sampling for the teacher questionnaires was based on participating students, these responses do not necessarily represent all of the teachers of the target grade in each of the TIMSS countries. Rather, they represent teachers of the representative samples of students assessed. It is important to note that when using the TIMSS data, the student is always the unit of analysis, even when information from the teacher questionnaires is being reported. Using the student as the unit of analysis makes it possible to describe the instruction received by representative samples of students. Although this approach may provide a different perspective from that obtained by simply collecting information from teachers, it is consistent with the TIMSS goals of illuminating students' educational contexts and performance.

The data obtained from the science teachers can be used in two ways. Some of the general information can be used together for all science teachers in each country. The data for information specific to the science subject, such as preparation to teach the sciences, instructional time in the sciences, and emphasis on experiments, should most likely be treated separately both for the general/integrated science and for the separate science subject area teachers. Tracking information provided by schools has been included in the database (ITCOURSE) and can be used to identify

▼ ▼ ▼

[1] In Slovenia and the Slovak Republic, background questionnaires were administered to only one of the separate science subject area teachers for the sampled mathematics classes. As a result, science teacher background data are not available for more than half of the relevant science teachers, and Slovenia and the Slovak Republic are not included in the exhibits based on science teacher data.

teachers by the type of course taught to the sampled students - mathematics, physics, biology, chemistry, earth science, or integrated science.

Another consequence of the TIMSS design was that since students were usually taught mathematics and science by different teachers and sometimes were taught one subject by more than one teacher, they had to be linked to more than one teacher for reporting purposes. When a student is taught a subject by more than one teacher, the student's sampling weight is distributed among those teachers. The student's contribution to student population estimates thus remains constant regardless of the number of teachers. This is consistent with the policy of reporting attributes of teachers and their classrooms in terms of the percentages of students taught by teachers with these attributes. Some analyses of this type might involve computing the sum or determining the highest value reported across all of a student's teachers for one subject area. The composite values obtained can then be used to produce the reported student-weighted statistics (e.g., total instructional time in the subjects and the degree of content coverage in mathematics or science).

Reporting School Background Data

The principals of the participating schools in TIMSS completed questionnaires on the school contexts in which the learning and teaching of mathematics and science occurred. Although schools constituted the first stage of sampling, the TIMSS school sample was designed to optimize the student sample, not to provide an optimal sample of schools.[2] Therefore, like the teacher data, the school-level data is most usefully reported using the student as the unit of analysis to describe the school contexts for the representative samples of students.

Summary Indices from Background Data

In an effort to summarize the information obtained from the background questionnaires concisely and focus attention on educationally relevant support and practice, TIMSS sometimes combined information to form an index that was more global and reliable than the component questions (e.g., students' home educational resources and attitudes towards mathematics or science; teachers' emphasis on reasoning and problem-solving, and confidence in their preparation to teach mathematics or science; availability of school resources for mathematics or science instruction). According to the responses of students, their teachers, and/or their schools, students were placed in a "high," "medium," or "low" category for such an index, with the high level being set so that it corresponds to conditions or activities generally associated with higher academic achievement.

▼ ▼ ▼

[2] See Chapter 5 for a description of the TIMSS sample design.

For example, a three-level index of home educational resources was constructed from students' responses to three questions: number of books in the home, educational aids in the home (computer, study desk/table for own use, dictionary), and parents' education. Students were assigned to the high level if they reported having more than 100 books, all three educational aids, and at least one parent had finished university. Students at the low level reported having 25 or fewer books in the home, fewer than all three educational aids, and the highest level of education for both parents as some secondary or less. Students with all other response combinations were assigned to the middle category.

The 17 indices computed for the TIMSS 1999 report are listed in Exhibit 3.5, which identifies the name of the index; the label used to identify it in the international report and database; the mathematics or science exhibit where the index data were reported; and the method used to compute the index. Further descriptions of these indices are presented in Supplement 3 of this User Guide.

Exhibit 3.5 Summary Indices from Background Data in the TIMSS 1999 International Reports

Name of Index	Label	Exhibit[a]	Analysis Method
Index of Home Educational Resources	HER	4.1 (M) 4.1 (S)	Index based on students' responses to three questions about home educational resources: number of books in the home; educational aids in the home (computer, study desk/table for own use, dictionary); parents' education. High level indicates more than 100 books in the home; all three educational aids; and either parent's highest level of education is finished university. Low level indicates 25 or fewer books in the home; not all three educational aids; and both parents' highest level of education is some secondary or less or is not known. Medium level includes all other possible combinations of responses. Response categories were defined by each country to conform to their own educational system and may not be strictly comparable across countries.
Index of Out-of-School Study Time	OST	4.5 (M) 4.5 (S)	Index based on students' responses to three questions about out-of-school study time: time spent after school studying mathematics or doing mathematics homework; time spent after school studying science or doing science homework; time spent after school studying or doing homework in school subjects other than mathematics and science. Number of hours based on: no time = 0, less than 1 hour = 0.5, 1-2 hours = 1.5, 3-5 hours = 4, more than 5 hours = 7. High level indicates more than three hours studying all subjects combined. Medium level indicates more than one hour to three hours studying all subjects combined. Low level indicates one hour or less studying all subjects combined.
Index of Students' Self-Concept in Mathematics	SCM	4.8 (M)	Index based on students' responses to five statements about their mathematics ability: 1) I would like mathematics much more if it were not so difficult; 2) Although I do my best, mathematics is more difficult for me than for many of my classmates; 3) Nobody can be good in every subject, and I am just not talented in mathematics; 4) Sometimes, when I do not understand a new topic in mathematics initially, I know that I will never really understand it; 5) Mathematics is not one of my strengths. High level indicates student disagrees or strongly disagrees with all five statements. Low level indicates student agrees or strongly agrees with all five statements. Medium level includes all other possible combinations of responses.
Index of Students' Self-Concept in the Sciences*	SCS-G SCS-E SCS-B SCS-P SCS-C	4.8 (S)	Index based on students' responses to four statements about their science ability: 1) I would like science much more if it were not so difficult; 2) Although I do my best, science is more difficult for me than for many of my classmates; 3) Nobody can be good in every subject, and I am just not talented in science; 4) Science is not one of my strengths. In countries where science is taught as separate subjects, students were asked about each subject area separately. High level indicates student disagrees or strongly disagrees with all four statements. Low level indicates student agrees or strongly agrees with all four statements. Medium level includes all other possible combinations of responses.
Index of Positive Attitudes Towards Mathematics	PATM	4.10 (M)	Index based on students' responses to five statements about mathematics: 1) I like mathematics; 2) I enjoy learning mathematics; 3) Mathematics is boring (reversed scale); 4) Mathematics is important to everyone's life; 5) I would like a job that involved using mathematics. Average is computed across the five items based on a 4-point scale: 1 = strongly negative; 2 = negative; 3 = positive; 4 = strongly positive. High level indicates average is greater than 3. Medium level indicates average is greater than 2 and less than or equal to 3. Low level indicates average is less than or equal to 2.
Index of Positive Attitudes Towards the Sciences*	PATS-G PATS-E PATS-B PATS-P PATS-C	4.10 (S)	Index based on students' responses to five statements about science: 1) I like science; 2) I enjoy learning science; 3) Science is boring (reversed scale); 4) Science is important to everyone's life; 5) I would like a job that involved using science. Average is computed across the five items based on a 4-point scale: 1 = strongly negative; 2 = negative; 3 = positive; 4 = strongly positive. In countries where science is taught as separate subjects, students were asked about each subject area separately. High level indicates average is greater than 3. Medium level indicates average is greater than 2 and less than or equal to 3. Low level indicates average is less than or equal to 2.
Index of Confidence in Preparation to Teach Mathematics	CPTM	6.3 (M)	Index based on teachers' responses to 12 questions about how prepared they feel to teach different mathematics topics based on a 3-point scale: 1 = not well prepared; 2 = somewhat prepared; 3 = very well prepared. Average is computed across the 12 items for topics for which the teacher did not respond "do not teach". High level indicates average is greater than or equal to 2.75. Medium level indicates average is greater than or equal to 2.25 and less than 2.75. Low level indicates average is less than 2.25.
Index of Confidence in Preparation to Teach Mathematics	CPTM	6.3 (M)	Index based on teachers' responses to 12 questions about how prepared they feel to teach different mathematics topics based on a 3-point scale: 1 = not well prepared; 2 = somewhat prepared; 3 = very well prepared. Average is computed across the 12 items for topics for which the teacher did not respond "do not teach". High level indicates average is greater than or equal to 2.75. Medium level indicates average is greater than or equal to 2.25 and less than 2.75. Low level indicates average is less than 2.25.

a Exhibit number in the international report where data based on the index were presented. An (M) indicates mathematics report; (S) indicates science report.

Exhibit 3.5 Summary Indices from Background Data in the TIMSS 1999 International Reports (continued)

Name of Index	Label	Exhibit[a]	Analysis Method
Index of Confidence in Preparation to Teach Mathematics	CPTM	6.3 (M)	Index based on teachers' responses to 12 questions about how prepared they feel to teach different mathematics topics based on a 3-point scale: 1 = not well prepared; 2 = somewhat prepared; 3 = very well prepared. Average is computed across the 12 items for topics for which the teacher did not respond "do not teach". High level indicates average is greater than or equal to 2.75. Medium level indicates average is greater than or equal to 2.25 and less than 2.75. Low level indicates average is less than 2.25.
Index of Confidence in Preparation to Teach Science	CPTS	6.3 (S)	Index based on teachers' responses to 10 questions about how prepared they feel to teach different science topics (see reference exhibit R3.2) based on a 3- point scale: 1 = not well prepared; 2 = somewhat prepared; 3 = very well prepared. Average is computed across the 10 items for items for which the teacher did not respond "do not teach". High level indicates average is greater than or equal to 2.75. Medium level indicates average is greater than or equal to 2.25 and less than 2.75. Low level indicates average is less than 2.25.
Index of Teachers' Emphasis on Scientific Reasoning and Problem-Solving	ESRPS	6.12 (S)	Index based on teachers' responses to five questions about how often they ask students to: 1) explain the reasoning behind an idea; 2) represent and analyze relationships using tables, charts, graphs; 3) work on problems for which there is no immediately obvious method of solution; 4) write explanations about what was observed and why it happened; 5) put events or objects in order and give a reason for the organization. Average is computed across the five items based on a 4-point scale: 1 = never or almost never; 2 = some lessons; 3 = most lessons; 4 = every lesson. High level indicates average is greater than or equal to 3. Medium level indicates average is greater than or equal to 2.25 and less than 3. Low level indicates average is less than 2.25.
Index of Teachers' Emphasis on Mathematics Reasoning and Problem-Solving	EMRPS	6.13 (M)	Index based on teachers' responses to four questions about how often they ask students to: 1) explain the reasoning behind an idea; 2) represent and analyze relationships using tables, charts, or graphs; 3) work on problems for which there is no immediately obvious method of solution; 4) write equations to represent relationships. Average is computed across the four items based on a 4- point scale: 1 = never or almost never; 2 = some lessons; 3 = most lessons; 4 = every lesson. High level indicates average is greater than or equal to 3. Medium level indicates average is greater than or equal to 2.25 and less than 3. Low level indicates average is less than 2.25.
Index of Emphasis on Conducting Experiments in Science Classes*	ECES-G ECES-E ECES-B ECES-P ECES-C	6.14 (S)	Index based on teachers' reports on the percentage of time they spend demonstrating experiments; teachers' reports on the percentage of time students spend conducting experiments; students' reports on how often the teacher gives a demonstration of an experiment in science lessons; students' reports on how often they conduct an experiment or practical investigation in class. In countries where science is taught as separate subjects, students were asked about each subject area separately, and only teachers who teach a particular subject are included in the index shown for that subject. High level indicates teacher reported that at least 25% of class time is spent on the teacher demonstrating experiments or students conducting experiments, and the student reported that the teacher gives a demonstration of an experiment or the student conducts an experiment or practical investigation in class almost always or pretty often. Low level indicates the teacher reported that less than 10% of class time is spent on the teacher demonstrating experiments or students conducting experiments, and student reported that the teacher gives a demonstration of an Experiment and the student conducts an experiment or practical investigation in class once in a while or never. Medium level includes all other possible combinations of responses.
Index of Emphasis on Calculators in Mathematics Class	ECMC	6.16 (M)	Index based on students' reports of the frequency of using calculators in mathematics lessons and teachers' reports of students' use of calculators in mathematics class for five activities, checking answers, tests and exams, routine computation, solving complex problems, and exploring number concepts. High level indicates the student reported using calculators in mathematics lessons almost always or pretty often, and the teacher reported students use calculators at least once or twice a week for any of the tasks. Low level indicates the student reported using calculators once in a while or never, and the teacher reported students use calculators never or hardly ever for all of the tasks. Medium level includes all other possible combinations of responses.

a Exhibit number in the international report where data based on the index were presented. An (M) indicates mathematics report; (S) indicates science report.

* Separate indices were computed for general/integrated science (G), earth science (E), biology (B), physics (P), and chemistry (C).

Exhibit 3.5 Summary Indices from Background Data in the TIMSS 1999 International Reports (continued)

Name of Index	Label	Exhibit[a]	Analysis Method
Index of Teachers' Emphasis on Science Homework	ESH	6.18 (S)	Index based on teachers' responses to two questions about how often they usually assign science homework and how many minutes of science homework they usually assign students (see reference exhibit R3.16). High level indicates the assignment of more than 30 minutes of homework at least once or twice a week. Low level indicates the assignment of less than 30 minutes of homework less than once a week or never assigning homework. Medium level includes all other possible combinations of responses.
Index of Teachers' Emphasis on Mathematics Homework	EMH	6.21 (M)	Index based on teachers' responses to two questions about how often they usually assign mathematics homework and how many minutes of mathematics homework they usually assign students (see reference exhibit R3.15). High level indicates the assignment of more than 30 minutes of homework at least once or twice a week. Low level indicates the assignment of less than 30 minutes of homework less than once a week or never assigning homework. Medium level includes all other possible combinations of responses.
Index of Availability of School Resources for Mathematics Instruction	ASRMI	7.1 (M)	Index based on schools' average response to five questions about shortages that affect general capacity to provide instruction (instructional materials; budget for supplies; school buildings and grounds; heating/cooling and lighting systems; instructional space), and the average response to five questions about shortages that affect mathematics instruction (computers; computer software; calculators; library materials; audio-visual resources) (see reference exhibits R4.1–R4.2). High level indicates that both shortages, on average, affect instructional capacity none or a little. Medium level indicates that one shortage affects instructional capacity none or a little and the other shortage affects instructional capacity some or a lot. Low level indicates that both shortages affect instructional capacity some or a lot.
Index of Availability of School Resources for Science Instruction	ASRSI	7.1 (S)	Index based on schools' average response to five questions about shortages that affect general capacity to provide instruction (instructional materials; budget for supplies; school buildings and grounds; heating/cooling and lighting systems; instructional space), and the average response to six questions about shortages that affect science instruction (laboratory equipment and materials; computers; computer software; calculators; library materials; audio-visual resources) (see reference exhibits R4.1–R4.2). High level indicates that both shortages, on average, affect instructional capacity none or a little. Medium level indicates that one shortage affects instructional capacity none or a little and the other shortage affects instructional capacity some or a lot. Low level indicates that both shortages affect instructional capacity some or a lot.
Index of Good School and Class Attendance	SCA	7.5 (M) 7.5 (S)	Index based on schools' responses to three questions about the seriousness of attendance problems in school: arriving late at school; absenteeism; skipping class (see exhibit 7.6). High level indicates that all three behaviors are reported to be not a problem. Low level indicates that two or more behaviors are reported to be a serious problem, or two behaviors are reported to be minor problems and the third a serious problem. Medium level includes all other possible combinations of responses.

a Exhibit number in the international report where data based on the index were presented. An (M) indicates mathematics report; (S) indicates science report.

* Separate indices were computed for general/integrated science (G), earth science (E), biology (B), physics (P), and chemistry (C).

3.5 Reporting Response Rates for Background Questionnaire Data

While it is desirable that all questions included in a data collection instrument be answered by all intended respondents, a certain percentage of non-response is inevitable. Not only do some questions remain unanswered; sometimes, entire questionnaires are not completed or not returned.

The handling of non-responses varies depending on how the data are to be reported. For the purpose of reporting, the exhibits in the TIMSS 1999 international reports contain special notations on response rates for the background variables.

Although in general the response rates for the student and school background variables were high, some variables and some countries exhibited less than acceptable rates. The non-response rates are somewhat higher for the teacher background data, particularly in cases where teachers were required to complete more than one questionnaire. Since the student is the unit of analysis, the non-response rates given in the international reports always reflect the percentage of students for whom the required responses from students, teachers, or schools were not available. The following response rates were used as thresholds when reporting background data in the TIMSS international reports.

- For a country where student, teacher or school responses were available for 70% to 84% of the students, the data were annotated with an "r" before the data column potentially affected by the response rate.
- When student, teacher, or school responses were available for 50% to 69% of the students, the data were annotated with an "s" before the data column potentially affected by the response rate.
- When student, teacher, or school responses were available for fewer than 50% of the students, the data were not reported and an "x" was placed within the appropriate data column.
- When the percentage of students in a particular category fell below 2%, achievement data were not reported in that category and a '~' was placed within the appropriate data column.

As a user, you can choose any threshold for annotating the data; however, caution is recommend, as it is important to disclose the response rates. Regardless of the missing rate, all responses are included in the database.

4 Data Collection, Scoring, Materials Processing, and Database Creation

4.1 Overview

Each country participating in TIMSS was responsible for collecting its national data and processing the materials in accordance with the international standards. In each country, a national research center and National Research Coordinator (NRC) were appointed to implement these activities. One of the main ways in which TIMSS sought to achieve uniform project implementation was by providing clear and explicit instructions on all operational procedures. Such instructions were provided primarily in the form of operations manuals, supported where possible by computer software systems that assisted NRCs in carrying out the specified field operations procedures. Forms accompanying some of the manuals served to document the implementation of the procedures in each country. Many of these forms were used to track schools, students, and teachers, and to ensure proper linkage of schools, students, and teachers in the database.

As part of the TIMSS quality assurance efforts, there also was a program of site visits by trained Quality Assurance Monitors representing the International Study Center.

4.2 Data Collection and Field Administration

For the sake of comparability, all testing in TIMSS was conducted as closely as possible to the end of the school year. For countries in the Southern Hemisphere, the school year ends in November or December, and thus testing was conducted in September through November. For countries in the Northern Hemisphere, the school year ends in May or June, and thus testing was conducted during those two months.

In TIMSS 1999, nine countries tested on a Southern Hemisphere school schedule - Australia, Chile, Korea, Malaysia, New Zealand, the Philippines, South Africa, and Thailand. The remaining countries tested their students on the Northern Hemisphere schedule, with the exception of Lithuania, which tested its students in October 1999, when students were beginning grade 9.

In TIMSS 1995, four countries tested on the Southern Hemisphere school schedule – Australia, Korea, New Zealand, and Singapore. The remaining countries tested on the Northern Hemisphere schedule.

In addition to selecting the sample of students to be tested, the NRCs were responsible for working with the School Coordinators to organize data collection within schools, translating the test instruments, assembling and printing the test booklets, and packing and shipping the necessary materials to the designated School Coordinators. They also were responsible for arranging for the return of the testing materials from the school sites to the national center, preparing for and implementing the free-response scoring, entering the results into data files, conducting on-site quality assurance observations for a 10 percent of schools, and preparing a report on survey activities.

The *Survey Operations Manual* was prepared by the International Study Center at Boston College and the IEA Data Processing Center for the NRCs and their colleagues who were responsible for implementing the TIMSS data collection and processing procedures. This manual describes the activities and responsibilities of the NRCs from the receipt of international testing materials at the national center to the delivery/shipment of cleaned data sets and accompanying documentation to the IEA Data Processing Center. In addition to detailed within-school sampling instructions, the manual included:

- Procedures for translating and assembling the test instruments and questionnaires
- Instructions for obtaining cooperation from the selected schools
- Explicit procedures for packing and sending materials to the schools
- Preparations for test administration
- Instructions for data entry and verification.

Included in this manual were a set of survey tracking forms that were completed at various stages of the study to track schools, classrooms, and students, and ensure proper linkage among them in the database.

Each school was asked to appoint a coordinator to be the liaison between the national research center and the school. The *School Coordinator Manual* describes the steps the School Coordinator followed. Essentially, the School Coordinator was responsible for receiving the shipment of testing materials from the national center and providing for their security before and after the test date. The School Coordinator administered the teacher questionnaires, arranged the testing accommodations, trained Test Administrators, and arranged for make-up sessions when necessary.

The *Test Administrator Manual* covered the procedures from the beginning of testing to the return of the completed tests, questionnaires, and tracking forms to the School Coordinator. The manual includes a Test Administration Script to be read by the Test Administrator. The Test Administrators were responsible for activities preliminary to the testing session, including maintaining the security of the test booklets and ensuring adequacy of supplies and the testing environment. Activities during the testing session included distribution of the test booklets to the appropriate students (using the Student Tracking Form), timing of the testing and breaks, and accurately reading the test administration script.

Assigning Testing Materials to Students and Teachers

Eight different test booklets were distributed among the students in each sampled class. Each student was required to complete one booklet and the student questionnaire. Booklets were assigned to students using a random assignment procedure. Each mathematics or science teacher of the selected students was assigned a mathematics or a science teacher questionnaire. Where teachers taught both mathematics and science to the students, every effort was made to collect information about both topics. However, NRCs had the final decision as to how much response burden to place on such teachers.

National Quality Control Program

The International Study Center implemented an international quality control program whereby international quality control monitors visited a sample of 15 schools in each country and observed the test administration. In addition, NRCs were expected to organize a national quality control program, based upon the international model. This national program required Quality Control Observers to document data collection activities in their country. They visited a 10 percent sample of TIMSS 1999 schools, observed actual testing sessions, and recorded compliance of the test administration with prescribed procedures.

The International Study Center prepared *The Manual for National Quality Control Observers* (TIMSS, 1998c), which contained information about TIMSS 1999 and detailed the role and responsibilities of the National Quality Control Observers.

4.3 Free-Response Scoring

Upon completion of each testing session, the School Coordinator shipped the booklets, questionnaires, and forms to the National Research Center. The NRCs then organized the instruments for scoring and data entry. These procedures were designed to maintain identification information that linked students to schools and teachers, minimize the time and effort spent handling the booklets, ensure reliability in the free-response scoring, and document the reliability of the scoring. Since approximately one-third of the written test time was devoted to free-response items, the free-response scoring was a considerable task.

The free-response items were scored using item-specific rubrics. Scores were represented by two-digit codes. The first digit designates the correctness level of the response. The second digit, combined with the first, represents a diagnostic code used to identify specific types of approaches, strategies, or common errors and misconceptions. This coding approach was used with all free-response items, including both the short-answer and extended-response items.[1]

The number of points specified in each rubric varies by item since each item is unique in terms of answer, approach, and types of misconceptions generated by students. Most items are worth one point. In rubrics for one-point items, correct student responses were coded as 10, 11, 12, and so on through 19. The type of response in terms of the approach used or explanation provided is denoted by the second digit.

Regardless of the number of possible score points for an item, incorrect student responses were coded as 70, 71, and so on through 79 and were awarded zero score points. However, as in the approach used for correct scores, the second digit in the code represents the type of misconception displayed, incorrect strategy used, or incomplete explanation given.

The rubrics for two-point items were more complicated, having categories for full and partial credit. On such items students received two points for a fully correct answer and one point for a partially correct answer.

4.4 Documenting the Reliability of the Free-Response Scoring

In order to demonstrate the quality of the TIMSS 1999 data, it was important to document the agreement between scorers. To establish the scoring reliability, NRCs were required to have a random sample of about 200 of each booklet type independently scored by two scorers. This constituted a 25% sample in most countries. The degree of agreement between the two scores is a measure of the reliability of the scoring process. The information collected from this double scoring procedure is available in this database.

Since the purpose of the double scoring was to document the consistency of scoring, the procedure used for scoring the booklets in the reliability sample had to be as close as possible to that used for scoring the booklets in general. The recommended procedure was designed to blend the scoring of the reliability sample with the normal scoring activity, to take place throughout the scoring process, and to be systematically implemented across student responses and scorers.

▼ ▼ ▼

1 See Chapter 2 of this User Guide for more detailed information regarding the scoring rubrics used in mathematics and science.

To meet the goal of reliable scoring, TIMSS had an ambitious training program. The training sessions were designed to assist representatives of the national centers who were then responsible for training personnel in their respective countries to apply the two-digit codes reliably. TIMSS recommended that scorers be organized into teams of about six, headed by a team leader who monitored the progress and reliable use of the codes. The team leaders were to continually check and reread the responses coded by their team, systematically covering the daily work of each scorer.

4.5 Data Entry

The DPC provided an integrated computer program for data entry and data verification known as the DATAENTRYMANAGER (DEM). This program worked on all IBM-compatible personal computers running under DOS, OS/2 or Windows 3.x, 95 or 98. It facilitated data entry directly from the tracking forms and test instruments and provided a convenient checking and editing mechanism. DEM also offered data and file management capabilities, interactive error detection, reporting, and quality control procedures. Detailed information and operational instructions were provided in the DATAENTRYMANAGER manual. Since DEM incorporated the international codebooks describing all variables, use of the software ensured that the data files were produced according to the TIMSS 1999 rules and standards for data entry.

Although use of DEM for all data entry tasks was strongly recommended, NRCs were permitted to use their own procedures and computer programs, as long as all data files conformed to the specifications of the international codebooks. NRCs who chose not to use DEM were responsible for ensuring that all data files were delivered to the DPC in the international format.

After entering data in files in accordance with the international procedures, countries submitted their data files to the IEA Data Processing Center.

4.6 Data Checking and Editing

The TIMSS data underwent an exhaustive cleaning process designed to identify, document, and correct deviations from the international instruments, file structures, and coding schemes. The process also emphasized consistency of information with national data sets and appropriate linking among the many data files. The national centers were contacted regularly throughout the cleaning process and were given multiple opportunities to review the data for their countries.

4.7 Linking Students, Teachers, and Classes

Within each school, a class identification number (ID) was assigned to each class in the target grades listed on the Class Tracking Form. The class ID consisted of the school ID plus an identification number for the class within the school. Each student tested was also assigned a student identification number.

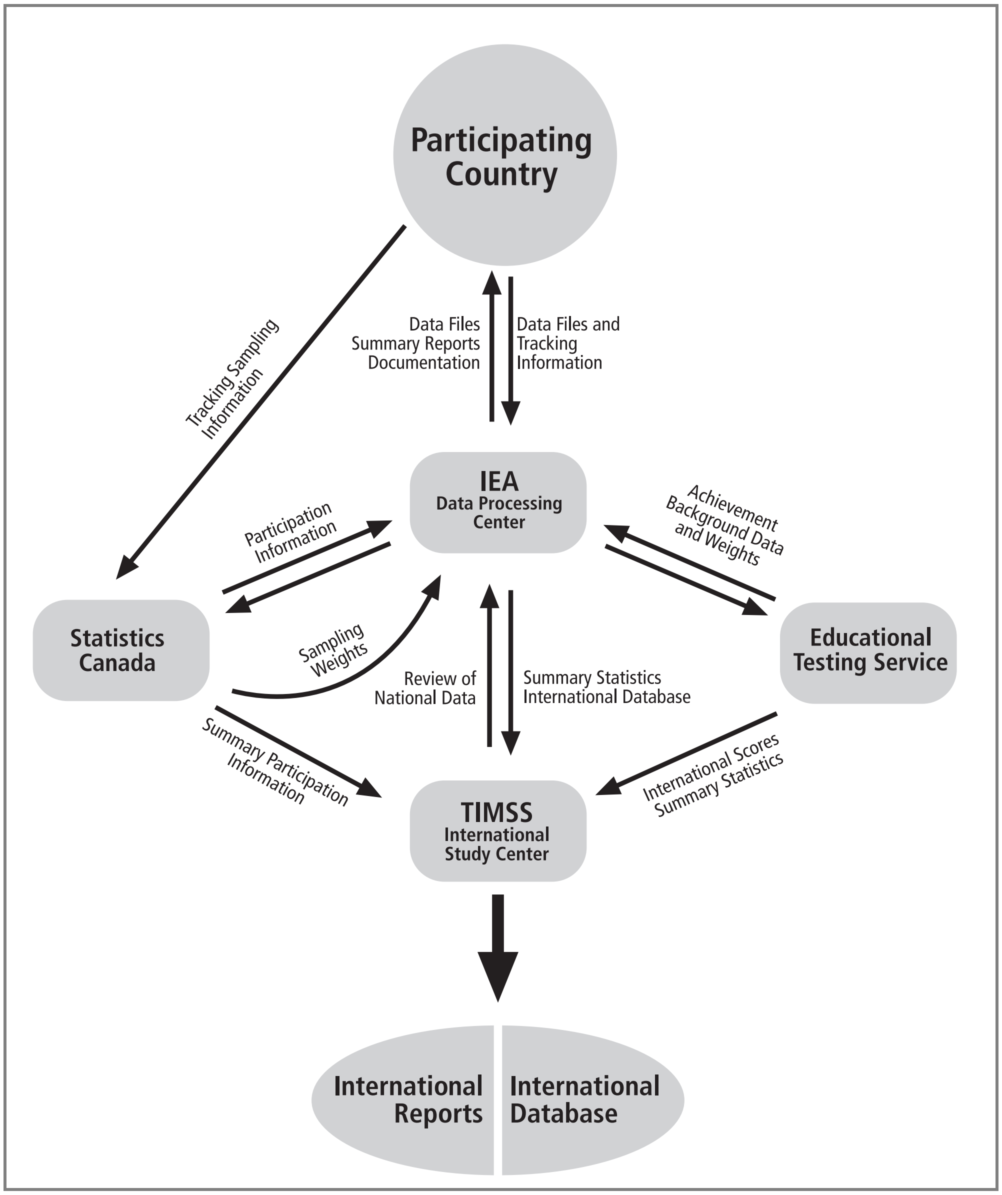
All mathematics and science teachers of the selected classes were assigned a teacher ID that consisted of the school ID plus a two-digit number of the teacher within the school. Since a teacher could be teaching both mathematics and science to some or all of the students in a class, it was necessary to have a unique identification number for each teacher/class and teacher/subject combination. This was achieved by adding a two-digit link number to the five digits of the teacher ID, giving a unique seven-digit teacher/class identification number. Careful implementation of these procedures was necessary to be able to link classes to a teacher and analyze student outcomes in relation to teacher-level variables.

4.8 Data Flow

The data collected in the TIMSS 1999 survey were entered into data files with a common international format at the national research centers of the participating countries. These data files were then submitted to the IEA Data Processing Center for cleaning and verification. The major responsibilities of the Data Processing Center were to check that the data files matched the international standard and make modifications where necessary, to apply standard cleaning rules to the data to verify the consistency and accuracy of their processing, and to interact with the NRCs to ensure their accuracy. Other responsibilities were to produce summary statistics of the background and achievement data for review by the TIMSS International Study Center and, upon feedback from the individual countries and the TIMSS International Study Center, to construct the international database. The IEA Data Processing Center was also responsible for distributing the national data files to each of the participating countries.

Once verified and in the international file format, the achievement data were sent to the International Study Center where basic item statistics were produced and reviewed. At the same time, the Data Processing Center sent data files containing information on the participation of schools and students in each country's sample to Statistics Canada. This information, together with data provided by the NRC from tracking forms and the within-school sampling software, was used by Statistics Canada to calculate sampling weights, population coverage, and school and student participation rates. The sampling weights were sent to the TIMSS International Study Center for verification and then forwarded to the Educational Testing Service for use in scaling the achievement data. When the review of the item statistics was completed and the Data Processing Center had updated the database, the student achievement files were sent to the Educational Testing Service, which conducted the scaling and generated proficiency scores in mathematics and science for each participating student. Once the sampling weights and the proficiency scores were verified at the International Study Center, they were sent to the Data Processing Center for inclusion in the international database and then distributed to the national research centers. The International Study Center prepared the exhibits for the international reports and published the results of the study. Exhibit 4.1 is a pictorial representation of the flow of the data files.

Exhibit 4.1 Flow of Data Files



5 Sampling and Sampling Weights

5.1 Overview

This chapter summarizes the selection of school and student samples and the sampling weights included in the TIMSS data files. The TIMSS 1999 sample design and weighting procedures are fully detailed in Foy and Joncas (2000) and Foy (2000). Sampling information for TIMSS 1995 is presented in Foy, Rust, and Schleicher (1996).

To be acceptable for TIMSS, national sample designs had to result in probability samples that gave accurately weighted estimates of population parameters, and for which estimates of sampling variance could be computed. The TIMSS 1999 sample design was very similar to that of its predecessor, TIMSS 1995. The TIMSS design was chosen so as to balance analytical requirements and operational constraints, while keeping it simple enough for all participants to implement.

The international project management provided manuals and expert advice to help National Research Coordinators (NRCs) adapt the TIMSS 1999 sample design to their national system and to guide them through the phases of sampling. The TIMSS 1999 *School Sampling Manual* (TIMSS, 1997) described how to implement the international sample design and offered advice on planning, working within constraints, establishing appropriate sample selection procedures, and fieldwork. The *Survey Operations Manual* (TIMSS, 1998a) and *School Coordinator Manual* (TIMSS, 1998b) discussed sample selection and execution within schools, the assignment of test booklets to selected students, and administration and monitoring procedures used to identify and track respondents and non-respondents. NRCs also received software designed to automate the sometimes complex within-school sampling procedures.

NRCs were allowed to adapt the basic TIMSS sample design to the needs of their education system by using more sampling information or more sophisticated designs and procedures. These adjustments, however, had to be approved by the International Study Center at Boston College and monitored by Statistics Canada.

5.2 Target Populations and Exclusions

In IEA studies, the target population for all countries is known as the *international desired population*. The international desired population for TIMSS 1999 was as follows:

- All students enrolled in the upper of the two adjacent grades that contain the largest proportion of 13-year-olds at the time of testing.

The TIMSS 1999 target grade was the upper grade of the TIMSS 1995 population 2[1] and was expected to be the eighth grade in most countries. This would allow countries participating in both TIMSS 1995 and TIMSS 1999 to establish a trend line of comparable achievement data. Exhibits 5.1 and 5.2 show the grades that meet this definition within each country for TIMSS 1999 and TIMSS 1995, respectively.

Exhibit 5.1 Grades Tested in TIMSS 1999

Country	Country's Name for Grade Tested	Years of Formal Schooling	Average Age of Students Tested
Australia	8 or 9	8 or 9	14.3
Belgium (Flemish)	2A & 2P	8	14.1
Bulgaria	8	8	14.8
Canada	8	8	14.0
Chile	8	8	14.4
Chinese Taipei	2nd Grade Junior High School	8	14.2
Cyprus	8	8	13.8
Czech Republic	8	9	14.4
England	Year 9	9	14.2
Finland	7	7	13.8
Hong Kong, SAR	Secondary 2	8	14.2
Hungary	8	8	14.4
Indonesia	2nd Grade Junior High School	8	14.6
Iran, Islamic Rep.	9	8	14.6
Israel	9	8	14.1
Italy	3rd Grade Middle School	8	14.0
Japan	2nd Grade Lower Secondary	8	14.4
Jordan	8	8	14.0
Korea, Rep. of	2nd Grade Middle School	8	14.4
Latvia (LSS)	8	8	14.5
Lithuania[1]	9	8.5	15.2
Macedonia, Rep. of	8	8	14.6
Malaysia	Form 2	8	14.4
Moldova	8	9	14.4
Morocco	7	7	14.2
Netherlands	Secondary 2	8	14.2
New Zealand	Year 9	8.5 to 9.5	14.0
Philippines	1st Year High School	7	14.1
Romania	8	8	14.8
Russian Federation	8	7 or 8	14.1
Singapore	Secondary 2	8	14.4
Slovak Republic	8	8	14.3
Slovenia	8	8	14.8
South Africa	8	8	15.5
Thailand	Secondary 2	8	14.5
Tunisia	8	8	14.8
Turkey	8	8	14.2
United States	8	8	14.2

1 Lithuania tested the same cohort of students as other countries, but later in 1999, at the beginning of the next school year.

▼ ▼ ▼

1 For the TIMSS 1995 population definition, see Foy, Rust, & Schleicher (1996).

Exhibit 5.2 Grades Tested in TIMSS 1995 – Population 2

Country	Lower Grade		Upper Grade	
	Country's Name for Lower Grade	Years of Formal Schooling Including Lower Grade[1]	Country's Name for Upper Grade	Years of Formal Schooling Including Upper Grade[1]
Australia[2]	7 or 8	7 or 8	8 or 9	8 or 9
Austria	3. Klasse	7	4. Klasse	8
Belgium (Fl)	1A	7	2A & 2P	8
Belgium (Fr)	1A	7	2A & 2P	8
Bulgaria	7	7	8	8
Canada	7	7	8	8
Colombia	7	7	8	8
Cyprus[3]	7	7	8	8
Czech Republic	7	7	8	8
Denmark	6	6	7	7
England	Year 8	8	Year 9	9
France	5ème	7	4ème (90%) or 4ème Technologique (10%)	8
Germany	7	7	8	8
Greece	Secondary 1	7	Secondary 2	8
Hong Kong	Secondary 1	7	Secondary 2	8
Hungary	7	7	8	8
Iceland	7	7	8	8
Iran, Islamic Rep.	7	7	8	8
Ireland	1st Year	7	2nd Year	8
Israel	–	–	8	8
Japan	1st Grade Lower Secondary	7	2nd Grade Lower Secondary	8
Korea, Republic of	1st Grade Middle School	7	2nd Grade Middle School	8
Kuwait	–	–	9	9
Latvia	7	7	8	8
Lithuania	7	7	8	8
Netherlands	Secondary 1	7	Secondary 2	8
New Zealand[3, 4]	Form 2	7.5 - 8.5	Form 3	8.5 - 9.5
Norway[3]	6	6	7	7
Philippines[3]	Grade 6 Elementary	6	1st Year High School	7
Portugal	Grade 7	7	Grade 8	8
Romania	7	7	8	8
Russian Federation[5]	7	6 or 7	8	7 or 8
Scotland	Secondary 1	8	Secondary 2	9
Singapore	Secondary 1	7	Secondary 2	8
Slovak Republic	7	7	8	8
Slovenia	7	7	8	8
Spain	7 EGB	7	8 EGB	8
South Africa[3]	Standard 5	7	Standard 6	8
Sweden[3]	6	6	7	7
Switzerland[3]				
(German)	6	6	7	7
(French and Italian)	7	7	8	8
Thailand	Secondary 1	7	Secondary 2	8
United States	7	7	8	8

1 Years of schooling based on the number of years children in the grade level have been in formal schooling, beginning with primary education (International Standard Classification of Education Level 1). Does not include preprimary education.

2 Australia: Each state/territory has its own policy regarding age of entry to primary school. In 4 of the 8 states/territories students were sampled from grades 7 and 8; in the other four states/territories students were sampled from grades 8 and 9.

3 Indicates that there is a system-split between the lower and upper grades. In Cyprus, system-split occurs only in the large or city schools. In Switzerland there is a system-split in 14 of 26 cantons.

4 New Zealand: The majority of students begin primary school on or near their 5th birthday so the "years of formal schooling" vary.

5 Russian Federation: 70% of students in the seventh grade have had 6 years of formal schooling; 70% in the eighth grade have had 7 years of formal schooling.

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95. Information provided by TIMSS National Research Coordinators.

Excluded Students

TIMSS 1999 expected all participating countries to define their *national desired population* to correspond as closely as possible to its definition of the international desired population. Sometimes, however, NRCs had to make changes. For example, some countries had to restrict geographical coverage by excluding remote regions or excluding a segment of their education system. The technical reports document any deviations from the international definition of the TIMSS target population.

Using their national desired population as a basis, participating countries had to operationally define their population for sampling purposes. This definition, known in IEA terminology as the *national defined population*, is essentially the sampling frame from which the first stage of sampling takes place. The national defined population could be a subset of the national desired population. All schools and students from the former excluded from the latter are referred to as the *excluded population.*

TIMSS 1999 participants were expected to keep the excluded population to no more than 10 percent of the national desired population. Exclusions could occur at the school level, within schools, or both. Because the national desired population was restricted to schools that contained the target grade, schools not containing this grade were considered to be outside the scope of the sampling frame, and not part of the excluded population. Participants could exclude schools from the sampling frame for the following reasons:

- They were in geographically remote regions.
- They were of extremely small size.
- They offered a curriculum, or school structure, that was different from that of the mainstream education system(s).
- They provided instruction only to students in the exclusion categories defined as "within-sample exclusions."

Within-school exclusions were limited to students who, because of some disability, were unable to take the TIMSS 1999 tests. NRCs were asked to define anticipated within-school exclusions. Because these definitions can vary internationally, NRCs were also asked to follow certain rules adapted to their jurisdictions. In addition, they were to estimate the size of such exclusions so that compliance with the 10 percent rule could be gauged in advance.

The general TIMSS 1999 rules for defining within-school exclusions included:

- **Educable mentally disabled students.** These are students who were considered, in the professional opinion of the school principal or other qualified staff members, to be educable mentally disabled, or students who had been so diagnosed by psychological tests. This included students who were emotionally or mentally unable to follow even the general instructions of the TIMSS test. It did not include students who merely exhibited poor academic performance or discipline problems.
- **Functionally disabled students**. These are students who were permanently physically disabled in such a way that they could not perform in the TIMSS tests. Functionally disabled students who could perform were included in the testing.
- **Non-native-language speakers**. These are students who could not read or speak the language of the test and so could not overcome the language barrier of testing. Typically, a student who had received less than one year of instruction in the language of the test was excluded, but this definition was adapted in different countries.

The size of the excluded population was documented and served as an index of the coverage and representativeness of the selected samples.[2]

5.3 Sample Design

The basic sample design for TIMSS 1999 is generally referred to as a two-stage stratified cluster sample design. The first stage consisted of a sample of schools[3], which may be stratified; the second stage consisted of a single classroom selected at random from the target grade in sampled schools. Participating countries were asked to define this class on the basis of mathematics instruction.

Although in the second sampling stage the sampling units were intact mathematics classrooms, the ultimate sampling units were students. Consequently, it was important that each student from the target grade be a member of one and only one of the mathematics classes in a school from which the sampled classes were to be selected. In most education systems, the mathematics class coincided with a science

▼ ▼ ▼

[2] For the 1999 assessment see Foy and Joncas (2000), "Implementation of the Sample Design", in Martin, M.O., Gregory, K.D., and Stemler, S.E. (Eds.). *TIMSS 1999 Technical Report*, Chestnut Hill, MA: Boston College.

[3] In some very large countries, it was necessary to include an extra preliminary stage in which school districts were sampled first, and then schools.

class or classes. In some systems, however, it may have been the case that some of the students in the selected mathematics class were not enrolled in a science class, and possibly some students in the science class were not enrolled in any mathematics class. In any case, participating countries were asked to define the classrooms on the basis of mathematics instruction. If not all students in the national desired population belonged to a mathematics or science class, then an alternative definition of the classroom was required for ensuring that the non-mathematics or non-science students had an opportunity to be selected.

Sampling Precision and Sample Size

Sample sizes for TIMSS had to be specified so as to meet the analytic requirements of the study. Since students were the principal units of analysis, the ability to produce reliable estimates of student characteristics was important. The TIMSS standard for sampling precision required that all population samples have an effective sample size of at least 400 students for mathematics and science achievement. In other words, the samples should have sampling errors no greater than those that would be obtained from a simple random sample of 400 students.

Furthermore, since TIMSS was designed to allow for analyses at the school and classroom levels, at least 150 schools were to be selected from the target population. A sample of 150 schools results in 95 percent confidence limits for school-level and classroom-level mean estimates that are precise to within plus or minus 16 percent of their standard deviations. To ensure sufficient sample precision for these units of analysis, some participants had to sample more schools than they would have selected otherwise.

The precision of multistage cluster sample designs is generally affected by the so-called clustering effect. A classroom as a sampling unit constitutes a cluster of students who tend to be more like each other than like other members of the population. The *intraclass correlation* is a measure of this similarity. For example, sampling 30 students from a single classroom when the intraclass correlation is positive will yield less information than a random sample of 30 students spread across all classrooms in a school. Such sample designs are less efficient, in terms of information per sampled student, than a simple random sample of the same size. This clustering effect has to be considered when computing standard error and confidence intervals with the TIMSS data. The traditional assumptions of simple random sampling statistical methods do not hold true with these data.

5.4 School Sample Selection

The sample-selection method used for first-stage sampling was based on a systematic probability-proportional-to-size (PPS) technique. The probability of selection for a school was proportional to the number of eighth-grade students in the school (see Foy and Joncas, 2000). It was sometimes the case that a sampled school was unable to participate in the assessment. In such cases, the originally sampled school was replaced by a *replacement school*. The mechanism for selecting replacement schools, established a priori, identified the next school on the ordered school-sampling list as the replacement for each particular sampled school. The school after that was a second replacement, should it be necessary. Ordering of the school sampling frame by stratification variables and by size ensured that any original sampled school's replacement would have similar characteristics.

5.5 Classroom and Student Sampling

In the second sampling stage, classrooms of students were sampled. Generally, in each school, one classroom was sampled from the target grade, although some countries opted to sample two classrooms in order to be able to conduct special analyses. Most participants tested all students in the sampled classrooms, and in these instances the classrooms were sampled with equal probabilities.

As an optional third sampling stage, participants with particularly large classrooms in their schools could decide to subsample a fixed number of students per selected classroom. This was done using a simple random sampling method whereby all students in a sampled classroom were assigned equal selection probabilities. This was done only in 1995.

Excluding Students from Testing

Although all students enrolled in the target grade were part of the target population and were eligible to be selected for testing as already described in this chapter, TIMSS recognized that some students in every school would be unable to take part in the assessment because of some physical disability, mental disability, or a language difference. Accordingly, the sampling procedures provide for the exclusion of students with any of several disabilities or language differences. Countries were required to track and account for all excluded students, and were cautioned that excluding an excessive proportion would lead to their results being annotated in the international reports.

5.6 Sampling Weights

Appropriate estimation of population characteristics based on the TIMSS samples requires that the TIMSS sample design be taken into account in all analyses. This is accomplished in part by assigning a weight to each respondent, where the sampling weight properly accounts for the sample design, takes into account any stratification or disproportional sampling of subgroups, and includes adjustments for non-response (Foy, 2000).

The students within each country were selected using probability sampling. This means that the probability of each student being selected as part of the sample is known. The inverse of this selection probability is the sampling weight. In a properly selected and weighted sample, the sum of the weights for the sample approximates the size of the population. As is the case in TIMSS, the sum of the sampling weights for a sample is an estimate of the size of the population of students within the country in the sampled grades. The sampling weights must be used whenever population estimates are required. The use of the appropriate sampling weights ensures that the different subgroups that constitute the sample are properly and proportionally represented in the computation of population estimates.

Weight Variables Included in the Student Data Files [4]

Each student's sampling weight is a composite of five factors: the school weighting factor, the school weighting adjustment, the class weighting factor, the student weighting factor, and the student weighting adjustment. In addition, three versions of each student's weight are provided- the "total student" weight, the "senate" weight, and the "house" weight- each with its own particular uses.

The variables described in this section are included in the Student Background and Student Achievement files. The meaning and interpretation of the weights in each of the files is the same. The weighting factors included in the student-level data files and their adjustment factors are as follows.

WGTFAC1 School Weighting Factor

This variable corresponds to the inverse of the probability of selection for the school where the student is enrolled.

▼ ▼ ▼

[4] For the user not familiar with the data files included in the TIMSS International Database reading Chapter 8 is recommended before proceeding with this section.

WGTADJ1 School Weighting Adjustment

This is an adjustment that is applied to WGTFAC1 to account for non-participating schools in the sample. Multiplying WGTFAC1 by WGTADJ1 gives the sampling weight for the school, adjusted for non-participation.

WGTFAC2 Class Weighting Factor

This is the inverse of the probability of selection of the classroom within the school. Since in general only one classroom was selected per grade within each school, there was no need to compute an adjustment factor for the classroom weight.

WGTFAC3 Student Weighting Factor

This is the inverse of the probability of selection for the individual student within a classroom. In cases where an intact classroom was selected, the value is set to 1 for all members of the classroom.

WGTADJ3 Student Weighting Adjustment

This is an adjustment applied to the variable WGTFAC3 to account for non-participating students in the selected school and/or classroom. Multiplying the variables WGTFAC2, WGTFAC3, and WGTADJ3 and adding them up within each school, gives an estimate of the number of students within the sampled school.

The five variables listed above are combined to give a student's overall sampling weight. The probability of selecting an individual student is the product of three independent events: selecting the school, the classroom, and the student. To obtain the probability of selection for an individual student multiply three selection probabilities – school, classroom, and student – and their respective adjustment factors. Inverting this probability gives the sampling weight for the student.

Three versions of the students' sampling weight are provided in the user database. All three give the same figures for statistics such as means and proportions, but vary for statistics such as totals and population sizes. Each one has particular advantages in certain circumstances. These three versions are as follows.

TOTWGT Total Student Weight

This is obtained by simply multiplying the variables WGTFAC1, WGTADJ1, WGTFAC2, WGTFAC3, and WGTADJ3 for the student. The sum of these weights within a sample provides an estimate of the size of the population. Although this is a commonly used sampling weight, it sometimes adds to a very large number, and to a different number within each country. This is not always desirable. For example, if you want to compute a weighted estimate of the mean achievement in the population across all countries, using the variable TOTWGT as your weight variable will lead each country to contribute proportionally to its population size, with the large countries counting more than small countries. Although this is desirable in some circumstances (e.g., when computing the 75th percentile for mathematics achievement for students around the world), in general TOTWGT is not the student weight of choice for cross-country analyses, since it does not treat countries equally, and gives inflated results in significance tests when the proper adjustments are not used.

A key property of the sampling weights is that the same population estimates for means and proportions will be obtained as long as you use a weight variable proportional to the original weights (TOTWGT). For example, you could take the sampling weights for a large country and divide them by a constant to make them smaller. You could also take the weights of a smaller country and multiply them by a constant to make them bigger. Regardless of which constant is used within a country, the weighted estimates obtained from each of these proportional transformations of the weights will be exactly the same. Two other weight variables are computed and included in the student data files. Each of these is computed for a specific purpose and will yield exactly the same results within each country, but will have some desirable properties when estimates across countries are computed or significance tests are performed.

SENWGT **Senate Weight**

This variable is computed as

$$SENWGT_{g_i i} = TOTWGT_{g_i i}\left(\frac{500}{\sum_{i=1}^{I} TOTWGT_{g_i i}}\right)$$

for each student, by grade within each country, where *i* is the individual student and *g* is the grade of the student. The transformation of the weights will be different within each country, but in the end the sum of the variable SENWGT within each country will be 500 per grade or 1,000 for both grades. The variable SENWGT, within each country, is proportional to TOTWGT multiplied by the ratio of 500 divided by the sum of the weights over all students in the grade. These sampling weights can be used when international estimates are sought and you want to have each country contribute the same amount to the international estimate. When this variable is used as the sampling weight for international estimates, the contribution of each country is the same, regardless of the size of the population.

HOUWGT **House Weight**

This variable is computed as

$$HOUWGT_{g_i i} = TOTWGT_{g_i i}\left(\frac{n}{\sum_{i=1}^{I} TOTWGT_{g_i i}}\right)$$

for each student, by grade within each country, where *i* is the individual student and *g* is the grade of the student. The transformation of the weights will be different within each country, but in the end the sum of the variables HOUWGT over all schools within each country will add up to the sample size for that country. The variable HOUWGT is proportional to TOTWGT multiplied by the ratio of the sample size (*n*) divided by sum of the weights over all students in the grade. These sampling weights can be used when you want the actual sample size to be used in performing significance tests. Although some statistical computer software packages allow you to use the sample size as the divisor in the computation of standard errors, others will use the sum of the weights, and this results in severely deflated standard errors for the statistics if TOTWGT is used as the weighting variable. When performing analyses using such software, it is recommended to use the variable HOUWGT as the weight variable. Because of the clustering effect in most TIMSS samples, it may also be desirable to apply a correction factor such as a design effect to the HOUWGT variable.

Weight Variables Included in the Student-Teacher Linkage Files

The individual student sampling weights generally should be used when you want to obtain estimates at the student level. The exception is when student and teacher data are to be analyzed together. In this case, a separate set of weights have been computed to account for the fact that a student could have more than one mathematics or science teacher. These weight variables are included in the Student-Teacher Linkage file and are listed below.

MATWGT

This weight is computed by dividing the sampling weight for the student by the number of mathematics teachers that the student has. This weight should be used whenever you want to obtain estimates regarding students and their mathematics teachers.

SCIWGT

This weight is computed by dividing the sampling weight for the student by the number of science teachers that the student has. This weight should be used whenever you want to obtain estimates regarding students and their science teachers.

TCHWGT

This weight is computed by dividing the sampling weight for the student by the number of mathematics and science teachers that the student has. This weight should be used whenever you want to obtain estimates regarding students and their mathematics and science teachers combined.

The Student-Teacher Linkage file also includes variables that indicate the number of mathematics teachers, the number of science teachers, and the number of mathematics and science teachers the student has.

Weight Variables Included in the School Data Files

The TIMSS samples are samples of students within countries. Although they are made up of a sample of schools within the countries, the samples of schools are selected so that the sampling of students, rather than the sampling of schools, is optimized. In particular, the probability-proportional-to-size sampling methodology causes large schools to be oversampled. Several weight variables are included in the school files, as follows:

WGTFAC1 School Weighting Factor

This variable corresponds to the inverse of the probability of selection for the school where the student is enrolled.

WGTADJ1 School Weighting Adjustment

This is an adjustment that is applied to WGTFAC1 to account for non-participating schools in the sample. If you were to multiply WGTFAC1 by WGTADJ1 you would obtain the sampling weight for the school, adjusted for non-participation.

SCHWGT School-level Weight

The school sampling weight is the inverse of the probability of selection for the school, multiplied by its corresponding adjustment factor. It is computed as the product of WGTADJ1 and WGTFAC1. Although this weight variable can be used to estimate the number of schools with certain characteristics, it is important to keep in mind that the sample selected for TIMSS is a good sample of students, but not necessarily an optimal sample of schools. Schools are selected with probability proportional to their size, so it is expected that there is a greater number of large schools in the sample. For countries that sampled by track within school, the SCHWGT is based on the track size rather than the total school size. This may lead to invalid school-weighted analyses.

STOTWGTL Sum of the Student Weights for the Lower Grade

STOTWGTU Sum of the Student Weights for the Upper Grade

These variables are the sums of the weights of the students in the corresponding grade level within the school. If there are no students at this grade level, then the variable is set to zero.

5.7 Other Sampling Variables Included in the Student and Student-Teacher Link Files

With complex sampling designs that involve more than simple random sampling, as in the case of TIMSS where a multi-stage cluster design was used, there are several methods for estimating the sampling error of a statistic that avoid the assumption of simple random sampling. One such method is the jackknife repeated replication (JRR) technique (Wolter, 1985). The particular application of the JRR technique used in TIMSS is termed a paired selection model because it assumes that the sampled population can be partitioned into strata, with the sampling in each stratum consisting of two primary sampling units (PSU), selected independently.

The following variables capture the information necessary to estimate correct standard errors using the JRR technique.

JKZONE

The variable JKZONE indicates the sampling zone or stratum to which the student's school is assigned. The sampling zones can have values from 1 to 75 in the Student Background and Student Achievement data files. This variable is included in the Student Background and the Student Achievement data files.

JKREP

The variable JKREP indicates the PSU and its value are used to determine how the student is to be used in the computation of the replicate weights. This variable can have values of either 1 or 0. Those student records with a value of 0 should be excluded from the corresponding replicate weight, and those with a value of 1 should have their weights doubled. This variable is included in the Student Background and the Student Achievement data files. For each individual student, this variable is identical in these two files.

Additionally, the variables JKCZONE and JKCREP are included in the school file.

JKCZONE

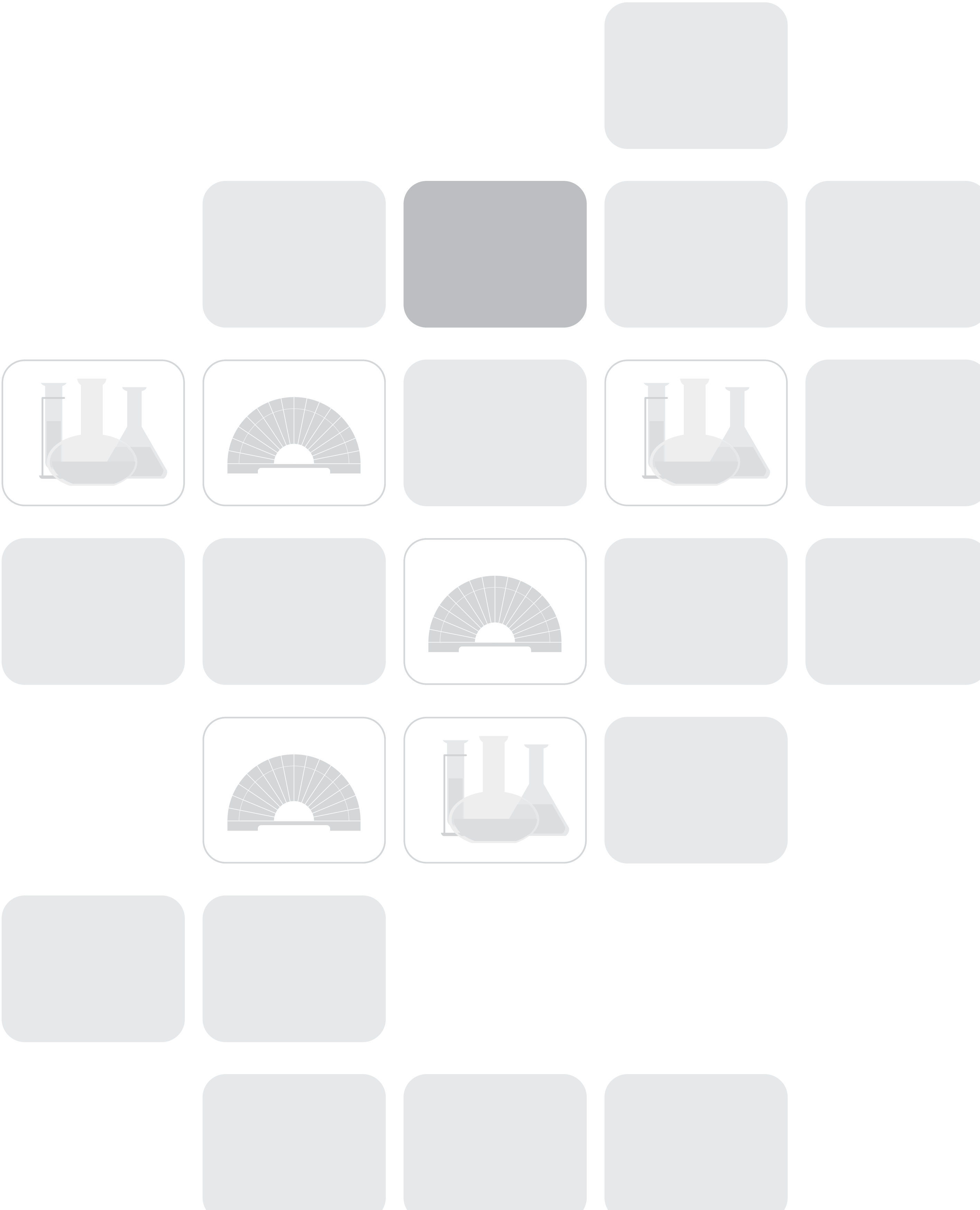
The variable JKCZONE indicates the sampling zone to which the school is assigned. The sampling zones can have values from 1 to 75 in the School data files. Schools selected with certainty receive the value of 0 for this variable.

JKCREP

The variable JKCREP can have values of either 1 or 0. It indicates whether this school is to be dropped or have its weight doubled when estimating standard errors. Those school records with a value of 0 should be excluded from the corresponding replicate weight, and those with a value of 1 should have their weights doubled.

5.8 Using Sampling Variables in the Analysis

It is necessary to use the information captured by the sampling variables whenever you are conducting analyses using the TIMSS data. Chapter 7 describes how to use these variables and information, and Chapters 9 and 10 provide further examples.



6 Scaling Methodology and Achievement Scores

6.1 Overview

The TIMSS achievement test design makes use of matrix-sampling techniques to divide the assessment item pool so that each sampled student responds to just a portion of the items, thereby achieving wide coverage of the mathematics and science subject areas while keeping the response burden on individual students to a minimum[1]. TIMSS relies on a sophisticated form of psychometric scaling known as IRT (Item Response Theory) scaling to combine the student responses in a way that provides accurate estimates of achievement. The TIMSS IRT scaling uses the multiple imputation or "plausible values" method to obtain proficiency scores in mathematics and science and their content areas for all students, even though each student responded to only a part of the assessment item pool.

The scaling methodology and procedures for TIMSS are described in detail by Yamamoto and Kulick (2000). What follows in the present chapter is a brief summary of the scaling, together with a description of the achievement scores available in the TIMSS database.

6.2 TIMSS 1999 Scaling Methodology

Three distinct scaling models, depending on item type and scoring procedure, were used in the analysis of the TIMSS assessment data. A three-parameter model was used with multiple-choice items, which were scored as correct or incorrect, and a two-parameter model was used for free-response items with just two response options, which also were scored as correct or incorrect. Since each of these item types has just two response categories, they are known as dichotomous items. A partial credit model was used with polytomous free-response items, i.e., those with two or more score points.

Two- and Three- Parameter IRT Models for Dichotomous Items

The fundamental equation of the three-parameter (3PL) model gives the probability that a person, whose proficiency on a scale k is characterized by the unobservable variable θ, will respond correctly to item i:

▼ ▼ ▼

[1] The TIMSS 1999 achievement test design is described in Chapter 2.

(1)

$$P\left(x_i = 1 \middle| \theta_k, a_i, b_i, c_i\right) = c_i + \frac{\left(1 - c_i\right)}{1.0 + \exp\left(-1.7 a_i \left(\theta_k - b_i\right)\right)}$$

where

x_i is the response to item i, 1 if correct and 0 if incorrect;

θ_k is the proficiency of a person on a scale k (note that a person with higher proficiency has a greater probability of responding correctly);

a_i is the slope parameter of item i, characterizing its discriminating power;

b_i is its location parameter, characterizing its difficulty;

c_i is its lower asymptote parameter, reflecting the probability a respondent of very low proficiency has of selecting the correct answer.

The probability of an incorrect response to an item is defined as

(2)

$$P_{i0} \equiv P\left(x_i = 0 \middle| \theta_k, a_i, b_i, c_i\right) = 1 - P_{i1}\left(\theta_k\right)$$

The two-parameter (2PL) model was used for the short free-response items that were scored as correct or incorrect. The form of the 2PL model is the same as the above equations with the c_i parameter fixed at zero.

The IRT Model for Polytomous Items

In TIMSS, free-response items requiring an extended response were scored for partial credit, with 0, 1, and 2 as the possible score levels. These polytomous items were scaled using a generalized partial credit model (Muraki, 1992). The fundamental equation of this model gives the probability that a person with proficiency θ_k on scale k will have, for the i-th item, a response x_i that is scored in the l-th of m_i ordered score categories:

(3)

$$P(x_i = l \mid \theta_k, a_i, b_i, d_{i,1}, \ldots, d_i, m_{i,mi-1}) = \frac{\exp\left[\sum_{v=0}^{l} 1.7 a_i \left(\theta_k - b_i + d_{i,v}\right)\right]}{\sum_{g=0}^{m-1} \exp\left[\sum_{v=0}^{g} 1.7 a_i \left(\theta_k - b_i + d_{i,v}\right)\right]} = P_{il}(\theta_k)$$

where

m_i is the number of response categories for item i;

x_i is the response to item i, possibilities ranging between 0 and m_i-1;

θ_k is the proficiency of a person on a scale k;

a_i is the slope parameter of item i, characterizing its discrimination power;

b_i is its location parameter, characterizing its difficulty;

$d_{i,l}$ is category l threshold parameter.

The indeterminacy of model parameters of the polytomous model are resolved by setting $d_{i,0}$ =0 and setting

(4)

$$\sum_{l=1}^{m_i-1} d_{i,l} = 0.$$

In TIMSS analyses, estimates of both dichotomous and polytomous item parameters were obtained by the NAEP BILOG/PARSCALE program, which combines Mislevy and Bock's (1982) BILOG and Muraki and Bock's (1991) PARSCALE computer programs. The item parameters in each scale were estimated independently of the parameters of other scales. Once items were calibrated in this manner, a likelihood function for the proficiency θ_k was induced from student responses to the calibrated items. This likelihood function for the proficiency θ_k is called the posterior distribution of the θs for each respondent.

Scaling Mathematics and Science Domains and Content Areas

In order to estimate student proficiency scores in TIMSS for the subject domains of mathematics and science, all items in each subject domain were calibrated together. This approach was chosen because it produced the best summary of student proficiency across the whole domain for each subject. Treating the entire mathematics or science item pool as a single domain maximizes the number of items per respondent, and the greatest amount of information possible is used to describe the proficiency distribution. This was found to be a more reliable way to compare proficiency across countries than to make a scale for each of the content areas such as algebra, geometry, etc., and then form a composite measure of mathematics by combining the content area scales. The domain-scaling approach was also found to be more reliable for assessing change from TIMSS 1995 to TIMSS 1999.

A disadvantage of this approach is that differences in content scales may be under-emphasized as they tend to regress toward the aggregated scale. Therefore, to enable comparisons of student proficiency on content scales, TIMSS also provides separate scale scores of each content area in mathematics and science. If each content area is treated separately when estimating item parameters, differential profiles of content area proficiency can be examined, both across countries and across subpopulations within a country.

Omitted Responses and Not-Reached Items

Some missing data occurs by design in cases where items were not administered to a student. Missing data could also occur because a student did not answer an item because the student did not know the answer, omitted an item by mistake, or did not have time to attempt the item. In TIMSS, not-reached items were treated differently in estimating item parameters and in generating student proficiency scores. In estimating the values of the item parameters, items that were considered not to have been reached by students were treated as if they had not been administered. This approach was optimal for parameter estimation. However, since the time allotment for the TIMSS tests was generous, and enough for even marginally able respondents to complete the items, not-reached items were considered to have incorrect responses when student proficiency scores were generated.

Proficiency Estimation Using Plausible Values and Conditioning

Most cognitive skills testing is concerned with accurately assessing the performance of individual respondents for the purposes of diagnosis, selection, or placement. Regardless of the measurement model used, classical test theory or item response theory, the accuracy of these measurements can be improved - that is, the amount of measurement error can be reduced - by increasing the number of items given to the individual. Thus, it is common that achievement tests designed to provide information on individual students contain more than 70 items. Since the uncertainty associated with each θ in such tests is negligible, the distribution of θ or the joint distribution of θ with other variables can be approximated using individual θ's.

For the distribution of proficiencies in large populations, however, more efficient estimates can be obtained from a matrix-sampling design like the one used in TIMSS. This design solicits relatively few responses from each sampled respondent while maintaining a wide range of content representation by aggregating responses across all respondents. With this approach, however, the advantage of estimating population characteristics more efficiently is offset by the inability to make precise statements about individuals. The uncertainty associated with individual θ estimates becomes too large to be ignored. In this situation, aggregations of individual student scores can lead to seriously biased estimates of population characteristics (Wingersky, Kaplan, & Beaton, 1987).

Plausible values methodology was developed as a way to address this issue by using all available data to estimate directly the characteristics of student populations and subpopulations, and then generate imputed scores or plausible values from these distributions that can be used in analyses with standard statistical software. A detailed review of plausible values methodology is given in Mislevy (1991).[2]

The following is a brief overview of the plausible values approach. Let y represent the responses of all sampled students to background questions or background data of sampled students collected from other sources, and let $\underline{\theta}$ represent the vector of proficiency values of interest. If $\underline{\theta}$ were known for all sampled students, it would be possible to compute a statistic $t(\underline{\theta},\underline{y})$ - such as a sample mean or sample percentile point - to estimate a corresponding population quantity T.

Because of the latent nature of the proficiency, however, θ values are not known even for sampled respondents. The solution to this problem is to follow Rubin (1987) by considering $\underline{\theta}$ as "missing data" and approximate $t(\underline{\theta},\underline{y})$ by its expectation given $(\underline{x},\underline{y})$, the data that actually were observed, as follows:

(5)

$$\begin{aligned} t^*(\underline{x},\underline{y}) &= E\left[t(\underline{\boldsymbol{\theta}},\underline{y})\middle|\underline{x},\underline{y}\right] \\ &= \int t(\underline{\boldsymbol{\theta}},\underline{y})p\left(\underline{\boldsymbol{\theta}}\middle|\underline{x},\underline{y}\right)d\boldsymbol{\theta}. \end{aligned}$$

It is possible to approximate t^* using random draws from the conditional distribution of the scale proficiencies given the student's item responses x_j, the student's background variables y_j, and model parameters for the student. These values are referred to as imputations in the sampling literature, and as plausible values in large-scale surveys such as National Assessment of Educational Progress (NAEP), National Adult Literacy Study (NALS), and International Adult Literacy and Life Skills Survey (IALLS). The value of θ for any respondent that would enter into the computation of t is thus replaced by a randomly selected value from his or her conditional distribution. Rubin (1987) proposed repeating this process several times so that the uncertainty associated with imputation can be quantified by "multiple imputation." For example, the average of multiple estimates of t, each computed

▼ ▼ ▼

2 Along with theoretical justifications, Mislevy (1991) presents comparisons with standard procedures, discusses biases that arise in some secondary analyses, and offers numerical examples.

from a different set of plausible values, is a numerical approximation of t^* of the above equation; the variance among these values reflects uncertainty due to not observing $\underline{\theta}$. It should be noted that this variance does not include the variability of sampling from the population.

Plausible values for each respondent j are drawn from the conditional distribution $P(\underline{\theta}_j \mid \underline{x}_j, \underline{y}_j, \Gamma, \Sigma)$,where Γ is a matrix of regression coefficients for the background variables, and Σ is a common variance matrix for residuals. Using standard rules of probability, the conditional probability of proficiency can be represented as

(6)

$$\begin{aligned} P(\underline{\theta}_j \mid \underline{x}_j, \underline{y}_j, \Gamma, \Sigma) &\propto P(\underline{x}_j \mid \underline{\theta}_j, \underline{y}_j, \Gamma, \Sigma) P(\underline{\theta}_j \mid \underline{y}_j, \Gamma, \Sigma) \\ &= P(\underline{x}_j \mid \underline{\theta}_j) P(\underline{\theta}_j \mid \underline{y}_j, \Gamma, \Sigma) \end{aligned}$$

where $\underline{\theta}_j$ is a vector of scale values, $P(\underline{x}_j \mid \underline{\theta}_j)$ is the product over the scales of the independent likelihoods induced by responses to items within each scale, and $P(\underline{\theta}_j \mid \underline{y}_j, \Gamma, \Sigma)$ is the multivariate joint density of proficiencies of the scales, conditional on the observed value y_j of background responses and parameters Γ and Σ. Item parameter estimates are fixed and regarded as population values in the computations described in this section.

A multivariate normal distribution was assumed for $P(\underline{\theta}_j \mid \underline{y}_j, \Gamma, \Sigma)$, with a common variance, Σ, and with a mean given by a linear model with regression parameters, Γ. Since in large-scale studies like TIMSS where there are many hundreds of background variables, it is customary to conduct a principal components analysis to reduce the number to be used in Γ. Typically, components representing 90% of the variance in the data are selected. These principal components are referred to as the conditioning variables and denoted as y^c. The following model is then fit to the data.

(7)

$$\underline{\theta} = \Gamma' \underline{y}^c + \underline{\varepsilon},$$

where $\underline{\varepsilon}$ is normally distributed with mean zero and variance Σ. As in a regression analysis, Γ is a matrix in which each of the columns is the effects for each scale and Σ is the matrix of residual variance between scales.

Generating Proficiency Scores

After completing the EM algorithm, and estimating Γ and Σ, the plausible values were drawn in a three-step process from the joint distribution of the values of Γ for all sampled. First, a value of Γ was drawn from a normal approximation to $P(\Gamma,\Sigma | x_j,y_j)$ that fixed Σ at the value $\hat{\Sigma}$ (Thomas, 1993). Second, conditional on the generated value of Γ (and the fixed value of $\Sigma = \hat{\Sigma}$), the mean θ, and variance Σ_j^p of the posterior distribution in equation (2) were computed. In the third step, the proficiency values were drawn independently from a multivariate normal distribution with mean θ and variance Σ_j^p. These three steps were repeated five times, producing five imputations of θ for each sampled respondent.

Creating IRT Scales for Mathematics and Science Content Areas for 1995 and 1999 Data

The primary function of the IRT scales in the mathematics and science content areas is to portray student achievement in each country as a profile of performance in each area, relative to student performance in other areas. Such profiles should, for example, show countries where performance in algebra was relatively better than in geometry, or better in life science than in chemistry. Although it would have been desirable to establish a link between 1995 performance and 1999 performance in each content area, there were not enough common items in the two assessments to do this reliably. However, the numbers of items in each content area were considered sufficient to develop separate content area scales for each assessment. The five content areas in mathematics and six areas in science for which scales were developed are presented in Exhibit 6.1.

Exhibit 6.1 Number of Items in Mathematics and Science Content Areas (1995 and 1999 Combined)

Mathematics Content Area	Number of Items	Science Content Area	Number of Items
Fractions/Number	104	Earth Science	34
Measurement	39	Life Science	70
Data Representation	33	Physics	64
Geometry	41	Chemistry	39
Algebra	57	Environmental and Resource Issues	17
		Scientific Inquiry and the Nature of Science	12
Total	274	Total	236

The calibration samples used for the joint 1995-1999 scaling were also used to estimate the item parameters for each of the content area scales. The principal components produced for the conditioning of the joint 1995-1999 mathematics and science domain scales were used for the 1999 content area plausible value analyses as well. Plausible values were generated for all countries for both assessments using the new, jointly estimated item parameters under multivariate conditions.

The indeterminacy of the content area scales in mathematics was resolved by setting the mean of each mathematics content area scale over all of the 38 TIMSS 1999 countries equivalent to the mean of the domain scale for mathematics. The same approach was taken for science.

It should be noted that since there were far fewer items in each content area scale than in the domain scales (for example, 57 algebra items compared with 274 mathematics items); and a relatively greater proportion of the variance in the content area scales was due to measurement error. In the scaling, the total variance for content area scales and domain scales was set to be equal, and therefore the measurement error plays a relatively greater role in the variance of the content area scales. This implies that the content area scale means of each country tend to be regressed toward the grand mean, and that the regression is more noticeable for very high- or very low-achieving countries.

6.3 Student Achievement Scores

The TIMSS international database contains several student-level achievement scores, including the plausible values described in this chapter. These scores were computed at different stages of the study to serve specific purposes. This chapter presents a description of these achievement scores, how they were derived, how they were used by TIMSS, and how users of the database can use them. For identification purposes, the first letter for the variable name identifies the population for which the score was computed. The scores computed for Population 2 have the letter B as the first character. This convention has been followed with other background and derived variables and with the files included in the database.

Achievement Scores in the Student Files

Four types of achievement scores are included in the student data files: plausible values, raw scores, standardized raw scores, and national Rasch scores. Each type is described below.

Mathematics and Science Scores: Plausible Values

As described earlier in this chapter, TIMSS made use of multiple imputation or "plausible values" methodology to provide estimates of student proficiency in mathematics and science. Because of the error involved in the imputation process, TIMSS produced not one but five imputed values for each student. The original TIMSS scaling in 1995 produced five plausible values for mathematics overall and science overall. The revised scaling in 1999 produced five plausible values for each of the content area subscales within each subject area as well as overall scores for mathematics and science. The plausible values available for each of the scales are listed in Table 6.2.

The need for plausible values arises from the fact that any student was administered only a fraction of the items in the assessment, as described in Chapter 2. Time constraints did not allow for all the items to be administered to each student. A plausible value is an estimate of how the individual student would have performed on a test that included all possible items in the assessment. Since no student responded to all items, this estimate is based on the responses to the items that were included in the test booklet that the student actually took and the performance of students with similar characteristics based on their responses to the background questionnaire.

Overall mathematics and science plausible values were standardized so that the mean of the 8^{th} graders in the 1995 sample is equal to 500 and the standard deviation equals 100. These scores were computed across years. This is not the case for the subscale scores. In the 1995 sample, the subscales scores were standardized to have a mean of 500 and a standard deviation of all 8^{th} graders. In the 1999 sample, it was not possible to link these to the 1995 scales so the mean and standard deviation were set to the overall mean and standard deviation of the overall scale, respectively. The subscale scores are not comparable across assessment years.

Within a subscale or subject area and across the sample, one set of plausible values can be considered as good as another. Each of these sets is equally well designed to estimate population parameters, although the estimates will differ somewhat. This difference is attributable to imputation variance, which contributes to the uncertainty of the estimate. Five sets of plausible values are provided so that analyses can be replicated as many as five times. Results which vary considerably from replication to replication may be influenced by unreliability in the achievement measures, and considered to be suspect. When reporting international achievement all five plausible values are used. Chapter 7 details the procedure by which the information from the five plausible values is summarized. The plausible values are included in the Student Background, Student Achievement, and Student-Teacher Link files.

Exhibit 6.2 Plausible Values Based on TIMSS Content Area Subscales

Variable[1]	Content Area Subscale
BSMMAT01 – BSMMAT05	Mathematics Overall
BSMDAP01 – BSMDAP05	Data Representation, Analysis and Probability
BSMALG01 – BSMALG05	Algebra
BSMFNS01 – BSMFNS05	Fractions and Number Sense
BSMGEO01 – BSMGEO05	Geometry
BSMMEA01 – BSMMEA05	Measurement
BSSSCI01 – BSSSCI05	Science Overall
BSSEAS01 – BSSEAS05	Earth Science
BSSNOS01 – BSSNOS05	Scientific Inquiry and the Nature of Science
BSSPHY01 – BSSPHY05	Physics
BSSCHE01 – BSSCHE05	Chemistry
BSSERI01 – BSSERI05	Environmental and Resource Issues

1 Scientific Inquiry and the Nature of Science and Environmental and Resource Issues subscale plausible values were computed with the 1999 data only.

Raw Scores

BSMSCPT

Number of raw score points obtained on the mathematics items.

BSSSCPT

Number of raw score points obtained on the science items.

After the items were recoded as right or wrong, or adjusted to their level of correctness in the case of the open-ended items, raw scores were computed by adding the number of points obtained across the items for a subject. Multiple-choice items received a score of either 1 or 0. Open-ended response items received score points from 0 to 3 depending on their scoring guide. The value of the first digit of the code determines the number of score points assigned to an open-ended item. Open-ended items scored with a first digit of 7 or 9, indicating an incorrect/incomplete answer, were given zero points. A description of the algorithm used to score the items can be found in Chapters 9 and 10 in the section "Scoring the Items."

Although these scores can be used to compare students' performances on the same booklet in the same year, they should not be used to compare students' performances across different booklets or across years. Different booklets contain different numbers of items for each subject, and the specific items contained in one booklet had varying difficulties. It is recommended that these scores be used only to verify whether the items have been recoded correctly when a user decides to recode the items to their level of correctness. Raw scores for the individual students can be found in the Student Background and Student Achievement data files.

Standardized Raw Scores

BSMSTDR

Standardized mathematics raw score.

BSSSTDR

Standardized science raw score.

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 year in preliminary analyses. The standardized score was computed so that the weighted mean score within each booklet in a year was equal to 50, and the weighted standard deviation was equal to 10.

The standardized raw scores were used in the initial item analysis for computing the discrimination coefficients for each of the items in the test. This initial item analysis was conducted prior to scaling the test items. The standardized raw scores for the individual students can be found in the Student Background and Student Achievement data files.

National Rasch Scores

BSMNRSC

National Rasch Mathematics Score (ML).

BSSNRSC

National Rasch Science Score (ML).

The national Rasch scores were also computed for preliminary analyses. These provided a basic Rasch score for preliminary analyses within countries, but cannot be used for international comparisons since each country has been assigned the same mean score. The national Rasch scores were computed by standardizing mathematics and science logit scores to have a weighted mean of 150 and a standard deviation of 10 within each country. The logit scores were computed using the Quest Rasch analysis software; Quest provides maximum likelihood (ML) estimates of a scaled score, based on the one-parameter Rasch model, for the performance of students on a set of items. The computation took into account the varying difficulty of the items across test booklets, and the performance and ability of the students responding to each set of items. These logit scores were obtained using item difficulties that were computed for each country using all available item responses for the country and centering the item difficulty around zero. When computing the item difficulties, responses marked as “not reached” were treated as items that were not administered. This avoided giving inflated item difficulties to the items located at the end of the test in cases where students systematically do not reach the end of the test. These item difficulties were then used to compute logit scores for each student.

When computing the student logit scores the responses marked as “not reached” were treated as incorrect responses. This avoided unfairly favoring students who started answering the test and stopped as soon as they did not know the answer to a question. Logit scores for the students generally ranged between -4 and +4. When using maximum likelihood estimation procedures it is not possible to obtain finite logit scores for students who correctly answered all or none of the items. Scores for these students were set to +5 and -5 logits, respectively. These logit scores were then standardized to have a weighted mean of 150 and a standard deviation of 10 for each country.

The national Rasch scores should not be used for international comparisons for two reasons: they were computed with a different set of item difficulties for each country, and the weighted mean score within each country is always equal to 150. National Rasch scores can be found in the Student Background and Student Achievement data files.

6.4 International and National Benchmarks of Achievement

In order to provide more information about student achievement, TIMSS in 1999 identified four points on the mathematics and science scales for use as international benchmarks. The Top 10% Benchmark was defined as the 90th percentile on the TIMSS scale, computed across all students in all participating countries, with countries weighted in proportion to the size of their eighth-grade population. This point on each scale (mathematics and science) is the point above which the top 10 percent of students in the 1999 TIMSS assessment scored. The Upper Quarter Benchmark is the 75th percentile on the scale, above which the top 25 percent of students scored. The Median Benchmark is the 50th percentile, above which the top half of students scored. Finally, the Lower Quarter Benchmark is the 25th percentile, the point reached by the top 75 percent of students.

If all countries had the same distribution of student achievement, approximately 10 percent of students within each country would be above the 90th percentile in the international distribution, regardless of the country's population size.

Because of the imputation technology used to derive the proficiency scores, the international benchmarks had to be computed once for each of the five plausible values and the results averaged to arrive at the final figure. The international benchmarks are presented in Exhibits 6.3 and 6.4 for mathematics and science, respectively.

Exhibit 6.3 TIMSS 1999 International Benchmarks of Mathematics Achievement for the Eighth Grade

Proficiency Score	25th Percentile	50th Percentile	75th Percentile	90th Percentile
Plausible Value 1	396.86	479.20	554.49	615.15
Plausible Value 2	395.76	478.79	554.74	615.37
Plausible Value 3	395.62	478.56	554.83	616.23
Plausible Value 4	394.57	478.09	554.03	615.02
Plausible Value 5	396.30	479.10	554.56	615.76
Mean Plausible Value	395.82	478.75	554.53	615.51

Exhibit 6.4 TIMSS 1999 International Benchmarks of Science Achievement for the Eighth Grade

Proficiency Score	25th Percentile	50th Percentile	75th Percentile	90th Percentile
Plausible Value 1	409.03	487.76	558.66	617.01
Plausible Value 2	409.87	487.61	557.60	615.88
Plausible Value 3	410.38	488.04	557.27	616.12
Plausible Value 4	410.05	487.54	557.47	615.82
Plausible Value 5	410.87	487.59	557.79	615.88
Mean Plausible Value	410.04	487.71	557.76	616.14

The database for the 1999 assessment contains a set of variables that indicate which international benchmark the student met. Because of the plausible value technology used in TIMSS, five variables are included for each subject indicating, for each plausible value, which benchmark was reached. These variables are the following:

BSMIBM01-5

International Mathematics Benchmark reached with each of the five plausible values.

BSSIBM01-5

International Science Benchmark reached with each of the five plausible values.

The following codes are used for these variables:

Code 1	Student performed below the 25^{th} international percentile.
Code 2	Student performed at or above the 25^{th} international percentile, but below the 50^{th} international percentile.
Code 3	Student performed at or above the 50^{th} international percentile but below the 75^{th} international percentile.
Code 4	Student performed at or above the 75^{th} international percentile but below the 90^{th} international percentile.
Code 5	Student performed at or above the 90^{th} international percentile.

Additionally, four points on the mathematics and science scales were identified within each country for use as national benchmarks. The Top 10% national benchmark for each country was defined as the 90^{th} percentile on the TIMSS scale, computed across all students within each country. This point on each scale (mathematics and science) is the point above which the top 10% of students within the country scored. The upper quarter national benchmark is the 75^{th} percentile on the scale, above which the top 25% of students within the country scored. The median national benchmark is the 50^{th} percentile, above which the top half of students within the country scored. Finally, the lower quarter national benchmark is the 25^{th} percentile, the point reached by the top 75% of students within the country.

Because of the nature of these national benchmarks, approximately 10% of students within each country would be above the 90th percentile in the national distribution, 25% would be above the 75^{th} percentile, and so on. Because of the imputation technology used to derive the proficiency scores, the national benchmarks also had to be computed once for each of the five plausible values.

The database for the 1999 assessment also contains a set of variables that indicate the national benchmarks me by each student. Because of the plausible value technology used in TIMSS, five variables are included for each subject indicating, for each plausible value, which national benchmark was reached. These variables are the following:

BSMNBM01-5

National Mathematics Benchmark reached with each of the five plausible values

BSSNBM01-5

National Science Benchmark reached with each of the five plausible values

The following codes are used for these variables:

Code 1	Student performed below the 25th national percentile.
Code 2	Student performed at or above the 25th national percentile, but below the 50th national percentile.
Code 3	Student performed at or above the 50th national percentile but below the 75th national percentile.
Code 4	Student performed at or above the 75th national percentile but below the 90th national percentile.
Code 5	Student performed at or above the 90th national percentile.

7 Estimating Sampling and Imputation Variance

7.1 Overview

When analyzing data from complex designs such as TIMSS, it is important to compute correct error variance estimates for the statistics of interest. In TIMSS this error variance comes from two sources: the sampling process and the imputation process. This chapter describes the methods used to estimate each of these error variance components.

7.2 Estimating Sampling Variance

With complex sampling designs that involve more than simple random sampling, as in the case of TIMSS where a multi-stage cluster design was used, there are several methods for estimating the sampling error of a statistic that avoid the assumption of simple random sampling. One such method is the jackknife repeated replication (JRR) technique (Wolter, 1985). The particular application of the JRR technique used in TIMSS is termed a paired selection model because it assumes that the sampled population can be partitioned into strata, with the sampling in each stratum consisting of two primary sampling units (PSU), selected independently. Following this first-stage sampling, there may be any number of subsequent stages of selection that may involve equal or unequal probability selection of the corresponding elements. The TIMSS design called for a total of 150 schools for the target population. These schools constituted the PSUs in most countries, and were paired sequentially after sorting by a set of implicit stratification variables. This resulted in the implicit creation of 75 strata, with two schools selected per stratum.

The jackknife repeated replication (JRR) method is suitable for estimating sampling errors in the TIMSS design because it provides approximately unbiased estimates of the sampling error arising from the complex sample selection procedure for estimates such as percentages, totals and means. In addition, this method can also be readily adapted to the estimation of sampling errors for parameters estimated using other statistical modeling procedures, such as percent-correct technology. The general use of the JRR entails systematically assigning pairs of schools to sampling zones, and the random selection of one of these schools to have its contribution doubled, and the other zeroed, so as to construct a number of "pseudo-replicates" of the original sample. The statistic of interest is computed once for all of the original sample, and once more for each of the pseudo-replicate samples. The variation between the original sample estimate and the estimates from each of the replicate samples is the jackknife estimate of the sampling error of the statistic.

Computing Sampling Variance Using the JRR Method

When implementing the JRR method in TIMSS, for each country it was assumed that there were up to 75 strata or zones (H) within each country, each one containing two sampled schools selected independently. When computing a statistic "*t*" from the sample for a country, the formula for the JRR variance estimate of the statistic *t* is then given by the following equation:

(1)

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

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

As can be seen from the above formula, the computation of the JRR variance estimate for any statistic from the TIMSS database requires the computation of any statistic up to 76 times for any given country: once to obtain the statistic for the full sample, and up to 75 times to obtain the statistics for each of the jackknife replicates (J_h). The number of times a statistic needs to be computed for a given country will depend on the number of implicit strata or sampling zones defined for the sample.

Note that when using the JRR technique for the estimation of sampling variability, the approach will appropriately reflect the combined effect of the between- and within-sampling zone contributions to the variance.

Doubling and zeroing the weights of the selected units within the strata or "zones" is accomplished effectively with the creation of replicate weights that are then used in the calculations. Chapters 9 and 10 show how this approach allows standard statistical software such as SAS or SPSS to be used to compute JRR estimates of sampling variability. The replicate weight approach requires you to temporarily create a new set of weights for each pseudo-replicate sample. Each replicate weight is equal to *k* times the overall sampling weight, where *k* can take values of zero, one, or two, depending on whether or not the case is to be removed from the computation, left as it is, or have its weight doubled. The value of k for an individual student record for a given replicate depends on the assignment of the record to the specific PSU and zone.

Construction of Sampling Zones for Sampling Variance Estimation

An important step in applying the JRR technique to the estimation of sampling variability consists of assigning the schools to implicit strata known as sampling zones. Most TIMSS sample designs in the participating countries called for a total of 150 sampled schools per target population. Each of these 150 schools was assigned to one of 75 sampling zones. These zones were constructed by pairing the sequentially sampled schools and assigning these pairs to a sampling zone. Since schools were generally sorted by a set of implicit stratification variables, the resulting assignment of sequentially sampled schools to sampling zones takes advantage of any benefit due to this implicit stratification. In cases when more than 75 pairs of schools were sampled within a country, schools were then assigned to sampling zones in a way such that no more than 75 sampling zones were defined. In some cases this meant assigning more than two schools to the same sampling zone.

Sampling zones were constructed within explicit strata. In cases when there was an odd number of schools in an explicit stratum, either by design or because of school-level non-response, the students in the remaining school were randomly divided into two "quasi" schools for the purposes of calculating the jackknife standard error. Each zone then consisted of a pair of schools or "quasi" schools. When computing replicate weights for the estimation of JRR sampling error, one member of each pair of schools is randomly selected to have its weights doubled, while the weights of the other member are set to zero.

The variable JKZONE indicates the sampling zone to which the student's school is assigned. The sampling zones can have values from 1 to 75 in the Student Background and Student Achievement data files.

The variable JKREP indicates how the student is to be used in the computation of the replicate weights. This variable can have values of either 1 or 0. Those student records with a value of 0 should be excluded from the corresponding replicate weight, and those with a value of 1 should have their weights doubled. This variable is included in the Student Background and the Student Achievement data files.

Computing the JRR Replicate Weights

Having assigned the schools to zones, if it is desired to use standard statistical software such as SAS or SPSS for sampling variance estimation, a convenient way to proceed is to construct a set of replicate weights for each pseudo-replicate sample. In TIMSS, the schools in the sample were assigned in pairs to one of the 75 zones indicated by the variable JKZONE, and within each zone the pair members were randomly assigned an indicator (u_i) represented by the variable JKREP, coded randomly to 1 or 0 so that one of the members of each pair had values of 1 on the variable u_i, and the remaining one a value of 0. This indicator determined whether the weights for the elements in the school in this zone were to be doubled or zeroed.[1] The replicate weight ($W_h^{g,i,j}$) for the elements in a school assigned to zone h is computed as the product of k_h times their overall sampling weight, where k_h could take values of zero, one, or two depending on if the case was not to contribute, be left unchanged, or have it counted twice for the computation of the statistic of interest. In the TIMSS database, the replicate weights are not permanent variables. They need to be created by the sampling variance estimation program. An example program that makes use of replicate weights in computing JRR estimates is provided in Chapters 9 and 10.

When creating the replicate weights the following procedure is followed:

1. Each sampled student is assigned a vector of 75 weights or $W_h^{g,i,j}$, where h takes values from 1 to 75.
2. The value of $W_0^{g,i,j}$ is the overall sampling weight which is simply the product of the final school weight, the appropriate final classroom weight, and the appropriate final student weight as defined in Chapter 3.
3. The replicate weights for a single case are then computed as:

(2)

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

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

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

▼ ▼ ▼

[1] Please note that, depending on the program, this information can be used differently. For example, in the examples described in this User Guide, students with a value of JKREP = 0 have their weights zeroed and those with value of 1 have their weights doubled. Programs such as Wesvar do the inverse. Both yield unbiased estimates.

7.3 Estimating Imputation Variance

The general procedure for estimating the imputation variance using plausible values is as follows: first estimate the statistic (t), each time using a different set of the plausible values (M). The statistics t_m can be anything estimable from the data, such as a mean, the difference between means, percentiles, etc. If all of the (M=5) plausible values in the TIMSS database are used, the parameter will be estimated five times, once using each set of plausible values. Each of these estimates will be called t_m, where m=1,2,...,5.

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

(3)

$$Var_{imp} = \left(1 + \frac{1}{M}\right) \times Var\left(t_m\right)$$

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

7.4 Combining Sampling and Imputation Variance

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

(4)

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

where $Var_{jrr}(t_1)$ is the sampling variance for the first plausible value.

This User Guide contains programs in SAS and SPSS that compute each of these variance components for the TIMSS data.

7.5 Achievement Differences Across Countries and Multiple Comparisons

An aim of the TIMSS 1999 international database is to provide data to do fair and accurate comparisons of student achievement across the participating countries. In comparisons of performance across countries, standard errors are necessary to assess the statistical significance of the difference between the summary statistics.

For differences between countries, which can be considered as independent samples, the standard error of the difference in means is computed as the square root of the sum of the squared standard errors of each mean:

(5)

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

where se_1 and se_2 are the standard errors of the means.

However, when using the data to make multiple simultaneous comparisons between countries, it will be necessary to make additional adjustments.

As an example, the multiple comparison charts presented in the TIMSS 1999 international reports allow the comparison of average performance of a country with that of other participating countries. The significance tests reported in these charts include a Bonferroni adjustment for multiple comparisons that holds to 5 percent the probability of erroneously declaring the mean of one country to be different from that of another country. The Bonferroni adjustment is necessary because that probability greatly increases as the number of simultaneous comparisons increases.

If repeated samples were taken from two populations with the same mean and variance, and in each one the hypothesis that the two means are significantly different at the α= .05 level (i.e., with 95% confidence) was tested, then it would be expected in about 5 percent of the comparisons significant differences would be found between the sample means even though no difference exists in the populations. The probability of finding significant differences when none exist (the so-called type I error) is given by α= .05. Conversely, the probability of not making such an error is 1 - α, which in the case of a single test is .95. However, comparing the means of three countries involves three tests (country A versus country B, country B versus country C, and country A versus country C). Since these are independent tests, the probability of not making a type I error in any of these tests is the product of the individual probabilities, which is $(1 - \alpha)(1 - \alpha)(1 - \alpha)$. With α= .05, the overall probability of not making a type I error is only .873, which is considerably less than the probability for a single test. As the number of tests increases, the probability of not making a type I error decreases rapidly, and conversely, the probability of making such an error increases.

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

The TIMSS 1999 international reports contain multiple comparison exhibits that show the statistical significance of the differences between all possible combinations of the 38 participating countries. There were (38*37)/2 = 703 possible differences. In the Bonferroni procedure the significance level (α) of a statistical test is adjusted by the number of comparisons that are planned and then looking up the appropriate quantile from the normal distribution. In deciding on the appropriate adjustment of the significance level for TIMSS, it was necessary to decide how the multiple comparison exhibits would most likely be used. A very conservative approach would be to adjust the significance level to compensate for all of the 703 possible comparisons among the 38 countries concerned. However, this risks an error of a different kind, that of concluding that a difference in sample means is not significant when in fact there is a difference in the population means.

Since you are likely to be interested in comparing a single country with all 37 other countries, rather than in making all possible between- country comparisons at once, the more realistic approach of using the number of countries minus one to adjust the significance level can be adopted. This means that the number of simultaneous comparisons to be adjusted for would be 37 instead of 703. The critical value for a 95 percent significance test adjusted for 37 simultaneous comparisons is 3.2049, from the appropriate quantiles from the normal (Gaussian) distribution. You will need to adjust accordingly.

7.6 Comparing Achievement with the International Mean

Many of the data exhibits in the TIMSS 1999 international reports show countries' mean achievement compared with the international mean, and you might be interested in replicating some of these analyses or doing some analysis of your own. Since this resulted in 38 simultaneous comparisons, one for each country, the critical value was adjusted to 3.2125 using the Dunn-Bonferroni procedure.

When comparing a country's mean with the international average, TIMSS took into account the fact that the country contributed to the international standard error. To correct for this contribution the sampling component of the standard error of the difference for country j was computed as follows:

(6)

$$se_{s\_dif\_j} = \sqrt{\frac{\left(\left(N-1\right)^2 - 1\right)se_j^2 + \sum_{k=1}^{N} se_k^2}{N}}$$

where $se_{s\_dif\_j}$ is the standard error of the difference due to sampling when country *j* is compared to the international mean. N is the number of countries used to compute the international mean. se_k^2 is the sampling standard error for country *k*. and se_j^2 is the sampling standard error for country *j*.

The imputation component of the standard error was computed by taking the square root of the imputation variance calculated as follows:

(7)

$$se_{s\_dif\_j} = \sqrt{1 + \frac{1}{NPV} Var\left(d_{1\ldots5}\right)}$$

where NPV is the number of plausible values used in the analysis, and d_l is the difference between the international mean and the country mean for plausible value *l*.

Finally, the standard error of the difference is calculated as:

(8)

$$se_{dif\_j} = \sqrt{se_{i\_dif\_j}^2 + se_{s\_dif\_j}^2}$$

7.7 Comparisons between 1999 and 1995

TIMSS 1999 was designed to enable comparisons between a country's achievement and background variables on the 1995 and 1999 assessments. A total of 26 countries participated at the eighth grade in both assessments.[2] When assessing whether eighth-grade achievement had significantly changed from 1995 to 1999, the differences between year, even within the same country, can be considered as differences between independent samples. The standard error of the differences in this case can also be computed as the square root of the sum of the squared standard errors of each statistic.

▼ ▼ ▼

[2] Although all countries had acceptable sampling participation in 1999, three countries – Israel, South Africa, and Thailand – failed to meet sampling guidelines in 1995, and were omitted from the calculation of trends.

Because of differences from 1995 to 1999 in the sampling of student populations, the results for Israel and Italy of trend data differ from those containing just 1999 data. In TIMSS 1995, Israel tested only Hebrew-speaking students, while in 1999 the target population included both Hebrew- and Arab-speaking students. To provide meaningful trend analysis, TIMSS compared 1995 and 1999 using the Hebrew-speaking part of the population only. In Italy, the 1995 assessment sampled students from most but not all provinces, whereas in 1999 all provinces were included. The TIMSS 1999 trend data for Italy represents only those provinces that participated in both years. Trend samples can be identified by means of the variable INTMS95.

7.8 Gender Differences

Because in most countries boys and girls attend the same schools, the samples of boys and girls cannot be treated as independent for the purpose of statistical tests. Accordingly, a jackknife procedure applicable to correlated samples for estimating the standard error of the difference should to be applied. This involves computing the differences between boys and girls once for each of the 75 replicate samples, and five more times, once for each plausible value. These differences are then summarized into an error of the difference using the formula presented earlier in this chapter.

Chapters 9 and 10 include macros that can be easily used to compute these differences by using contrast variables and regression models.

8 Content and Format of Database Files

8.1 Overview

The database contains achievement data and student, teacher, and school background data collected in the 38 countries that participated in TIMSS 1999. It also contains the data for the 42 countries that took part in the 1995 Population 2[1] assessment. In total, the TIMSS 1999 database contains responses of nearly 500,000 students, over 50,000 mathematics and science teachers, and over 12,000 school principals collected between 1995 and 1999.This chapter describes the content and format of the TIMSS 1999 international database. This chapter is organized in seven major sections corresponding to the types of files included in the database. Within each section, the contents of the files are described. These file types are:

Data Files

The TIMSS *Data Files* reflect the result of an extensive series of data management and quality control steps taken to insure the international comparability, quality, accuracy, and general utility of the database in order to provide a strong foundation for secondary analyses. They contain responses to background questionnaires administered to students, their mathematics and science teachers, and the principals of their schools. As part of the international data files, variables derived for reporting in the international reports are included. In addition, analysis notes are provided for other computed variables, allowing users to replicate these computations. These analysis notes are included in Supplement 3 of this User Guide.

Data Almanacs

Data Almanac Files contain weighted summary statistics for each participating country, on each achievement item and each variable in the student, teacher, and school background questionnaires.

Codebook Files

Also included in the database are *Codebook Files*. These specifically document the structure of the data files as well as information about the format of all variables in each of the data files.

Program Files

Several *Program Files* are provided for use in secondary analyses, including Data Access Control files for converting the raw data files provided in ASCII format into SAS or SPSS files, macro programs for computing statistics using the jackknife repeated replication method and using the plausible values discussed in Chapter 7,

▼ ▼ ▼

[1] This corresponds to 7th and 8th grades in most countries.

and macro programs for converting achievement item response codes to score values used in the computation of international scores.

Test-Curriculum Matching Analysis Files

To investigate the match between the TIMSS achievement tests and the curriculum in participating countries, country representatives indicated whether or not each test item addressed a topic in their curriculum. The *Test-Curriculum Matching Analysis Files* contain this information for each country.

Item Information Files

The *Item Information Files* contain information about the cognitive test items including item identifications, mathematics or science content area, performance expectation category, test subject, subscale, item label, format, response key, and maximum points. The following sections describe each of the file types and how they can be used to access and analyze the TIMSS international data for students, teachers, and schools.

8.2 The Data Files

There are four basic types of data files in the TIMSS International Database:

- Background Files with information from students, their mathematics teachers, their science teachers and the principals of their school.
- Achievement Files with responses to the TIMSS test
- Scoring Reliability Files
- Student Teacher-Link Files

These files and the variables contained in each are described below. Data files are provided for each country that participated in TIMSS 1999 or TIMSS Population 2 in 1995 and for which internationally comparable data are available.

Data File and Variable Naming Convention

Before proceeding to describe the different data files, this section will describe the file and variable-naming conventions used in the TIMSS databases. It also describes how the variables are organized in the database.

Data File Naming Convention

The datafiles included with this database are named using the following convention:

- The first character of the files is always "B." This indicates that the file refers to eighth grade in 1999, and seventh and eighth grade (Population 2) in 1995.

- The second character indicates the source or level of the information in the file:

  C indicates a school level file

  T indicates a teacher level file

  S indicates a student level file.

- The third character indicates the subject and/or type of the data in the file. The following abbreviations are used, listed in alphabetical order:

  A Student Achievement Booklets

  G General Background Questionnaires (School and Student Questionnaires)

  M Mathematics Teacher Background Questionnaire

  R Free-Response Scoring Reliability (scoring reliability sample of student test booklets)[2]

  S Science Teacher Background Questionnaire

  T Student-Teacher Link File

- The next three digits identify the country with the three-character alphanumeric country abbreviation following the ISO coding scheme. Exhibit 8.2 shows these ISO codes for each participating country.

- The seventh and eighth character indicates the study cycle

  M1 stands for the 1995 files

  M2 stands for the 1999 files

- The 3 character file extensions used for the data files is .DAT.

For each file type, a separate data file is provided for each participating country. Countries for which data is not available have an empty file. The data files for each country may be identified by a three-character country abbreviation represented by "<country>" in Exhibit 8.1, which summarizes the data files available.

▼▼▼

[2] Since there were no free-response items administered in both the 1995 and 1999 assessments, these data are not included for the 1995 assessment in this release of the TIMSS database. The 1995 scoring reliability data are available from the 1995 release of the TIMSS database (Gonzalez & Smith, 1997).

Exhibit 8.1 TIMSS Data Files

File Type	1995 Files	1999 Files
Student Achievement test File	BSA<Country>M1	BSA<Country>M2
Student Background File	BSG<Country>M1	BSG<Country>M2
Mathematics Teacher Background File	BTM<Country>M1	BTM<Country>M2
Science Teacher Background File	BTS<Country>M1	BTS<Country>M2
School Background File	BCG<Country>M1	BCG<Country>M2
Student – Teacher Linkage File	BST<Country>M1	BST<Country>M2
Student Achievement Scoring Reliability File	BSR<Country>M1	BSR<Country>M2

The three-character abbreviations used for each TIMSS country are listed in Exhibit 8.2, along with a numeric code for each country that is used to identify the country in the data files (see the following section discussing identification variables). Each file that is available within a country is annotated with a "•" in Exhibit 8.2.

Please note that even though the 1995 assessment had a performance assessment component, this was not repeated in 1999 and is not included in this release of the TIMSS international database. If you are interested in looking at these data you will need to obtain the TIMSS 1995 International Database and accompanying User Guide (Gonzalez and Smith, 1997).

Exhibit 8.2 Country Identification and Availability of Data Files

Country	ISO Code	Numeric Code	BSG 95	BSA 95	BTM 95	BTS 95	BST 95	BCG 95	BSG 99	BSA 99	BTM 99	BTS 99	BST 99	BCG 99	BSR 99
Australia	AUS	036	●	●	●	●	●	●	●	●	●	●	●	●	●
Austria	AUT	040	●	●	●	●	●	●							
Belgium (Fl)	BFL	956	●	●	●	●	●	●	●	●	●	●	●	●	●
Belgium (Fr)	BFR	957	●	●	●	●	●	●							
Bulgaria	BGR	100		●					●	●	●	●	●	●	●
Canada	CAN	124	●	●	●	●	●	●	●	●	●	●	●	●	●
Chile	CHL	152							●	●	●	●	●	●	●
Chinese Taipei	TWN	158							●	●	●	●	●	●	●
Colombia	COL	170	●	●	●	●	●	●							
Cyprus	CYP	196	●	●	●	●	●	●	●	●	●	●	●	●	
Czech Republic	CZE	203	●	●	●	●	●	●	●	●	●	●	●	●	●
Denmark	DNK	208	●	●	●	●	●	●							
England	ENG	926	●	●	●	●	●	●	●	●	●	●	●	●	●
Finland	FIN	246							●	●	●	●	●	●	●
France	FRA	250	●	●	●	●	●	●							
Germany	DEU	280	●	●	●	●	●	●							
Greece	GRC	300	●	●	●	●	●	●							
Hong Kong	HKG	344	●	●	●	●	●	●	●	●	●	●	●	●	●
Hungary	HUN	348	●	●	●	●	●	●	●	●	●	●	●	●	
Iceland	ISL	352	●	●	●	●	●	●							
Indonesia	IDN	360							●	●	●	●	●	●	●
Iran, Islamic Rep.	IRN	364	●	●	●	●	●	●	●	●	●	●	●	●	●
Ireland	IRL	372	●	●	●	●	●	●							
Israel	ISR	376	●	●	●	●	●	●	●	●	●	●	●	●	
Italy	ITA	380	●	●	●	●	●	●	●	●	●	●	●	●	●
Japan	JPN	392	●	●	●	●	●	●	●	●	●	●	●	●	●
Jordan	JOR	400							●	●	●	●	●	●	●
Korea	KOR	410	●	●	●	●	●	●	●	●	●	●	●	●	
Kuwait	KWT	414	●	●	●	●	●	●							
Latvia	LVA	428	●	●	●	●	●	●	●	●	●	●	●	●	
Lithuania	LTU	440	●	●	●	●	●	●	●	●	●	●	●	●	
Macedonia	MKD	807							●	●	●	●	●	●	●
Malaysia	MYS	458							●	●	●	●	●	●	●
Moldova	MDA	498							●	●	●	●	●	●	●
Morocco	MAR	504							●	●	●	●	●	●	●
Netherlands	NLD	528	●	●	●	●	●	●	●	●	●	●	●	●	●
New Zealand	NZL	554	●	●	●	●	●	●	●	●	●	●	●	●	●
Norway	NOR	578	●	●	●	●	●	●							
Philippines	PHL	608							●	●	●	●	●	●	●
Portugal	PRT	620	●	●	●	●	●	●							
Romania	ROM	642	●	●	●	●	●	●	●	●	●	●	●	●	
Russian Fed.	RUS	643	●	●	●	●	●	●	●	●	●	●	●	●	●
Scotland	SCO	927	●	●	●	●	●	●							
Singapore	SGP	702	●	●	●	●	●	●	●	●	●	●	●	●	●
Slovak Republic	SVK	201	●	●	●	●	●	●	●	●	●	●	●	●	●
Slovenia	SVN	705	●	●	●	●	●	●	●	●	●	●	●	●	
South Africa	ZAF	710	●	●					●	●					
Spain	ESP	724	●	●	●	●	●	●							
Sweden	SWE	752	●	●	●	●	●	●							
Switzerland	CHE	756	●	●	●	●	●	●							
Thailand	THA	764	●	●	●	●	●	●	●	●	●	●	●	●	●
Tunisia	TUN	788							●	●	●	●	●	●	●
Turkey	TUR	792							●	●	●	●	●	●	●
United States	USA	840	●	●	●	●	●	●	●	●	●	●	●	●	●

"•" Indicates that file is available.

All TIMSS international data files are provided in ASCII format, enabling you to access the data directly without having to go through intermediate steps. All details of the file structure are provided in codebook files related to each of the data files. The use of these codebooks is described later in this chapter.

Background Variable Naming Convention

International background variables obtained from the student, teacher, and school questionnaires are provided in the corresponding background data files. In general, the background variables are provided for all countries where the data are considered internationally comparable. In a few cases, some slightly modified specific country options were retained in the international variables. The codebook files contain the international background variable names, descriptive labels, response code definitions, formats, and field locations corresponding to each questionnaire item.

In addition to the background variables contained in the questionnaires, a number of derived variables were computed for use in the international reports. These derived variables, many of which use data from more than one source, are also included in the International Database for use in secondary analyses. All derived variables are described in Supplement 3. The derived variables based on background variables can, in general, be used to reproduce the values shown in the international report tables by applying the appropriate school, teacher or student filters and weights.[3] Supplement 3 lists all derived variables used to produce the international reports, as well as analysis notes for those derived variables not included in the database. Descriptions of each of the derived variables, their associated exhibit reference number in the international report, and analysis notes indicating how they were computed using source variables are provided in Supplement 3. Documentation describing how specific national deviations were handled in the derived variables is also summarized in Supplement 3.

The naming convention for the background variables permits the identification of the survey population and data source based on 7- or 8-character codes. These are the variables that capture the responses given to the survey instruments (background questionnaires). The following convention is followed in naming these variables:

- The first character indicates the population. “B” is used to indicate eighth grade in 1999 and seventh and eighth grade (Population 2) in 1995.

- The second character indicates the type of respondent. The following abbreviations are used:

▾ ▾ ▾

[3] Additional information about the methods used to report background data may be found in Martin, Gregory, and Stemler, 2000.

C School Principal

T Teacher

S Student

- The third character indicates the type of question. The following abbreviations are used:

B Background Questionnaire Question

D Derived Variable

- The fourth character indicates the subject or topic to which a questionnaire item refers. The following abbreviations are used:[4]

G General Questionnaire Item (not subject specific)

M Questionnaire item related to Mathematics

S Questionnaire item related to General Science

B Questionnaire item related to Biology

E Questionnaire Item related to Earth Science

P Questionnaire item related to Physics

C Questionnaire item related to Chemistry

Y Questionnaire item related to Physical Sciences[5]

- For Background Questionnaire variables, the fifth, sixth, seventh and eighth characters indicate a question name and are unique to each variable.

Background Variable Location Convention

To help identify the location of the background variables in the corresponding background questionnaire, a convention has been developed that identifies the questionnaire and location within the questionnaire for each variable. This convention is used in the data almanacs and in the description of the variables included in the Supplements of this User Guide. Following this questionnaire convention, each questionnaire has been assigned a unique identification code as shown in Exhibit 8.3. This unique code is followed by the sequence number of the question within the questionnaire. For example, if the location of a variable is indicated as SQ2-06, this refers to the 6th question in the student background questionnaire, General Science Version.

▼ ▼ ▼

[4] Please note that "B," "P," "E," and "C" are only used in the student background data file for variables corresponding to questions about separate sciences asked in the separate science version of the student questionnaire.

[5] These variables were included only in the 1995 assessment.

Exhibit 8.3 Background Questionnaire Variable Location Conventions

Questionnaire	Location Code
Student Questionnaire General Science Version	SQ2-***
Student Questionnaire Separate Science Subject Version	SQ2S-***
Mathematics Teacher Questionnaire (Part A and B)	TQM2A-*** and TQM2B-***
Science Teacher Questionnaire (Part A and B)	TQS2A-*** and TQS2B-***
School Questionnaire	SCQ2-***

*** = the sequential numbering of the question location in the questionnaire.

Achievement Item Naming Convention

The achievement item variable names are based on 7- or 8-character alphanumeric codes according to the general definitions given below. For the achievement test items, the variable names reflect permanent identification numbers that indicate the test subject, the TIMSS assessment year in which the item was first introduced, the TIMSS population for which it was originally intended, a unique sequential item number, and item part when applicable.[6]

- The first character indicates the subject of the item. The following codes are used:

  M Mathematics

  S Science

- The second and third character indicates the assessment year when the item was first used in TIMSS. The following codes are used:

  01 First used in TIMSS 1995

  02 First used in TIMSS 1999

- The fourth character indicates the population for which the item was developed originally. The following codes are used:

  1 3rd and 4th graders (Population 1)

  2 7th and 8th graders (Population 2)

- The fifth, sixth and seventh characters are a sequential item number ranging from 001 and possibly up to 999.

▼ ▼ ▼

6 The cognitive variable names have been revised since the original database release in 1996. The correspondence between the current permanent Ids and the original items can be found in the Item Information Files described in Section 8.7.

- The eighth character indicates the item part. This was used only for free-response items. In general a free-response items asked one question, but in a few cases two or three questions were asked. This character was used when more than one question was asked as part of an item. In some cases for analysis purposes two of these questions were collapsed into one variable. The individual parts and the collapsed variables are all included in the database. The following codes were used:

A	First question in the item
B	Second question in the item
C	Third question in the item
D	Derived variables based on combined scores of either parts A and B or parts B and C when appropriate

As an example, the variable M012001A is a Population 2 Mathematics item number 001, Part A, first administered in TIMSS 1995; the item S022249D is a Population 2 Science item number 249, derived from combining two of the responses to item S022249, and was first administered in TIMSS 1999.

Coding Convention

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

Background Item Response Code Values

The values assigned to each of the background item variables depend on the item format and the number of options available. For the multiple-choice items, 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 items such as "Number of students in a class" are coded with the actual number given as a response to the question.

Achievement Item Response Code Values

The values assigned to each of the achievement item variables also depend on the item format. For the multiple-choice items, one-digit numerical values of 1-5 are used to correspond to the response options A through E. For these items, the correct response is included as part of the item-variable label in the codebook files and program code is included as part of the database to score these items.

For the free-response achievement test items, two-digit numerical codes are used that correspond to the diagnostic scoring rubrics used to determine fully-correct, partially-correct, and incorrect responses for each item. As described in Chapter 2, the correctness score level may be determined by the first digit of these codes (codes 30 through 39 are worth 3 points; 20 through 29 are worth 2 points; 10 through 19 are worth 1 point; and 70 through 79 or 90 through 99 are worth 0 points).

For some free-response items, students are asked to provide an answer with supporting work, or to provide two reasons, examples, consequences, etc. The two parts of their answer (parts A and B) are scored separately. In addition to the score given for each part of their answer contained in the part A and part B variables, the total score for the item as a whole is derived. The total score for the item is contained in an associated derived variable, indicated by a final character of D. For example, derived variable S012135D contains the combined score for item parts S012135A and S012135B. For the majority of these items, each item part is worth one point, and the derived variables have the following code values:

Code 20	Full credit (1 point on both parts A and B);
Code 10	Partial credit (1 point on either part A or B);
Code 70	No credit (0 points on both parts A and B).

It is also possible for one item part to be worth a maximum of two points. In this case, the derived variable will have code values of: 30, 20, 10, or 70.

In addition to the correctness score information, specific missing codes are also defined that are described in the section discussing missing codes. Since all achievement item variables are included for all students in the achievement files regardless of which test booklet they completed, a 'Not Administered' code is given to all items that were not included in the test booklet assigned to each student.[7] However, only the variables for the items administered in a particular year are included in the corresponding files.

Missing Code Values

All values assigned to variables in the TIMSS international data files are numeric, and a subset of the numeric values for each of the variable types is reserved for specific codes related to different categories of missing data.[8] The missing categories defined below are assigned different values depending on the field width of the variable and the variable type. We recommend that you read the following section with particular care since the way in which you make use of these missing codes will have serious consequences for your analysis.

▼ ▼ ▼

7 See Chapter 2 for a presentation of the TIMSS Test Design.

8 The SAS and SPSS control statement files will recode these missing categories to special numeric missing codes in SAS and explicit missing codes in SPSS.

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

Omitted response codes are used for questions/items that a student, teacher, or school principal should have answered but did not answer. These are coded as "omitted" in the codebooks. 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 way were also recoded to "omitted." For achievement items, an omitted response code was given only in cases in which the item was left blank; a special code was used for invalid answers as described below. The length of the omitted response code given to a variable in the ASCII file depends on the number of characters needed to represent the variable. In all cases the space necessary to represent the variable was filled with 9's.

Uninterpretable Response Codes (ASCII: 7, 79; SAS: .N; SPSS: 7, 79)[9]

For the achievement test items, separate codes were established to distinguish between totally blank responses (omitted) and uninterpretable or invalid responses. For multiple-choice items, cases where more than one response option was checked were classified as uninterpretable and given a code 7. For the free-response items in the achievement test, uninterpretable student responses were given a code 97, which is distinguished from the code 99 given to items that were left blank.

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

Special codes were given for 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 depends on the length of the field for the variable. In general, the not administered code was used when an item was not administered as part of the questionnaire or test instruments either by design, such as in the case of some of the test items, or unintentionally, such as when the item was left out of the instrument or misprinted. The Not Administered Codes are used in the following cases:

- Test item not assigned to the student - only one of the eight rotated booklets used in the TIMSS study was assigned to each student. All variables corresponding to items that were not given to a student have been coded to "Not administered".

- Booklet not received / booklet lost - If a respondent did not receive the instruments assigned to him/her, or the instruments were lost after administration, all items have been coded to "Not administered."

▼ ▼ ▼

[9] In TIMSS 1995 the code for uninterpretable open–ended responses is 90.

- Student absent from session - If a student or individual was not present for a particular testing session, then all variables referring to that session have been coded to "Not administered". However, if a student participated in a session and did not answer any of the items, these questions have been coded to "Omit."

- Item left out or misprinted - If a particular question or item (or a whole page) was misprinted or not available to the student, teacher, or school, the corresponding variables have been coded to "Not administered."

- Achievement items omitted or mistranslated in student test booklets - Any items identified during the translation verification or item analysis processes that were mistranslated such that the nature of the question was altered were removed for a country.

- Background questionnaire items were omitted - Questions in the student, teacher, or school background questionnaires that were considered not applicable in some countries were not included in their questionnaires.

- Background questionnaire items were mistranslated or not internationally comparable - In some cases, questions in the international version of the questionnaires were mistranslated or modified to fit the national situation. Whenever possible, modified background questionnaire items were recoded to match as closely as possible the international version of the items. This could not be done in all cases, however, and some national data were recoded to "not administered" in order to include only the internationally comparable data.

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

The Not Applicable response codes are used only for the background questionnaire items in which responses are dependent on a filter question. The specific Not Applicable code given depends on the number of valid codes available for the item, as described above for the Omitted Response codes.

Not Reached Item Codes (ASCII: 6, 96 ; SAS: .R ; SPSS: 6, 96)

The Not Reached item codes are used only for achievement items. Test items at the end of each test booklet in each testing session which were left blank were considered "not reached" due to the fact that the student did not complete the test. These responses are distinguished from the "missing" responses, as they are handled differently during the item calibration process (see Chapter 5). They are treated as incorrect responses, however, in computing achievement scores. For the multiple-choice items, a Not Reached item code value of 6 is used. For the free-response written, a Not Reached item code value of 96 is used.

Types of Variables Included in the Data Files

Identification Variables

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

The Identification Variables included in Student, Teacher, and School Files are the following:

IDCNTRY

Five-digit country identification code. This variable should always be used as one of the link variables whenever files are linked within and across countries. For each country the 1st two digits are always zeroes (00).

IDPOP

Identifies the population. The value is always set to 2 in the TIMSS 1999 International database.

IDSCHOOL

Identification number that uniquely identifies the school within each country. These codes for the school are not unique across countries. Schools across countries can be uniquely identified only by the IDCNTRY and IDSCHOOL combination.

Additional identification variables in the student files include the following:

IDSTUD

Identification number that uniquely identifies each student in the country sampled. The variable IDSTUD is a hierarchical identification number. It is formed by the combination of the variables IDSCHOOL and IDCLASS, followed by a two-digit sequential number within each classroom. Students can be uniquely identified in the database by the combination of IDSTUD and IDCNTRY.

IDCLASS

Identification number that uniquely identifies the sampled class within the school.

IDBOOK

Identifies the specific test booklet (1-8) that was administered to the student.

Additional identification variables in the teacher files include the following:

IDTEACH

Identification number that uniquely identifies the selected teacher within the school. It is a hierarchical identification number formed by the combination IDSCHOOL and a two-digit sequential number within each school. This variable is unique to each teacher within each country but is not unique in the teacher file.

IDLINK

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

IDSUBJCT

The subject(s) taught by the teacher (mathematics or science). This identification variable IDSUBJCT may be used to identify teachers as either mathematics or science teachers. The following codes are used:

Code 1	Mathematics Teacher
Code 2	Science Teacher
Code 3	Mathematics and Science Teacher

In the Student Background file, the IDSTUD variable provides a unique identification number to identify each student within each country. Since teachers may teach more than one class, the IDTEACH and IDLINK combinations in the Teacher Background files provide a unique identification for each teacher teaching a specific class. Teacher background variables are linked to appropriate students using the Student-Teacher Linkage file. The variable IDSCHOOL, contained in all three background files, is a unique identification number for each school within a country that may be used to link school background data to corresponding students or teachers.

Linking and Tracking Variables

Information about students, teachers, and schools provided on the survey tracking forms[10] is included in linking or tracking variables. These variables have prefixes of IL or IT. Some of the important linking and tracking variables are listed below.

Linking and tracking variables in the Student Background Files include the following:

ITSEX	Gender of each student as stated in the student tracking form.

▼ ▼ ▼

10 Survey tracking forms are listings of students, teachers, or schools used for sampling and administration purposes.

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.

ITLANG[11] Language of testing for each student. Set to 1 for all countries that tested in a single language. For countries that administered the test in more than one language, codes of 1, 2 and 3 were used to correspond to the order of the languages shown in Exhibit 1.4.

ITPART Participation status variable indicating whether each student participated in any TIMSS testing session; only those students with ITPART equal to 3 (participated in a TIMSS testing session) are included in the Student Background files.

ITCOURSE Tracking variable indicating the type of course taught by teachers: Mathematics = 1; Physics or Physical science = 2 or 7; Biology or Life science = 3 or 8; Chemistry = 4; Earth science = 5; General/Integrated science = 6; and other country-specific science courses = 9. This variable was also used to identify teachers for inclusion in separate science panels in the international reports' exhibits (1999 Only).

ILRELIAB Linking variable indicating the inclusion status of each student in the reliability file containing double-coded free-response items. The following codes are used:

Code 1 Student questionnaire used for reliability analysis

Code 0 Student questionnaire NOT used for reliability analysis

Selection Variables

These variables are used to select cases for particular analyses. In the student files and in the Student-Teacher link files the following variables are included:

INTMS99 Students with INTMS99 set to "1" are those who participated in the TIMSS 1999 achievement test and who are part of the TIMSS 1999 international sample. Otherwise this variable is set to zero.

INTMS95 Students with INTMS95 set to "1" are those who participated in the TIMSS 1999 achievement test and who are part of the TIMSS 1999 sample that is comparable to the TIMSS 1995 sample in the

▼ ▼ ▼

[11] See Exhibit 1.4 for a list of the languages included in each country.

country or state. Otherwise this variable is set to zero. Although in most countries students who have INTMS99 set to "1" are the same as those who have INTMS95 set to "1", this is not always the case. For countries that did not participate in TIMSS 1995, this variable is coded to zero for all students.

Sampling and Weighting Variables

Several sampling and weighting variables are included in the student data files. These variables are described in Chapter 5. The variables are the following:

WGTFAC1	School Weighting Factor. This variable is included in the student and in the school background data files.
WGTADJ1	School Weighting Adjustment. This variable is included in the student and in the school background data files.
WGTFAC2	Class Weighting Factor. This variable is included in the student data files.
WGTFAC3	Student Weighting Factor. This variable is included in the student data files.
WGTADJ3	Student Weighting Adjustment. This variable is included in the student data files.
TOTWGT	Total Student Weight. This variable is included in the student data files.
SENWGT	Student Senate Weight. This variable is included in the student data files.
HOUWGT	Student House Weight. This variable is included in the student data files.
MATWGT	Mathematics Teacher Weight. This variable is included in the student-teacher link data files.
SCIWGT	Science Teacher Weight. This variable is included in the student-teacher link data files.
TCHWGT	Overall Teacher Weight. This variable is included in the student-teacher link data files.
SCHWGT	School-level Weight. This variable is included in the school background data file.
STOTWGTL	Sum of the Student Weights for the Lower Grade. This variable is included in the school background data file (1995 only).
STOTWGTU	Sum of the Student Weights for the Upper Grade. This variable is included in the school background data file (1995 only).

JKZONE	The sampling zone or stratum to which the student's school is assigned. This variable is included in the student and student teacher link data files.
JKREP	The PSU to which the student is assigned. This variable is included in the student and student teacher link data files.
JKCZONE	The sampling zone or stratum to which the school is assigned. This variable is included in the school background data files.
JKCREP	The PSU to which the school is assigned. This variable is included in the school background data files.

Achievement Variables

Several achievement variables are also included in the student data files. These variables are described in Chapter 6. The achievement variables included are the following:

BSMMAT01-5	Mathematics Overall Plausible Values 1-5
BSMDAP01-5	Data Representation, Analysis and Probability Plausible Values 1-5
BSMALG01-5	Algebra Plausible Values 1-5
BSMFNS01-5	Fractions and Number Sense Plausible Values 1-5
BSMGEO01-5	Geometry Plausible Values 1-5
BSMMEA01-5	Measurement Plausible Values 1-5
BSSSCI01-5	Science Overall Plausible Values 1-5
BSSEAS01-5	Earth Science Plausible Values 1-5
BSSNOS01-5	Scientific Inquiry and the Nature of Science Plausible Values 1-5
BSSLIS01-5	Life Science Plausible Values 1-5
BSSPHY01-5	Physics Plausible Values 1-5
BSSCHE01-5	Chemistry Plausible Values 1-5
BSSERI01-5	Environmental and Resource Issues Plausible Values 1-5
BSMSCPT	Number of raw score points obtained on the mathematics items
BSSSCPT	Number of raw score points obtained on the science items
BSMSTDR	Standardized mathematics raw score
BSSSTDR	Standardized science raw score
BSMNRSC	National Rasch Mathematics Score (ML)
BSSNRSC	National Rasch Science Score (ML)
BSMIBM01-5	International Mathematics Benchmark reached with plausible values

BSMIBM01-5 International Science Benchmark reached with plausible values

BSMNBM01-5 National Mathematics Benchmark reached with plausible values

BSMNBM01-5 National Science Benchmark reached with plausible values

Background Data Files

There are three different types of TIMSS background files – student, teacher, and school. These are described below.

Student Background File

Students who participated in TIMSS were administered a background questionnaire with questions related to home background and school experiences. The Student Background file contains students' responses to these questions. Two versions of the student questionnaire were administered in TIMSS. One version was for educational systems where science is taught as an integrated subject (general science version). The second version was for educational systems where the sciences (biology, earth science, physics, and chemistry) are taught separately (separate science version). The variable ITQUEST identifies the version of the questionnaire that was administered to the students in TIMSS 1995. Chapter 3 indicates which countries administered which student questionnaire.

For students who were administered the general science version, all questions that were given only in the separate science version were coded as not administered. For students who were assigned a separate science version of the questionnaire, all questions that were asked only in the general science version were coded as not administered. The Student Background files also contain a series of identification variables, link variables, sampling variables, achievement variables, and the derived variables that were used for the creation of the international reports.

Teacher Background File

The mathematics and science teachers of the students who were sampled for TIMSS were administered at least one 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 files contain one record for each of the classes taught either by a mathematics or a science teacher. If a teacher taught more than one class, they were expected to complete only one part A (general background questions) and a separate part B (class-specific questions) for each class they taught. In some cases, although the teacher was to respond to more than one questionnaire, responses to only one were obtained. In these cases, there were as many records entered in the teacher file as classes were taught by the teacher, and the background information in part A from the complete questionnaire was entered into these teacher records. There were two questionnaires administered in TIMSS –one for the mathematics teachers and one for the science teachers. The data from these questionnaires are found in separate files. Variable names for questions asked in both questionnaires are the same.

In the Teacher Background data files each teacher was assigned a unique identification number (IDTEACH) and a Teacher Link Number (IDLINK) that is specific to the class taught by the teacher and to which the information in the data record corresponds. The IDTEACH and IDLINK combination uniquely identifies a teacher teaching one specific class. So, for example, students linked to teachers identified by the same IDTEACH but different IDLINK are taught by the same teacher but in different classes. The Teacher Background files cannot be merged directly with the student data files and they do not contain sampling information or achievement scores. 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 a representative sample of students. The teacher data should therefore be analyzed only in conjunction with the Student-Teacher Linkage file. The Teacher Background data files contain a series of other identification variables, and link variables, as well as the responses of the teachers to the background questions.

School Background File

The principals or administrators of the schools in the TIMSS sample were administered a school background questionnaire with questions about school policy and school environment. The School Background data file contains the responses given to the questions in this questionnaire. That file also contains a series of identification variables, link variables, and sampling variables. The school data files can be merged with the student data files by using the country and school identification variables.

Student Achievement Data Files

Student achievement files contain the student response data for the individual achievement items in the TIMSS achievement test.

Students who participated in TIMSS were administered one of eight test booklets with questions in mathematics and science. Some of these questions were multiple-choice questions and some were open-ended. The responses to the open-ended questions were coded using a two-digit coding system. The achievement test data files contain the answers to the multiple-choice questions and the codes assigned by the coders to the student responses for open-ended questions. Since under the TIMSS test design a student received only a fraction of the total test item pool, the variables for the items that were not included in the test booklet that was administered to the student are coded as not administered. The specific test booklet that was administered to the student is coded in the variable IDBOOK. The student achievement data files also contain a series of identification variables, sampling variables, and achievement variables. The data contained in this file can be linked to the student background data files by using the variables IDCNTRY and IDSTUD.

Scoring Reliability Data Files

The scoring reliability files contain data that can be used to investigate the reliability of the TIMSS free-response item scoring. The scoring reliability file contains one record for each booklet that was double scored during the scoring reliability exercise of the free-response items.

For each free-response item in the achievement test, the following three variables are included:

- Original Code Variable (achievement item response codes obtained from the first coder)
- Second Code Variable (achievement item response codes obtained from second coder)
- Response Code Agreement Variable (degree of agreement between the two codes)

It should be noted that the Second Code Variable data were used only to evaluate the within-country coding reliability. Only the first codes contained in the student achievement files were used in computing the achievement scores reflected in the Student Background files and the international reports.

Scoring Reliability Variable Naming Convention

The variable names for the Original Code, Second Code, and Agreement Code variables are based on the same general naming system as that for the achievement item variables shown earlier. The second character in the variable name differentiates the three reliability variables:

S Original Code Variable (e.g. MS12001)

R Second Code Variable (e.g. MR12001)

I Agreement Code Variable (e.g. MI12001)

Reliability Variable Code Values

The values contained in both the Original Code and Second Code variables are the two-digit diagnostic codes obtained using the TIMSS scoring rubrics. The Agreement Code variable has three different values depending on the degree of agreement between the two coders:

Code 0	Identical codes (both digits in the diagnostic codes are identical)
Code 1	Identical score but different diagnostic code (first digits are the same; second digits are different)
Code 2	Different score (both the first and second digits are different)

In general, the response code data contained in the Original Code Variables are identical to those contained in the Student Achievement test files. In some cases, however, the response codes for specific items were recoded after a review of the international item statistics revealed inconsistencies in the original coding guides or showed that the original codes were not functioning as desired. The recoded diagnostic code values were used in computing the achievement scores reflected in the international reports.

The response codes in the Student Achievement test files reflect the recoded values. In contrast, the Original Code Variables in the coding reliability files contain the original unrecoded response codes. This was done so that the coding reliability measure indicated in the Agreement Code Variables was based on the original coding guides used during the free-response coding sessions conducted in each country. One exception to this is that any nationally defined diagnostic codes employed in individual countries (second digit of 7 or 8) were recoded to the "other" category (second digit of 9) within the same correctness level prior to the computation of the Code Agreement Variables.

In addition to the coding reliability variables, the reliability files also include identification variables to aid in case identification. Some tracking variables are also included that were used in conducting the coding reliability study within each country, including the reliability booklet set to which each student was assigned (ITBSET) and the identification of the first and second coders (IDSCORA and IDSCORR).

Student-Teacher Linkage Files

The Student-Teacher Linkage files for TIMSS contain information required to link the student and teacher files and to compute appropriately weighted teacher-level data using the student as the unit of analysis.

The Student-Teacher Linkage files contain one entry per student-teacher linkage combination in the data. In many cases, students are linked to more than one mathematics and/or science teacher, and in these cases there will be one record for each student-teacher link. In addition, in some countries many students may also have more than one teacher for each of the two subject areas. For instance, if three teachers are linked to a student, there are three entries in the file corresponding to that student.

National Data Issues Affecting the Use of International Data Files

In some cases, resources were not available to resolve database issues for specific countries in time for either the release of the international reports or the production of the international data files for TIMSS 1995. As a result, some international data are modified or not available for some countries. These general database issues as they pertain to the TIMSS 1995 and TIMSS 1999 data are documented below since they impact the ability to perform trend analyses in some cases.

TIMSS 1999 Data Issues

Chinese Taipei

In Chinese Taipei, biology is taught in grade 7, physics/chemistry in grade 8 and earth science in grade 9. Science teacher questionnaires were administered to the grade 7 biology and grade 8 physics/chemistry teachers of the sampled mathematics class. Data for the grade 7 teachers were retained in the international database but judged to be not comparable due to the different year. As a result, zero weights were assigned for the biology teachers in the international database and these data cannot be used to compute population estimates.

Slovak Republic

Grade 8 students are enrolled in four science courses (biology, physics, chemistry, and geography/geology), but only one science teacher corresponding to students in the sampled mathematics class was administered a questionnaire in most schools. Dummy records were created for the missing teachers based on the missing ITCOURSE values for each student. Appropriate weights and student-teacher links were assigned and all background variables were set to Not Administered for the missing science teachers. Due to the high level of missing data, science teacher data were not reported in the international report exhibits.

Slovenia

Grade 8 students are enrolled in three science courses (biology, physics, and chemistry), but only one science teacher corresponding to students in the sampled mathematics class was randomly selected to complete a questionnaire. Dummy records were created for the missing teachers based on the missing ITCOURSE values for each student. Appropriate weights and student-teacher links were assigned and all background variables were set to Not Administered for the missing science teachers. Due to the high level of missing data, science teacher data were not reported in the international report exhibits.

Belgium (Flemish)

Sampled an extra mathematics class in each school. Student background and mathematics teacher background data were obtained for the second class, but no science teacher data were collected. Dummy records were created for the missing science teachers for the second class. Zero weights were assigned and all background variables were set to Not Administered for the missing science teachers. Population estimates computed from science teacher data are based only on the students in the first sampled class, and missing rates are unaffected by the missing science teachers for the second class.

Lithuania

Lithuania tested the same cohort of students as other countries, but later in 1999, at the beginning of the next school year. The mathematics and science teacher questionnaires were, therefore, completed by the grade 9 teachers. Some teacher background data that pertain to the specific grade level (e.g., topic coverage) are not comparable and not included in the international database.

TIMSS 1995 Data Issues

Bulgaria

The student, teacher, and school background data submitted by Bulgaria were not deemed internationally comparable in 1995 and are thus not included in the International Database.

Philippines

The teacher and school data submitted by the Philippines were not deemed internationally comparable; however, they are included in the International Database.

Due to the use of unapproved school sampling procedures, the results presented in the international reports for the Philippines reflect unweighted data. Consequently, the sampling weights have been set to 1 for all cases in the files for the Philippines.

South Africa

The teacher and school data submitted by South Africa were not deemed internationally comparable and thus are not included in the International Database.

Switzerland

Sampling at Population 2 was achieved by track within school, and the school weights in the database reflect the size of the track rather than the total Population 2 enrollment in the school. The obtained total weights for students result in valid student-weighted estimates. The track-based school weights, however, are not appropriate for producing school-weighted estimates based on school-level variables. Although Switzerland was excluded from Population 2 tables in the TIMSS school report that are based on school-weighted analyses, all of their school background data are retained in the database.

Thailand

Information made available after publication of the Population 2 international reports required that the sampling weights for Thailand be recomputed. The adjusted sampling weights are included in the International Database. As a consequence, any computations using these new weights may be slightly different from those shown in the international report tables for Population 2.

8.3 Data Almanacs

Data Almanac Files contain weighted summary statistics for each participating country on each achievement item and on each variable in the student, teacher, and school background questionnaires and derived variables based on these questionnaires. Separate data almanacs are included for the background variables and for the achievement item variables. The data almanac files corresponding to each variable type are listed in Exhibit 8.4 and are described in the following sections.

The naming convention of the almanac files is as follows:

- The first character of the files is always “B.” This indicates that the file refers to eighth grade in 1999 and to seventh and eighth grades (Population 2) in 1995.

- The second character indicates the source or level of the information in the file:

  S indicates a background, student level almanac

  I indicates an achievement item almanac

- The third, fourth and fifth character identify the file as an almanac and it is always “ALM”.

- The sixth character identifies the information contained in the almanacs. The following codes are used for the sixth character:

  1 Student Background Questionnaire with Mathematics Achievement

  2 Student Background Questionnaire with Science Achievement

  3 School Background Questionnaire with Mathematics Achievement

  4 School Background Questionnaire with Science Achievement

  5 Mathematics Teacher Background Questionnaire with Mathematics Achievement

  7 Science Teacher Background Questionnaire with Science Achievement

  M Mathematics achievement items

  S Science achievement items

- The seventh character identifies the grade. This is used only for the 1995 data almanacs when two grades were tested at each of the populations. The following codes are used for the seventh character:

1	Lower Grade in the Population (Used only with the 1995 data files)
_	Upper Grade in the Population (Used with both the 1995 and 1999 data files)

- The last two characters in the file name indicates the study cycle. The following codes are used:

M1	stands for the 1995 almanac files
M2	stands for the 1999 almanac files

- The three character file extensions used for the codebook files indicate the file type. The following codes are used:

.DOC	Almanacs in MS Word format
.LST	Almanacs in ASCII format
.PDF	Almanacs in PDF format

Exhibit 8.4 Data Almanac Files for TIMSS 1999

Almanac File	Contents
BSALM1_M2	Student Background with Mathematics Achievement
BSALM2_M2	Student Background with Science Achievement
BSALM3_M2	School Background with Mathematics Achievement
BSALM4_M2	School Background with Science Achievement
BSALM5_M2	Mathematics Teacher Background with Mathematics Achievement
BSALM7_M2	Science Teacher Background with Science Achievement
BIALMM_M2	Percent of Responses by Item Category for the Mathematics items
BIALMS_M2	Percent of Responses by Item Category for the Science items

The data almanacs are provided as PDF files to be read with Adobe Acrobat Reader 4.0 or higher, as ASCII formatted files (LST), and as MS Word format files (DOC) to be read and edited with any text editor or word processing software that can read files of their size.

The files display student-weighted summary statistics, by grade, for each participating country on each variable. The almanacs also display the international averages for each variable, with each country weighted equally.

Background Item Data Almanacs

Background data almanacs include all student, teacher, and school background variables as well as derived variables based on these. The data presented in the almanacs use the student as the unit of analysis, even if the information is based on the teacher or school questionnaire items. Therefore, the weighted percentages in the teacher and school almanacs reflect the percentage of students to whom the data apply.

There are two types of displays in the background data almanacs, depending on whether data are categorical or continuous.

The display for categorical variables includes:

- Sample size (number of students, teachers or schools included in the sample)
- Valid number of cases (counts of students, teachers or schools for whom valid data were obtained)
- Weighted percentages of students corresponding to each valid response option (percentages based only on the students with valid data)
- Weighted percentages of students for whom none of the valid response options were selected (the variable is coded as "Not Administered" or "Omitted")

In cases where the question was Not Applicable based on the response to an explicit filter question, the percentage of students for whom the variable is coded as Not Applicable is also displayed. The percent Not Applicable is based on only the students with valid data. The last line in the almanac displays the international average in each category of response.

The display for continuous variables includes:

- Sample size (number of students, teachers or schools in the sample)
- Number of valid cases (counts of students, teachers or schools with valid data)
- Weighted percentages of students for whom the variable is coded as "Not Administered", "Not Applicable", or "Omitted"
- The mean, mode, minimum, maximum, and the 5^{th} , 10^{th} , 25^{th} , 50^{th} , 75^{th} , 90^{th} , and 95^{th} percentiles across students
- The last line in the almanac displays the international averages.

The background data almanacs also display for each variable the question that was asked, the location in the corresponding questionnaire and the variable name in the data file.

Achievement Item Data Almanacs

Achievement item data almanacs include summary information about the response types given by students. They contain, for each country, the percent of students choosing each option for multiple-choice items or the percent of students receiving each diagnostic code in the scoring rubric for the free-response items. They also contain the percent of boys and girls obtaining the maximum score on the item and display the international average percentages for each response category, with each country weighted equally.

There are two types of displays in the almanacs, depending on whether the item is a multiple-choice item or a free-response item. The statistics displayed in these almanacs are defined below:

- **N:** The number of students to whom the item was administered.
- **A, B, C, D and E:** Used for multiple-choice items only, these represent the percent of students choosing each one of the response options for the item. Not reached items are included in the denominator for these calculations.
- **Other Incorrect:** Used for multiple-choice items, it indicates the percent of students who omitted or gave an invalid response to the item.
- **Scoring Guide Codes (10, 11, 70, 71, etc.):** Used for free-response items only, these represent the percent of student responses assigned each of the codes in the scoring rubric for the test item. Not reached items are excluded from the denominator for these calculations.
- **Diff (Item Difficulty):** The percent of students that responded correctly to the item. This was used only for the multiple-choice items. When computing this statistic, not reached items were treated as administered and so the values might differ from those presented in the International Item Statistics Almanac.
- **V1, V2, V3:** Used only for the free-response items, these indicate the percent of students that scored 1 point or better on the item (V1), 2 points or better (V2) and 3 points or better (V3).
- **Invalid:** Used for multiple-choice items only, this indicates the percent of students that gave an invalid response to the item (e.g., multiple selections).
- **Not Reached:** Indicates the percent of students that did not reach the test item.
- **Omit:** Indicates the percent of students that omitted the test item.
- **Girl (% Right), Boy (% Right):** This indicates the percent of girls and boys that obtained the maximum score on the item.

8.4 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 are contained in codebook files. The codebook files corresponding to each variable are described in the following sections. One codebook file is provided for each of the data files listed earlier.

The naming convention of the codebook files is as follows:

- The first character of the files is always "B." This indicates that the file refers to eighth grade in 1999 and to the seventh and eighth grades (Population 2) in 1995.

- The second character indicates the source or level of the information in the file:

  C indicates a school level file

  T indicates a teacher level file

  S indicates a student level file

- The third character indicates the subject and/or type of the data in the file. The following abbreviations are used, listed in alphabetical order:

  A Student Achievement Booklets

  G General Background Questionnaires (*School and Student Questionnaires)*

  M Mathematics Teacher Background Questionnaire

  R Free-Response Scoring Reliability (scoring reliability sample of Student Test booklets)

  S Science Teacher Background Questionnaire

  T Student-Teacher Link File

- The next three characters identify the file as a codebook and it is always "CBK."

- The seventh and eighth characters indicate the study cycle

  M1 stands for the 1995 achievement files

  M2 stands for the 1999 achievement files

- The three-character file extensions used for the codebook files were the following:

  .CDT Codebook in ASCII text format

  .CDF Machine Readable ASCII format

  .SDB Dbase format

Accessing the Codebook Files

All three codebook file types are included in the database CD. The ASCII format codebooks can be read and edited with any text editor or word processing software that can read files of their size. Each is designed with a specific purpose in mind.

Printout Format (*.CDT)

The printout format is a text file containing the information from the codebook in a printout format. This format can be read with any word processing software and printed after some minor formatting. Using a mono-spaced font and a font size and page layout combination that will accommodate 132 characters per line is suggested. The information for each variable is presented in several lines of text. The lines for each variable are properly labeled.

Machine Readable Format (*.CDF)

A second formatted version of the codebooks is also included in the database. In this version each variable occupies one line and the following fields are included: variable name, question location, starting column, ending column, number of digits, number of decimals, variable label, and value labels. Those who want to use programming languages other than SAS or SPSS to conduct analysis with the TIMSS data can use these files. Semicolons separate the value labels in these files. Exhibit 8.5 below describes the structure of the machine-readable codebook files.

Exhibit 8.5 File Structure of Machine-Readable Codebook Files

Field Name	Location
Variable name	1-8
Question location	10-20
Starting location in data file	22-24
Ending location in data file	27-29
Number of digits	31-32
Number of decimals	34
Variable label	36-74
Value labels	77-200

Using the Codebooks

The variables in the codebooks appear in order by variable name within the section for each codebook type. The major sections of each codebook type are as follows:

Student, Teacher, and School Background File Codebooks

- Identification Variables
- Tracking/Linking Variables
- International Background Variables (in order of questionnaire item location)
- Derived Variables
- Sampling Variables (Student and School Files only)
- Score Variables (Student Files only)

Student Achievement Test File Codebooks

- Identification Variables
- Tracking/Linking Variables
- Achievement Item Variables (in order by item within clusters)
- Sampling Variables
- Score Variables

Student-Teacher Linkage File Codebooks

- Identification Variables
- Sampling Variables
- Score Variables
- Teacher Linking/Weighting Variables

Reliability File Codebooks

- Identification Variables
- Tracking Variables
- Reliability Variables (organized into sets of three variables described previously in order by item within cluster)

The fields are as follows:

Variable Number: The first column (Var. No.) contains a sequential number for each variable in each codebook file.

Question: The second column contains an abbreviated variable identifier providing descriptive information needed to identify the content of the question and/or the source for each type of variable.

Variable Name: The third column (Variable Name) contains the variable name associated with each variable included in the international data files. The naming system used for each variable type is described in the previous sections on the contents of data files.

Variable Label: The fourth column (Label) contains an extended textual variable label of up to 40 characters associated with each variable, providing more descriptive information about the content of each variable. For multiple-choice achievement items, the variable label includes the correct response option enclosed in brackets. During data analysis, the variable labels can be printed out to enhance understanding of the results.

Code: The fifth column (Code) contains the codes used for variable responses. For variables where numeric data are supplied in response to open-ended questions, the keyword VALUE is entered in the Code column. For categorical variables, all possible response options are listed. Any missing codes described in Section 7.2.5 are also included for either numerical or categorical variables. For example, for multiple-choice achievement items, the code options are A, B, C, D, and E while for the free-response achievement items, the code options are the two-digit numerical codes described in Chapter 4.

Option: The sixth column (Option) includes a textual description of each type of response option. For variables containing numeric data, it contains an explanation of the values contained in the variable.

Location/Format: The seventh column presents the location and format of each variable in the raw data files. The location/format indicates the pattern used to write each value of a numeric or categorical variable, with a general structure of XX-YY / <N or C> ZZ.Z.

The numbers preceding the forward slash (/) indicate the location of the variable and refer to its position in the raw data file (starting (XX) - ending (YY) column positions). The <N or C> after the slash identifies the variable as numerical (N) or categorical (C). The numeric codes after the forward slash (ZZ.Z) indicates the total number of digits (including the decimal point) and the number of decimal places associated with each variable (e.g. 2.0 = 2 integer digits, 0 decimal places; 6.2 = six total digits: 3 integer digits, decimal point, and two decimal digits).

8.5 Program Files

Three different types of program files are provided for use in analyses of the TIMSS data files:

- Data Access Control Files
- Jackknife Statistics Program Files
- Scoring Program Files

The Data Access Control files are provided to convert the ASCII-format raw data files into SAS data sets or SPSS system files. The Jackknife statistics program files are used to compute the statistics as well as the standard errors associated with these statistics, using the jackknife repeated replication method discussed in Chapter 7.

The Scoring Program files are required to convert achievement item response codes to the score values used in the computation of international scores. For all program files, two versions are provided: one for SAS programs and one for SPSS programs. The file extension (SAS or SPS) is used to identify the respective SAS and SPSS program files.

Each data file has its own control file. Control files are named according to following convention:

- The first character of the files is always "B." This indicates that the file refers to eighth grade in 1999 and to the seventh and eighth grades (Population 2) in 1995.
- The second character indicates the source or level of the information in the file:

  C indicates a school level file

  T indicates a teacher level file

  S indicates a student level file
- The third character indicates the subject and/or type of the data in the file. The following abbreviations are used, listed in alphabetical order:

  A Student Achievement Booklets

  G General Background Questionnaires (*School and Student Questionnaires)*

  M Mathematics Teacher Background Questionnaire

  R Free-Response Scoring Reliability (scoring reliability sample of Student Test booklets)

  S Science Teacher Background Questionnaire

  T Student-Teacher Link File

- The next three digits identify the file as a data access control file and is always "CTR."
- The seventh and eighth character indicates the study cycle:

  M1 stands for the 1995 achievement files

  M2 stands for the 1999 achievement files

File Extension Definitions

The three character file extensions used for the data access control files are listed below in alphabetic order:

.SAS SAS Data Access control file

.SPS SPSS Data Access control file

Chapters 9 and 10 further describe the SAS and SPSS program files and how they are applied through the use of specific example analyses using the TIMSS student, teacher, and school data files.

8.6 Test-Curriculum Matching Analysis Data Files

TIMSS 1999 developed international tests of mathematics and science that reflect, as far as possible, the various curricula of the participating countries. The subject matter coverage of these tests was reviewed by the TIMSS 1999 Subject Matter Item Replacement Committee, which consisted of mathematics and science educators and practitioners from around the world, and the tests were approved for use by the National Research Coordinators (NRCs) of the participating countries. Although every effort was made in TIMSS 1999 to ensure the widest possible subject matter coverage, no test can measure all that is taught or learned in every participating country. The question therefore arises how well the items on the tests match the curricula of the participating countries. To address this issue, TIMSS 1999 asked each country to indicate which items on the tests, if any, were inappropriate to its curriculum. For each country, TIMSS 1999 then took the list of remaining items and computed the average percentage correct on those items for that country and all other countries. This allowed each country to select only those items on the tests that they would like included, and to compare the performance of their students on those items with that of the students in the other participating countries. However, in addition to comparing the performance of all countries on the set of items chosen by each country, the Test-Curriculum Matching Analysis (TCMA) also shows each country's performance on the items chosen by each of the other countries. In these analyses, each country was able to see not only the performance of all countries on the items appropriate for its curriculum, but also the performance of its students on items judged appropriate for the curriculum in other countries. The analytical method of the TCMA is described in Beaton and Gonzalez (1997).

There are two files that contain data indicating the item selection status given by each country, at each grade level. These files are located in the subdirectory called TCMA in the CD. The two Test Curriculum Matching Analysis files are:

BTCMAMM2.CSV	1999 Mathematics item selection
BTCMASM2.CSV	1999 Science item selection

These files are in text format, with their fields separated by commas. The first two records for each file contain the labels for each field in the file. The revised item variable names (permanent IDs) are included in the first record. The corresponding item cluster locations are included in the second record. Each row in the file contains a country's selection status for all items in the appropriate subject area:

Code 1	Item Included
Code 0	Item Excluded

8.7 Item Information Files

Item Information files are provided to enable users of the database to readily produce summaries of item characteristics and cross-reference the different item identification numbers (e.g., item cluster location, original item variable name (1995), and new item variable name (1999 permanent ID). The Item Information files include the following information:

- Permanent ID number for the item
- Item Name and Cluster Location in the corresponding year
- Test Subject (Mathematics or Science)
- Item Type (Multiple choice or Free-Response)
- Response Key (correct response option for multiple-choice items only)
- Maximum Points (maximum score points possible on the item)
- Whether the item was used for scaling or not
- Release status for the year
- Number of options
- Content Area Subscale or Reporting Category

- Item Label
- Performance Expectation Category

There are two Item Information files containing this information for the test items in both years. These files are located in the subdirectory called ITEMS for the corresponding population. The two Item Information files are:

BITINFM1.TXT	Item Information for the 1995 Population 2 Assessment
BITINFM2.TXT	Item Information for the 1999 Population 2 Assessment

These files are in text format, with their fields separated by tabs, and there is one record for each achievement item part. They may, therefore, be opened and manipulated directly with a spreadsheet program.

9 Performing Analyses with the TIMSS Data Using SPSS

9.1 Overview

This chapter presents some basic examples of analyses that can be performed with the TIMSS International Database using the sampling weights and scores discussed in previous chapters. It also provides details on some SPSS programs to conduct such analyses, and the results of these analyses. The analyses presented here are simple in nature, and are designed primarily to familiarize you with the different files and their structure, as well as the relevant variables that need to be included in most analyses. The programs compute the percent of students in specified subgroups, the mean achievement on science in those groups, and the corresponding standard errors for the percent and mean statistics. Additionally, some examples of regression statistics are presented. These analyses, based on student, teacher and school data, replicate analyses that are included in the TIMSS 1999 Science International Report (Martin et. al, 2000). You are invited to compare the results from these analyses to the exhibits in the reports, and are encouraged to practice analyzing the TIMSS data by trying to replicate the exhibits that are presented in the international reports.[1]

In our examples we use macros written for SPSS that can be used to perform any of the analyses that are described below. These are general procedures that can be used for many purposes, provided you have some basic knowledge of the SPSS macro language. If you have some programming experience in this statistical package, then you will be able to make the necessary modifications to the macros to obtain the desired results. Before using these macros, you should create a system file in SPSS that contains the variables necessary for the analysis. As part of this chapter we describe the control files included in the CD that can be used to do this.

9.2 Contents of the CDs

There are two CDs that accompany this User Guide – one CD containing the TIMSS 1999 data and one containing the TIMSS 1995 data. Each CD has the following internal file structure:

- A main directory identifying the year (TIMSS95 or TIMSS99).
- Within each main directory, there are six sub-directories.

DATA:	Contains data files in ASCII format
PROGRAMS:	Contains SPSS and SAS programs
CODEBOOK:	Contains codebook files
ALMANACS:	Contains data almanacs

▼ ▼ ▼

1 Documentation regarding the computational methods used to obtain any derived variables included in the international reports is presented in Supplement 3.

TCMA:	Contains Test–Curriculum Matching Analysis Data
ITEMS:	Contains the Item Information files

The directory names within each CD and the file names generally follow the DOS 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 in the files contained in the CDs are the following.

.SAS	SAS Control file or program
.SPS	SPSS Control file or program
.DAT	ASCII Data file
.LST	Almanac
.CDT	Codebook in Printout format
.CDF	Codebook in Machine readable format
.CSV	Test Curriculum Matching Analysis
.TXT	Item Information Tables

The DATA sub-directory contains the TIMSS data files in ASCII format. The data files that are in this directory are described in Chapter 8 of this guide. Each of these files has two corresponding control files in the PROGRAMS sub-directory. One of these reads the ASCII data and creates a SAS data set, the other reads the ASCII data and creates an SPSS system file. This chapter will focus on the files that can be used with SPSS.

The following programs also can be found in this directory:

BSASCRM1.SPS and BSASCRM2.SPS

These files contain SPSS programs that can be used to convert the response codes to the cognitive items to their corresponding correctness score levels. The use of these programs is described in this chapter.

JACKPV.SPS

This macro program in SPSS can be used to compute weighted percentages of students within defined groups, and their mean achievement scores on an achievement scale using the available plausible values. This macro makes use of the plausible values provided for each subject in computing the mean achievement scores. This macro also generates replicate weights and computes the JRR sampling variances for the percentages of students within the group, and the JRR and imputation variances for the mean achievement scores. This macro should only be used when multiple plausible values are used in the analysis.

JACKGEN.SPS

This macro program in SPSS can be used to compute weighted percentages of students within defined groups, and their means on a specified continuous variable. This macro also generates replicate weights and computes jackknife repeated replication (JRR) sampling variances for the percentages and mean estimates. The variable can be any continuous variable in the file. How to use each of these macro programs is described later in this chapter. When computing the average of the achievement scores or when computing plausible values, you will need to use the macro JACKPV.SPS.

JACKREG.SPS

This macro program in SPSS can be used to compute weighted regression coefficients and their corresponding standard errors within defined groups. It also computes descriptive statistics on the variables. This macro can be used with any variable in the analysis but it does not make use of the five plausible values.

JACKREGP.SPS

This macro program in SPSS can be used to compute weighted regression coefficients and their corresponding standard errors when using plausible values as the predicted scores within defined groups. It also computes descriptive statistics on the variables.

Each of the four macros above has a corresponding sample program that calls these and prints out the results. These are discussed later in the chapter.

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

These are the programs used in the examples presented later in this chapter. These programs are included only in the CD with the 1999 data, although the same examples can be easily adapted to perform the same analyses with the 1995 data.

9.3 Creating SPSS System Files

The CD contains SPSS control code to read each one of the ASCII data files and create an SPSS system file. Each of these control files contain information on the location of each variable in the file, its format, a descriptive label for each variable and their categories (in the case of categorical variables), and code for handling missing data. The control and data files have been created to facilitate access of the data on a country by country basis. The command lines in the control files should be edited to produce programs that will create SPSS system files for any specified countries. While most of the program code is functional as provided, you will need to edit input and output directories. Performing analyses that require the data from more than one country will necessitate merging or appending the respective data files into a larger one. Alternatively, you can access the data and compute the necessary statistics on a country by country basis by reading one file at a time, computing the necessary statistics, and then moving on to the next country's data. The method chosen by you will depend greatly on the storage and processing capacity of the computer system that is used. For the examples that we present in this User Guide we have combined the data files of individual countries into one larger data file that contains the data for all participating countries. The 3-character identifier for this file is "ALL".

When creating a system file using SPSS, you will need to do the following:

1. Open the corresponding control file, for example, BSGCTRM2.SPS.

2. In the line where it reads '!let !datpath = !unquote("<path>")' you will need to enter the path where the raw data is located.

3. In the line where it reads '!let !savpath = !unquote("<path>")' you will need to enter the path where you want to store the system file that you will be creating.

4. At the bottom of the files in the line where it reads "create country =" select the three character identification code for each country whose data you want to create. If you leave this line unedited the data for all countries listed will be created.

5. Submit the code for processing. After processing is complete you will find the corresponding system file in the location you specified in step 3 above.

For example, in the extract presented in exhibit 9.1 the student background data for Australia, Belgium Flemish, Bulgaria, Canada, England and the United States will be read from the directory "/usr/isc/timss/idb/data/raw/." Separate SPSS system files for each country will be stored in the directory "/usr/isc/timss/idb/data" under the name BSG<3-character country code>M2. Only the sections that need to be modified are presented in this exhibit.

Exhibit 9.1 Extract from SPSS Control Code for Creating a Student Background SPSS System File

```
SET DECIMAL = DOT.

define create (country=!charend('/')).

!let !datpath = !unquote("/usr/isc/timss/idb/data/raw/").
!let !savpath = !unquote("/usr/isc/timss/idb/data/").

!do !cty !in(!country).

!let !indata  = !concat('"',!datpath,btm,!cty,'m2.dat"').
!let !savdata = !concat('"',!savpath,btm,!cty,'m2.sav"').

* NOTE: Assignment of variable names.
* NOTE: Please change <path> and <filename> according to your system.
*       Change this in two of the following lines of this program.


*NOTE: Assignment of variable names.

data list file = !indata

.
.
.
.
.
.

save outfile = !savdata.

!doend
!enddefine

create country = AUS BFL BGR ENG CAN USA.
```

9.4 Scoring the Items

There were several types of items administered as part of the TIMSS tests. There were multiple-choice items, in which the student was asked to select one of four or five options as the correct response. These were administered as part of the achievement test. The responses to these items are coded with one digit. The codes used to represent the responses to these items are as follows:

Code 1	Option A
Code 2	Option B
Code 3	Option C
Code 4	Option D
Code 5	Option E
Code 6	Not reached
Code 7	Invalid response (chose more than one of the options available)
Code 8	Not administered
Code 9	No response although the item was administered and was reached (i.e., item was omitted)

There were also free-response items where the students were asked to construct a response to a question, rather than choosing an answer from a list of options. Scorers trained to use the two-digit scoring rubrics described in Chapter 2 of this guide scored the answers to these questions. The first digit of the two-digit code indicates the score given to the question, and the second digit in conjunction with the first provides diagnostic information on the specific answer given by the student. These types of response codes were used for the free-response items administered as part of the achievement test. The codes used to represent the responses to these items are the following:

Code 30 to 39	Three-point answer. Second digit provides diagnostic information.
Code 20 to 29	Two-point answer. Second digit provides diagnostic information.
Code 10 to 19	One-point answer. Second digit provides diagnostic information.
Code 70 to 79	Zero-point answer. Second digit provides diagnostic information.
Code 90	Uninterpretable.[2]
Code 96	Not reached.
Code 98	Not administered.

The achievement data files contained in the CD include information about the answer given to each item administered to a student. You might want to work with these item data after they are recoded to the right-wrong format, in the case of multiple-choice items, or to the level of correctness in the case of the free-response items. To this effect, we have included in the CD a set of programs in SPSS that will allow you to recode the items from the achievement test to their right-wrong or correctness-level format. These programs contain a macro called SCOREIT and the necessary call to this macro so that all the items in the corresponding file are scored. This program will convert the response option codes for multiple-choice items to dichotomous score levels (0 or 1) based on scoring keys. For the free-response items the two-digit codes will be converted to the corresponding correctness score level (3, 2, 1, 0) based on the value of the first digit, as described in Chapter 2.

Two files are included to provide control code to perform the recodes of the test items in the achievement test file:

BSASCRM1	Written Assessment Files in 1995
BSASCRM2	Written Assessment Files in 1999

▼ ▼ ▼

[2] This code is used only in the 1995 achievement files.

When using these programs, you must first consider the recoding scheme that is desired. For example, under certain circumstances you might want to recode the not reached responses as incorrect (codes 6 and 96), whereas under other circumstances you might want to recode these responses as not administered or invalid. In the case of TIMSS, not reached responses were recoded as not administered (and effectively as missing responses) for the purpose of calibrating the items. But the not reached responses were then recoded as incorrect when scoring the item for the individual countries, and for the purpose of calculating the scale scores for the individuals. By default, the scoring program provided with the database recodes the items coded as not reached and those left blank as incorrect responses.

To use the SCOREIT macro you need to include it as part of the SPSS programs used for the analysis. This is done by using the INCLUDE statement in the corresponding program. When using SPSS, the scoring program code should be included after the system file containing the item responses has been read into memory and becomes the working file. Both of these programs recode the items onto themselves, so if you want to preserve the original answers and codes assigned to the questions, then the file with the recoded item variables needs to be saved under a different file name. A copy of the macro that scores the items in SPSS is presented in Exhibit 9.2.

Exhibit 9.2 Extracted Sections of SPSS Macro SCOREIT Used to Convert Cognitive Item Response Codes to Correctness-Score Levels

```
SET MPRINT=ON.
DEFINE SCOREIT (Type  = !charend('/')      /
                Item  = !charend('/')      /
                RIGHT = !charend('/')      /
                nr    = !charend('/')      /
                na    = !charend('/')      /
                om    = !charend('/')      /
                other = !charend('/')      ).

!If (!UPCASE(!Type)=MC) !Then
    !Do !I !In(!Item).
Recode !I       (!RIGHT  =1)
                (!nr     =0)
                (!na     =sysmis)
                (!om     =0)
                (!other  =0)
                (Else    =0).
    !DoEnd.
!IfEnd.

!If (!UPCASE(!Type)=OE) !Then
    !Do !I !In(!Item).
Recode !I    (10 thru 19=1)
             (20 thru 29=2)
             (30 thru 39=3)
             (70 thru 79=0)
             (!nr     = 0)
             (!na     = sysmis)
             (!om     = 0)
             (!other  = 0)
             (Else    = 0).
!DoEnd.
!IfEnd.
!enddefine.

SCOREIT  Type  = MC  /
         Item  =  <list items where option A is the correct one>  /
         RIGHT  = 1  /  nr  = 6  /  na  = 8  /  om  = 9  /  other  = 7.

SCOREIT  Type  = MC  /
         Item  =  <list items where option B is the correct one>  /
         RIGHT  = 2  /  nr  = 6  /  na  = 8  /  om  = 9  /  other  = 7.

SCOREIT  Type  = MC  /
         Item  =  <list items where option C is the correct one>  /
         RIGHT  = 3  /  nr  = 6  /  na  = 8  /  om  = 9  /  other  = 7.

SCOREIT  Type  = MC  /
         Item  =  <list items where option D is the correct one>  /
         RIGHT  = 4  /  nr  = 6  /  na  = 8  /  om  = 9  /  other  = 7.

SCOREIT  Type  = MC  /
         Item  =  <list items where option E is the correct one>  /
         RIGHT  = 5  /  nr  = 6  /  na  = 8  /  om  = 9  /  other  = 7.

SCOREIT  Type  = OE  /
         Item  =  <list open-ended items>  /
         RIGHT  =  0 /  nr  = 96 /  na  = 98 /  om  = 99 /  other  = 90.
```

9.5 Basic Analyses with the TIMSS Data: Means, Percentages, Regression Coefficients and their JRR Standard Errors

In this section four macros that can be used to compute the correct standard errors of sampling and imputation are described, including examples in which these macros are used to replicate exhibits in the TIMSS 1999 international reports and almanacs.

Computing Sampling and Imputation Variance for Plausible Values Using SPSS (JACKPV.SPS)

This section presents example SPSS code that can be used to compute the JRR standard errors for mean plausible values and percentages. This code is provided in the form of an SPSS macro that computes the percentages of students within subgroups defined by a set of classification variables, the JRR standard errors of these percentages, the means for the groups on one of the achievement scales using plausible values, and the standard errors of these means including the sampling and imputation variance components.

The JACKPV.SPS macro operates as follows:

1. Computes a set of replicate weights specified using the parameters NJKZ, JKZ, and JKR.

2. Aggregates or summarizes the file data by computing the sum of the weights for each subgroup, the sum of the weights overall, and the sum of a weighted set of plausible values.

3. Computes the percentages of people within each group, their means on the plausible values, and their corresponding standard errors. The resulting working file contains the corresponding statistics. The data set FINAL in the default directory contains the corresponding statistics.

When using this macro, you need to specify a set of classification variables, the name of the plausible values and how many there are, the number of replicate weights to be generated, the variables that contain the sampling information such as JKZONE and JKREP, and the sampling weight that is to be used for the analysis. You will also need to specify the data file that contains the data that is to be processed.

You need to know some basic SPSS macro language in order to use JACKPV.SPS. The macro should be included in the program file where it is going to be used. If you are operating in batch mode, then the macro should be called in every batch. If you are using SPSS interactively then the macro should be called once at the beginning of the session and it will remain active throughout the session. If the session is terminated and restarted at a later time the macro should be called once again. Once the macro is included in a specific session the word “JACKPV” should not be used within that session because doing so will invoke the macro.

The macro is included in the program file where it will be used by issuing the following command under SPSS:

```
include '<path>jackpv.sps'.
```

where <path> points to the specific drive and directory where the macro JACKPV.SPS can be found. The macro requires that several parameters be specified when it is invoked. These parameters are:

INFILE	The name of the data file that contains the variables necessary for the analysis. If the path location is included as part of the file name, the name of the file has to be enclosed in quotes. It is important to emphasize that this system file must include only those cases that are of interest in the analysis. If you want to have specific cases excluded from the analysis, as for example students with missing variables or selected students from a specific grade, this should be done prior to invoking the macro.
CVAR	This lists the variables that are to be used to classify the students in the data file. This can be a single variable, or a list of variables. The maximum number of variables will depend mostly on the computer resources available to you at the time. It is recommended to always include the variable that identifies the country. At least one variable has to be specified, usually IDCNTRY.
PVS	These are the plausible values to be used in the analysis. The plausible values need to be specified in the form “Plausible Value 1 to Plausible Value 5” as in “BSSSCI01 TO BSSSCI05”. Although in most cases you will want to use all five available plausible values, the program will also work when fewer are specified. You should always use at least two plausible values.
NPV	This is the number of plausible values that will be used for the analysis. Generally you will want to use all five plausible values for the analysis although under some circumstances fewer can be used. You should always use at least two plausible values for any analysis.
JKZ	The variable that captures the assignment of the student to a particular sampling zone. The name of this variable in all TIMSS files is JKZONE.

JKR The variable that captures whether the case is to be dropped or have its weight doubled for the corresponding replicate weight. The name of this variable in all TIMSS files is JKREP.

NJKZ This indicates the number of replicate weights to be generated when computing the JRR error estimates. When conducting analyses using the data from all countries, the value of NJKZ should be set to 75, the maximum possible value. When you are working with the data for only one country, you should set the NJKZ argument to as many replicates as are needed in the country. If the data from two or more countries is being used for an analysis, then the largest number of jackknife zones should be used. When in doubt about what number to set the NJKZ parameter to, it should be set to 75. The error variance will always be estimated correctly if more replicate weights than necessary are computed, but will be underestimated if you specify fewer replicate weights than necessary.

WGT The sampling weight to be used in the analysis, generally TOTWGT when using the student files, or MATWGT, SCIWGT, or TCHWGT when using the teacher files.

The simplest and most straightforward way to invoke the macro is by using the conventional SPSS notation for invoking macros. This involves listing the macro name followed by the corresponding list of arguments for the analysis, each separated by a slash. For example, if the macro is invoked using

```
include "d:\timss\programs\jackpv.sps".

jackpv infile= bsgallm2  /
      cvar  = idcntry itsex /
      pvs   = BSSSCI01 TO BSSSCI05 /
      npv   = 5        /
      jkz   = JKZONE   /
      jkr   = JKREP    /
      njkz  = 75       /
      WGT   = TOTWGT   .
```

it will compute the mean science achievement scores using five plausible values, and their corresponding standard errors, for boys and girls within each country, using the variable TOTWGT as the sampling weight. It will also compute the percentages of boys and girls within the country, and their corresponding standard errors. The data will be read from the data set BSGALLM2, and the standard error of the statistics will be computed based on 75 replicate weights.

The file that contains these results is then called FINAL and is saved to the default directory being used by SPSS. The variables that are contained in this file are:

Classification Variables

Each of the classification variables is kept in the resulting file. In the above example there are two classification variables in the FINAL data set, IDCNTRY and ITSEX. There is one unique occurrence for each combination of the categories for these variables.

Weight Variable

Contains the estimate in the population that belongs to the groups defined by each specific combination of the classification variable categories. In our example this variable is called TOTWGT.

N

Contains the number of cases in the groups defined by each specific combination of categories for the classification variables. In the example this is the number of boys and girls in the sample for each country.

PCT

Contains the weighted percentages of people in the groups for the classification variable listed last, within the specific combination of the categories defined by the classification variables. In the example, it is the percentage of boys and girls within each country.

PCT_SE

Contains the standard errors of PCT computed using the JRR method for computing the standard error.

MNX

Contains the weighted means for the first plausible value for the groups defined by the corresponding combinations of classification variable categories.

MNX_SE

Contains the JRR standard errors of the means for the first plausible value for the groups computed using the JRR method for computing the standard error. This does not include the imputation error that should be computed when using plausible values.

MNPV

Contains the means of the plausible values used in the analysis.

MNPV_SE

Contains the standard errors for the means of the plausible values. These standard errors contain the sampling and the imputation components of the errors of the estimates.

The file resulting from using this macro can be printed using a SPSS procedure of choice. An example call to this macro, and a printout of the resulting file is presented in Exhibit 9.3 below. The code is included in the file SampleJackPV.SPS.

Exhibit 9.3 SPSS Control Code and Extract of Output File for Using the Macro JACKPV.SPS

```
get file = "d:\timss\data\bm2\bsgallm2.sav"
/ keep=idcntry idstud idgrader jkrep jkzone totwgt
itsex bsssci01 to bsssci05 intms99.

select if (itsex=1 or itsex=2) and intms99=1 and idgrader = 2.
save outfile = student.

include "d:\timss\programs\jackpv.sps".
jackpv infile= student  /
       cvar  = idcntry itsex /
       pvs   = BSSSCI01 TO BSSSCI05 /
       npv   = 5        /
       jkz   = JKZONE   /
       jkr   = JKREP    /
       njkz  = 75       /
       WGT   = TOTWGT   .

print formats idcntry itsex n (F6.0) totwgt (f10.0)
      mnpv mnpv_se mnx mnx_se pct pct_se (f6.2).

report format=list
 / var = idcntry itsex n totwgt mnpv mnpv_se mnx mnx_se pct pct_se.

IDCNTRY   ITSEX       N    TOTWGT    MNPV   MNPV_SE     MNX    MNX_SE     PCT    PCT_SE

    36       1    2033    132651    531.54    5.06    531.89    4.63     50.99    2.44
    36       2    1999    127478    549.33    5.96    548.72    5.27     49.01    2.44
   100       1    1676     45121    511.16    5.83    509.95    5.62     51.12    1.45
   100       2    1593     43146    525.04    6.51    526.72    6.25     48.88    1.45
   124       1    4392    185897    526.22    3.21    527.82    2.60     50.16     .98
   124       2    4373    184743    540.26    2.42    539.56    2.33     49.84     .98
   152       1    2981    104646    408.90    4.27    409.75    3.70     50.09    1.88
   152       2    2926    104263    431.89    5.14    431.30    4.81     49.91    1.88
   158       1    2874    155846    560.66    3.94    563.86    3.28     50.20    1.13
   158       2    2898    154582    577.56    5.66    580.82    4.80     49.80    1.13
   196       1    1528      4774    454.92    3.09    456.74    2.41     48.80     .58
   196       2    1587      5009    465.36    3.02    466.58    2.65     51.20     .58
.
.
.
.
.

   764       1    3138    395196    480.78    4.59    480.39    4.37     54.35    1.45
   764       2    2594    331890    484.15    4.44    483.84    4.25     45.65    1.45
   788       1    2561     71001    417.02    3.33    416.08    2.09     50.85     .76
   788       2    2490     68638    442.43    4.33    439.53    2.30     49.15     .76
   792       1    3301    260041    431.21    4.81    430.79    3.93     42.07    1.52
   792       2    4540    358016    434.21    4.26    433.24    3.63     57.93    1.52
   807       1    1988     14928    457.70    6.03    455.85    5.02     49.30     .84
   807       2    2035     15352    458.47    5.42    460.49    4.43     50.70     .84
   840       1    4617   1666060    505.40    4.59    506.79    4.17     49.94     .75
   840       2    4455   1670235    524.40    5.47    522.55    5.12     50.06     .75
   926       1    1455    268833    521.97    6.20    526.33    5.42     48.70    1.76
   926       2    1504    283133    554.12    5.25    555.33    5.00     51.30    1.76
   956       1    2644     32632    525.75    4.61    525.98    4.42     49.79    3.15
   956       2    2615     32907    543.89    7.19    542.03    6.50     50.21    3.15
```

In this example, the mean of all five plausible values and the mean of the first plausible value and their corresponding standard errors for science achievement are calculated separately for boys and girls by country. In the listing of the results we can see that there are entries for each country corresponding to the results for females

(ITSEX = 1) and males (ITSEX = 2) after selecting only those cases with IDGRADER=2. The first column has the country code, the second column indicates the gender of the students, the third column has the number of students in each sample, and the fourth column has the total weight of this sample in the population. This is followed by the mean of all five plausible values in science achievement and the corresponding standard error. Following this is the mean of the first plausible value in science achievement and its corresponding standard error. The last two columns indicate the percentage of girls and boys in each country's sample and their corresponding standard errors.

For example, Australia (idcntry 36) sampled 2033 girls representing 132651 students in the whole population. The mean of the five plausible values for these girls is 531.54 with a standard error of 5.06. Girls made up 50.99 percent of Australia's sampled students. The mean of the first plausible value for the girls is 531.89 with a standard error of 4.63. Additionally, Australia sampled 1999 boys representing 127478 students in the whole population. The mean of the five plausible values for these boys is 549.33 with a standard error of 5.96. Boys made up 49.01 percent of Australia's sampled students. The mean of the first plausible value for the boys is 548.72 with a standard error of 5.27.

Computing Sampling Variance for Variables Other Than Plausible Values Using SPSS (JACKGEN.SPS)

In this section example SPSS code that can be used to compute the JRR standard errors for means and percentages of variables other than plausible values is described. This code is provided in the form of an SPSS macro called JACKGEN.SPS that computes the percentages of students within subgroups defined by a set of classification variables, the JRR standard errors of these percentages, the means for the groups on a variable of choice, and the JRR standard errors of these means. Although you can compute weighted percent and mean estimates using other basic SPSS commands, the macro JACKGEN.SPS also computes the JRR error estimate for these means and percentages.

This macro operates as follows:

1. Computes a set of replicate weights, usually 75, for each record using the procedure described in the previous chapter.

2. Aggregates or summarizes the file data by computing the sum of the weights for each subgroup, the sum of the weights overall, and the sum of a weighted set of analysis variable values.

3. Computes the percentages of people within each group, their means on the analysis variable, and their corresponding standard errors. The

resulting working file, called FINAL, contains the corresponding statistics.

When using this macro, you need to specify a set of classification variables, one analysis variable, the number of replicate weights to be generated, the variables that contain the sampling information such as JKZONE and JKREP, and the sampling weight that is to be used for the analysis. You will also need to specify the data file that contains the data that is to be processed.

You need to know some basic SPSS macro language in order to use the macro. It needs to be first included in the program file where it is going to be used. If you are operating in batch mode, then the macro needs to be called in every batch. If you are using SPSS interactively then the macro needs to be called once at the beginning of the session and it will remain active throughout the session. If the session is terminated and restarted at a later time the macro needs to be called once again. Once the macro is included in a specific session the word "JACKGEN" should not be used within that program because doing so will invoke the macro.

The macro is included in the program file where it will be used by issuing the following command under SPSS:

```
include '<path>jackgen.sps'.
```

where <path> points to the specific drive and directory where the macro JACKGEN.SPS can be found. The macro requires that several parameters be specified when it is invoked. These parameters are:

INFILE	The name of the data file that contains the variables necessary for the analysis. If the path location is included as part of the file name the name of the file has to be enclosed in quotes. It is important to emphasize that this system file must include only those cases that are of interest in the analysis. If you want to have specific cases excluded from the analysis, as for example students with missing variables or selected students from a specific grade, this should be done prior to invoking the macro.
CVAR	This lists the variables that are to be used to classify the students in the data file. This can be a single variable, or a list of variables. The maximum number of variables will depend mostly on the computer resources available to you at the time. It is recommended to always include the variable that identifies the country. At least one variable has to be specified, usually IDCNTRY.
DVAR	This is the variable for which means are to be computed. Only one variable can be listed here. If you want to examine, for

example, results in two different variables, then the macro needs to be invoked separately to generate each table.

NJKZ — This indicates the number of replicate weights to be generated when computing the JRR error estimates. When conducting analyses using the data from all countries, the value of NJKZ should be set to 75. When you are working with the data for only one country, you should set the NJKZ argument to as many replicates as there were in the country. If the data from two or more countries are being used for an analysis, then the larger number of jackknife zones should be used. When in doubt about what number to set the NJKZ parameter to, it should be set to 75. The error variance will always be estimated correctly if more replicate weights than necessary are computed, but will be underestimated if you specify fewer replicate weights than necessary.

JKZ — The variable that captures the assignment of the student to a particular sampling zone. The name of this variable in all TIMSS files is JKZONE.

JKR — The variable that captures whether the case is to be dropped or have its weight doubled for the corresponding replicate weight. The name of this variable in all TIMSS files is JKREP.

WGT — The sampling weight to be used in the analysis, generally TOTWGT when using the student files, or MATWGT, SCIWGT, or TCHWGT when using the teacher files.

The simplest and most straightforward way to invoke the macro is by using the conventional SPSS notation for invoking macros. This involves listing the macro name followed by the corresponding list of arguments for the analysis, each separated by a slash. For example, if the macro is invoked using

```
include " d:\timss\programs\jackgen.sps".

jackgen
      infile = bsgallm2          /
      cvar   = idcntry idgrader /
      dvar   = bsdage           /
      njkz   = 75               /
      jkz    = jkzone           /
      jkr    = jkrep            /
      wgt    = totwgt           .
```

it will compute means of age and their corresponding standard errors, by grade, within each country, using the variable TOTWGT as the sampling weight. The data will be read from the system file BSGALLM2.

The file that contains these results is then called FINAL and is saved to the default directory being used by SPSS. The variables that are contained in this file are:

Classification Variables

Each of the classification variables is kept in the resulting file. In our example above there are two variables in the resulting system file. These are IDCNTRY and IDGRADER. There is one unique occurrence for each combination of the categories for these variables.

Weight Variable

Contains the estimates in the population that belong to the group defined by each specific combination of the classification variable categories.

MNX

Contains the weighted means of the variable DVAR for the groups defined by the corresponding combinations of classification variable categories.

MNX_SE

Contains the standard errors of the MNX values computed using the JRR method for computing the standard error.

PCT

Contains the weighted percentages of people in the groups for the classification variable listed last, within the specific combination of the categories defined by the groups initially. In our example, we would obtain the percentage of students by grade for each country.

PCT_SE

Contains the standard errors of PCT computed using the JRR method for computing the standard error.

The file resulting from using this macro can then be printed using a SPSS procedure of choice. An example call to this macro, and a subset of the resulting file, is presented in Exhibit 9.4. In this example the macro will compute the percentages in each grade, by country, and the mean achievement scores in science. This code is included in the file SampleJackGEN.SPS.

Exhibit 9.4 SPSS Control Code and Extract of Output File for Using the Macro JACKGEN.SPS

```
get file = "d:\timss\data\bm2\bsgallm2.sav"
   / keep=idcntry idstud intms99 idgrader jkrep jkzone totwgt
          bsdage.

select if not(missing(bsdage)) and intms99=1 and idgrader = 2.

save outfile = student.

include " d:\timss\programs\jackgen.sps".

jackgen
      infile = student          /
      cvar   = idcntry idgrader /
      dvar   = bsdage           /
      njkz   = 75               /
      jkz    = jkzone           /
      jkr    = jkrep            /
      wgt    = totwgt           .

print formats idcntry idgrader n (f6.0) totwgt (f10.0) mnx mnx_se pct pct_se (F6.2).

report format=list automatic
 / var   = idcntry idgrader n totwgt mnx mnx_se pct pct_se.

IDCNTRY    IDGRADER      N        TOTWGT       MNX      MNX_SE     PCT      PCT_SE
_______    ________    ______   __________   ______    ______    ______    ______

     36          2      3652       242266     14.26       .02    100.00       .00
    100          2      3267        88189     14.79       .04    100.00       .00
    124          2      8711       368224     14.03       .01    100.00       .00
    152          2      5898       208619     14.35       .02    100.00       .00
    158          2      5772       310429     14.20       .01    100.00       .00
    196          2      3015         9458     13.80       .01    100.00       .00
    203          2      3448       119207     14.40       .01    100.00       .00
    246          2      2906        59415     13.84       .01    100.00       .00
    344          2      5156        78764     14.24       .03    100.00       .00
    348          2      3149       110043     14.44       .02    100.00       .00
    360          2      5842      1953607     14.57       .03    100.00       .00
    364          2      5290      1651634     14.56       .03    100.00       .00
    376          2      4122        80028     14.05       .03    100.00       .00
    380          2      3328       548711     13.96       .01    100.00       .00
    392          2      4684      1397517     14.38       .00    100.00       .00
    400          2      5047        89069     14.00       .01    100.00       .00
    410          2      6113       609393     14.43       .01    100.00       .00
    428          2      2821        17845     14.50       .02    100.00       .00
    440          2      2346        40155     15.19       .02    100.00       .00
    458          2      5577       397761     14.36       .02    100.00       .00
    498          2      3643        58888     14.41       .02    100.00       .00
    504          2      5099       328317     14.19       .02    100.00       .00
    528          2      2890       193356     14.21       .03    100.00       .00
    554          2      3584        51110     14.03       .01    100.00       .00
    608          2      6593      1077070     14.07       .04    100.00       .00
    642          2      3419       259136     14.81       .02    100.00       .00
    643          2      4328      2055768     14.10       .02    100.00       .00
    702          2      4966        41346     14.35       .01    100.00       .00
    703          2      3475        72053     14.27       .01    100.00       .00
    705          2      3092        23387     14.76       .01    100.00       .00
    710          2      7601       776084     15.48       .10    100.00       .00
    764          2      5727       726455     14.46       .02    100.00       .00
    788          2      5042       139393     14.78       .03    100.00       .00
    792          2      7836       617822     14.20       .02    100.00       .00
    807          2      4018        30229     14.62       .01    100.00       .00
    840          2      8776      3234450     14.18       .02    100.00       .00
    926          2      2842       529162     14.20       .01    100.00       .00
    956          2      5221        64979     14.12       .02    100.00       .00
```

In this example BSDAGE is used to calculate the mean age of the sampled students in each country. In the listing of the results we can see that there is one entry or line for each of the values of the variable IDCNTRY after selecting only those cases with IDGRADER=2. The first column has the country code, the second column has the grades, the third column has the number of students in each sample after selecting only those cases where BSDAGE is not missing, and the fourth column is the total weight this sample represents in the whole population. This is followed by the mean age of the sampled students and the corresponding standard error.

For example, from line one in the above results, we can determine that Australia sampled 3652 students from the population 2 grade representing 242266 students in the whole population, with a mean age of 14.26 and a standard error of .02. Since there is only one grade in the category, each set represents 100 percent with a zero standard error.

Computing Regression Coefficients and Their JRR Standard Errors for Variables Other Than Plausible Values (JACKREG.SPS)

In this section, example SPSS code that may be used to compute regression coefficients and their JRR standard errors is described. The CD containing the TIMSS International Database contains the SPSS macro program called JACKREG.SPS. The macro computes the multiple correlation between the specified dependent and independent variables within a subgroup defined by a set of classification variables, as well as the regression coefficients and the JRR standard error of the regression coefficients.

If you wish to conduct regression analyses using plausible values as the dependent variable, please refer to JACKREGP.SPS, which is described in the next section.

The JACKREGP.SPS macro operates as follows:

1. Computes a set of replicate weights, usually 75, for each record using the procedure described in the previous chapter.
2. Aggregates or summarizes the file data by computing the sum of the weights for each category, the sum of the weights overall, and the sum of a weighted set of analysis variable values.
3. Computes the regression coefficients within each group and their corresponding standard errors. The resulting working file, FINAL, contains the corresponding statistics.

When using this macro, you need to specify a set of classification variables, the analysis variable, the number of replicate weights to be generated, the variables that contain the sampling information such as JKZONE and JKREP, and the sampling weight that is to be used for the analysis. You will also need to specify the data file that contains the data that is to be processed.

You need to know some basic SPSS macro language in order to use the macro. It needs to be first included in the program file where it is going to be used. If you are operating in batch mode, then the macro needs to be called in every batch. If you are using SPSS interactively then the macro needs to be called once at the beginning of the session and it will remain active throughout the session. If the session is terminated and restarted at a later time the macro needs to be called once again. Once the macro is included in a specific session the word "JACKREG" should not be used within that program because doing so will invoke the macro.

The macro is included in the program file where it will be used by issuing the following command under SPSS:

```
include '<path>jackreg.sps'.
```

where <path> points to the specific drive and directory where the macro JACKREG.SPS can be found. The macro requires that several parameters be specified when it is invoked. These parameters are:

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

CVAR — This lists the variables that are to be used to classify the students in the data file. This can be a single variable, or a list of variables. The maximum number of variables will depend mostly on the computer resources available to you at the time. It is recommended to always include the variable that identifies the country. At least one variable has to be specified, usually IDCNTRY.

XVAR — This is a list of independent variables, at least one, that under the regression model will be used as predictors of the dependent variable specified in DVAR. These independent variables can be continuous or categorical, or any other type of coded variable. For example, it could be the variable ITSEX as originally coded in the data files, or dummy coded as 1 or 0.

DVAR — This is the dependent variable that under the regression model is predicted by the variable or variables specified by the XVAR parameter. Only one variable can be listed here. If you want to use the same set of predictor variables to predict two different variables, then the macro needs to be invoked separately to generate each set of results.

NJKZ This indicates the number of replicate weights to be generated when computing the JRR error estimates. When conducting analyses using the data from all countries, the value of NJKZ should be set to 75. When you are working with the data for only one country, you should set the NJKZ argument to as many replicates as there were in the country. If the data from two or more countries is being used for an analysis, then the larger number of jackknife zones should be used. When in doubt about what number to set the NJKZ parameter to, it should be set to 75. The error variance will always be estimated correctly if more replicate weights than necessary are computed, but will be underestimated if you specify fewer replicate weights than necessary.

JKZ The variable that captures the assignment of the student to a particular sampling zone. The name of this variable in all TIMSS files is JKZONE.

JKR The variable that captures whether the case is to be dropped or have its weight doubled for the corresponding replicate weight. The name of this variable in all TIMSS files is JKREP.

WGT The sampling weight to be used in the analysis, generally TOTWGT when using the student files, or MATWGT, SCIWGT, or TCHWGT when using the teacher files.

The simplest and most straightforward way to invoke the macro is by using the conventional SPSS notation for invoking macros. This involves listing the macro name followed by the corresponding list of arguments for the analysis, each separated by a slash. For example, if the macro is invoked using

```
include "d:\timss\programs\jackreg.sps".

jackreg  infile  = bsgallm2    /
   cvar    = idcntry idgrader /
   xvar    = regsex     /
   dvar    = bsdgstdt  /
   njkz    = 75         /
   jkz     = jkzone     /
   jkr     = jkrep      /
   wgt     = totwgt     .
```

it will compute the regression equation for the variable REGSEX as a predictor of the number of hours spent studying. The data will be read from the data set BSGALLM2 and the standard error of the statistics will be computed based on 75 replicate weights.

The file that contains these results is called REG and can be found in the default directory. The variables that are contained in this file are:

Classification Variables

Each of the classification variables is kept in the resulting file. In our example above there would be two variables in the resulting system file, IDCNTRY and IDGRADER. There is one unique occurrence for each combination of the categories for these variables.

Mult_RSQ

The squared multiple correlation coefficient for the model.

SS_Res, SS_Reg, SS_Total

The residual, regression, and total sum of squares for the model within each group as defined by the classification variables.

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

These are the regression coefficients for each of the predictor variables in the model and their corresponding jackknifed standard errors. The coefficient zero (B00) is the intercept for the model. The other coefficients receive a sequential number starting with 01. This sequential number corresponds to the order of the variables in the list of variables specified in the parameter XVAR.

The file resulting from using this macro can then be printed using a SPSS procedure of choice. An example call to this macro, and a subset of the resulting file is presented in Exhibit 9.5 This code is included in the file SampleJackREG.SPS.

Exhibit 9.5 SPSS Control Code and Extract of Output File for Using the Macro JACKREG.SPS

```
get file = "d:\timss\data\bm2\bsgallm2.sav" .

select if not(missing(itsex)).
compute regsex = itsex-1.

save outfile = student.

include "d:\timss\programs\jackreg.sps".

jackreg  infile  = student    /
   cvar    = idcntry idgrader /
   xvar    = regsex     /
   dvar    = bsdgstdt  /
   njkz    = 75        /
   jkz     = jkzone    /
   jkr     = jkrep     /
   wgt     = totwgt    .

select if idgrader = 2.

print formats idcntry n (F6.0) Mult_RSQ (f5.3)
      SS_Total SS_Reg SS_Res (F10.0) B00 B00.SE B01 B01.SE (f6.2).

list vars = idcntry n Mult_RSQ SS_Total SS_Reg SS_Res B00 B00.SE B01 B01.SE.

IDCNTRY      N  MULT_RSQ   SS_TOTAL     SS_REG     SS_RES    B00 B00.SE    B01 B01.SE

     36   3841     .009     453751       4186     442443   2.08    .05   -.26    .06
    100   2792     .049     248160      12204     235956   3.39    .08   -.80    .09
    124   8371     .019     775003      14716     760287   2.43    .05   -.41    .05
    152   5317     .015     525593       7775     517818   2.63    .05   -.41    .06
    158   5489     .016     836783      13569     823214   2.20    .06   -.43    .05
    196   2848     .043      27643       1187      26456   3.22    .06   -.73    .06
    203   3382     .042     165295       6927     158369   2.13    .05   -.49    .05
    246   2852     .012      48630        581      48049   1.88    .03   -.20    .04
    344   4970     .003     176706        554     176151   1.73    .06   -.17    .07
    348   3001     .051     257416      13006     244410   3.17    .05   -.71    .06
    360   5285     .021    5378045     115350    5262695   3.29    .05   -.51    .06
    364   3802     .010    3985216      41262    3943955   4.32    .07   -.39    .09
    376   3660     .020     183793       3694     180099   2.90    .06   -.46    .07
    380   3029     .055    1652205      91651    1560554   4.03    .05   -.86    .07
.
.
.
.
.
    643   3703     .020    4986756     101575    4885180   3.39    .05   -.48    .05
    702   4505     .024     118597       2799     115797   3.81    .05   -.55    .06
    703   3365     .051     123868       6348     117520   2.60    .05   -.60    .07
    705   2984     .039      51716       2009      49707   2.77    .05   -.60    .06
    710   5537     .001    2225748       1689    2224060   3.19    .07   -.11    .06
    764   5296     .026    1712802      44822    1667980   3.15    .04   -.52    .05
    788   3753     .010     397338       3855     393483   3.94    .06   -.39    .07
    792   6286     .021    1728855      37148    1691707   3.94    .05   -.55    .05
    807   3125     .019      77383       1437      75946   3.77    .05   -.50    .07
    840   8430     .031    6852641     214619    6638022   2.34    .04   -.52    .03
    956   4861     .029     148098       4274     143824   3.22    .05   -.53    .07


Number of cases read:  37    Number of cases listed:  37
```

In this example the variable REGSEX is created by subtracting one from the variable ITSEX. As a result the girls receive a code of 0 and the boys receive a code of 1 on this variable. In this particular model the variable REGSEX is used to predict the values of the variable BSDGSTDT by country and by grade. In the listing of the results we can see that there is one entry or line for each value of the variable IDCNTRY after selecting only those cases with IDGRADER=2. The first column has the country code and the second column has the number of students in each sample. Keep in mind that this number corresponds only to those students who had valid data for the variables in the regression model. Then the multiple R squared is listed for each country, followed by the sums of squares for the model and corresponding regression coefficients. Because of the way in which the variable REGSEX is coded, the constant (B00) is the mean value for the variable BSDSTDT for the girls and the error of the constant (B00_SE) is standard error of the estimate. The first regression coefficient (B01) is the difference in the average value of the variable BSDSTDT between the boys and the girls, and the standard error of this coefficient (B01_SE) is the jackknifed standard error of this difference.

For example, from the first line in the output we can say that in Australia (IDCNTRY = 36) data was available for 3841 cases. The coefficient of determination between gender and hours of study is .009 with girls studying, on average, about 2.08 hours per day and boys studying 15 minutes (0.26) less than girls on average. This difference is statistically significant (0.26 / .06 = 4.3).

Overall we can say that on average, and across all countries, girls report studying more hours than boys. This difference is significant in most cases, which can be determined by dividing the value of B01 by its standard error and comparing it to the appropriate critical value.

Computing Regression Coefficients and Their JRR Standard Errors with Plausible Values (JACKREGP.SPS)

In this section example SPSS code that can be used to compute the JRR standard errors for regression coefficients using plausible values as the dependent variable is described. This code is provided in the form of an SPSS macro called JACKREGP.SPS. This macro computes the multiple correlation coefficient between the specified plausible values and independent variables within subgroups defined by a set of classification variables, as well as the regression coefficients and their JRR standard errors.

The JACKREGP.SPS macro operates as follows:

1. Computes a set of replicate weights, usually 75, for each record using the procedure described in the previous chapter.

2. Aggregates or summarizes the file data by computing the corresponding sum of the weights for each category, the sum of the weights overall, and the sum of a weighted set of analysis variable values.

3. Computes the regression coefficients within each group and their corresponding standard errors. The resulting working file, called FINAL, contains the corresponding statistics.

When using this macro, you need to specify a set of classification variables, the analysis variable, the number of replicate weights to be generated, the variables that contain the sampling information such as JKZONE and JKREP, and the sampling weight that is to be used for the analysis. You will also need to specify the data file that contains the data that is to be processed.

You need to know some basic SPSS macro language in order to use the macro. It needs to be first included in the program file where it is going to be used. If you are operating in batch mode, then the macro needs to be called in every batch. If you are using SPSS interactively then the macro needs to be called once at the beginning of the session and it will remain active throughout the session. If the session is terminated and restarted at a later time the macro needs to be called once again. Once the macro is included in a specific session the word "JACKREGP" should not be used within that program because doing so will invoke the macro.

The macro is included in the program file where it will be used by issuing the following command under SPSS:

```
include '<path>jackregp.sps'.
```

where <path> points to the specific drive and directory where the macro JACKREGP.SPS can be found. The macro requires that several parameters be specified when it is invoked. These parameters are:

INFILE	The name of the data file that contains the variables necessary for the analysis. If the path location is included as part of the file name, the name of the file has to be enclosed in quotes. It is important to emphasize that this system file must include only those cases that are of interest in the analysis. If you want to have specific cases excluded from the analysis, as for example students with missing variables or selected students from a specific grade, this should be done prior to invoking the macro.
CVAR	This lists the variables that are to be used to classify the students in the data file. This can be a single variable, or a list of variables. The maximum number of variables will depend mostly on the computer resources available to you at the time. It is recommended to always include the variable that identifies the

country. At least one variable has to be specified, usually IDCNTRY.

XVAR	This is a list of independent variables, possibly one, which under the regression model will be used as predictors of the dependent variables specified by the plausible values. These independent variables can be continuous or categorical, or any other type of coded variable. For example, it could be the variable ITSEX as originally coded in the data files, or dummy coded as 1 or 0.
ROOTPV	This is the prefix used to identify the plausible values for the achievement scale of interest. This corresponds to the first 7 characters of the plausible value variables. For example, the root of the overall science plausible value is "BSSSCI0," the root of the geometry plausible values is "BSMGEO0."
NPV	This is the number of plausible values that will be used in the analysis. Generally you will want to use all five plausible values for the analysis although under some circumstances fewer can be used. You should always use at least two plausible value for any analysis.
NJKZ	This indicates the number of replicate weights to be generated when computing the JRR error estimates. When conducting analyses using the data from all countries, the value of NJKZ should be set to 75. When you are working with the data for only one country, you should set the NJKZ argument to as many replicates as there were in the country. If the data from two or more countries is being used for an analysis, then the larger number of jackknife zones should be used. When in doubt about what number to set the NJKZ parameter to, it should be set to 75. The error variance will always be estimated correctly if more replicate weights than necessary are computed, but will be underestimated if you specify fewer replicate weights than necessary.
JKZ	The variable that captures the assignment of the student to a particular sampling zone. The name of this variable in all TIMSS files is JKZONE.
JKR	The variable that captures whether the case is to be dropped or have its weight doubled for the corresponding replicate weight. The name of this variable in all TIMSS files is JKREP.
WGT	The sampling weight to be used in the analysis, generally TOTWGT when using the student files, or MATWGT, SCIWGT, or TCHWGT when using the teacher files.

The simplest and most straightforward way to invoke the macro is by using the conventional SPSS notation for invoking macros. This involves listing the macro name followed by the corresponding list of arguments for the analysis, each separated by a slash. For example, if the macro is invoked using

```
include "d:\timss\programs\jackregP.sps".

jackregp  infile  = bsgallm2  /
   cvar     = idcntry idgrader  /
   xvar     = regsex    /
   rootpv   = bsssci0   /
   npv      = 5         /
   njkz     = 75        /
   jkz      = jkzone    /
   jkr      = jkrep     /
   wgt      = totwgt    .
```

it will compute the regression equation for the variable REGSEX as a predictor of the plausible values in science. The data will be read from the data set BSGALLM2 and the standard error of the statistics will be computed based on 75 replicate weights.

The file that contains these results is then called REG and is saved to the default directory being used by SPSS. The variables that are contained in this file are:

Classification Variables

Each of the classification variables is kept in the resulting file. In our example above there would be two variables in the resulting system file, IDCNTRY and IDGRADER. There is one unique occurrence for each combination of the categories for these variables.

Mult_RSQ

The squared multiple correlation coefficient for the model.

SS_Res, SS_Reg, SS_Total

The residual, regression, and total sum of squares for the model within each group as defined by the classification variables.

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

These are the regression coefficients for each of the predictor variables in the model and their corresponding jackknifed standard errors combined with the imputation error. The coefficient zero (B00) is the intercept for the model. The other coefficients receive a sequential number starting with 01. This sequential number corresponds to the order of the variables in the list of variables specified in the parameter XVAR.

The file resulting from using this macro can then be printed using a SPSS procedure of choice. An example call to this macro, and a subset of the resulting file, is presented in Exhibit 9.6. This code is included in the file SampleJackREGP.SPS.

Exhibit 9.6 SPSS Control Code and Extract of Output File for Using the Macro JACKREGP.SPS

```
get file = "d:\timss\data\bm2\bsgallm2.sav" .

select if not(missing(itsex)).
compute regsex = itsex-1.

save outfile = student.

include "d:\timss\programs\jackregP.sps".

jackregp  infile  = student  /
   cvar     = idcntry idgrader  /
   xvar     = regsex    /
   rootpv  = bsssci0   /
   npv      = 5         /
   njkz     = 75        /
   jkz       = jkzone    /
   jkr       = jkrep     /
   wgt      = totwgt    .

print formats idcntry n (F6.0) Mult_R Mult_RSQ (f5.3)
      SS_Total SS_Reg SS_Res (F12.0) B00 B00.SE B01 B01.SE (f6.2).

select if idgrader = 2.
list vars = idcntry N Mult_R Mult_RSQ SS_Reg SS_Total SS_Res b00 b00.se b01 b01.se .
IDCNTRY    N  MULT_RSQ     SS_REG     SS_TOTAL       SS_RES    B00 B00.SE    B01 B01.SE
   36   4032     .011    21533145   1987336973   1965803828 531.54   5.06  17.80   6.85
  100   3269     .006     4404020    766810619    762406599 511.16   5.83  13.88   6.15
  124   8765     .008    18552616   2247376884   2228824268 526.22   3.21  14.04   3.87
  152   5907     .017    28127046   1610357882   1582230836 408.90   4.27  23.00   6.18
  158   5772     .009    22268514   2454575435   2432306921 560.66   3.94  16.90   4.21
  196   3115     .004      270839     69085171     68814332 454.92   3.09  10.44   3.90
  203   3453     .044    33353184    768755509    735402325 523.26   4.83  33.33   4.76
  246   2920     .004     1622161    365451796    363829635 530.13   4.00  10.23   4.97

  344   5178     .011     4226667    383047085    378820418 522.21   4.38  14.43   6.12
  348   3166     .023    17660238    784863720    767203483 540.08   4.05  25.20   4.20
  360   5848     .011   157243045  13942407927  13785164882 426.97   6.46  17.20   6.78
  364   5301     .033   386210898  11705877120  11319666222 429.58   5.66  31.00   7.59
  376   4192     .005     4418074    897239201    892821127 461.15   6.02  14.39   6.09
  380   3328     .011    47595532   4198936976   4151341445 484.39   4.13  18.17   5.83
  392   4745     .009    69695742   8163373496   8093677754 542.75   2.79  13.63   4.59
  400   5052     .007     7002932    941584856    934581924 459.73   5.03 -17.66   8.22
  410   6114     .015    67765990   4384393239   4316627249 538.11   3.98  20.87   5.07
  428   2872     .010     1064028    110217611    109153583 495.29   5.57  15.20   3.96
  440   2361     .017     4611645    278696318    274084673 477.85   4.44  21.29   4.57
  458   5577     .004    10038330   2665511485   2655473156 488.11   5.47   9.47   7.02
  498   3711     .003     1845789    543263230    541417440 454.21   4.37  10.63   5.44
  504   5392     .008    30338714   3640771963   3610433250 312.26   5.94  17.95   8.31
  528   2958     .014    16409182   1162465799   1146056617 536.21   7.13  18.07   4.06
  554   3613     .002      677691    446197763    445520073 506.06   5.45   7.03   7.82
  608   6599     .003    46484473  15698586279  15652101806 350.90   8.25 -12.14   8.37
  642   3424     .002     3821308   2453033016   2449211708 468.20   6.43   7.21   5.44
  643   4332     .012   213644280  17946910067  17733265787 519.46   7.09  20.35   3.94
  702   4966     .011     4392472    390675927    386283454 557.36   7.93  20.46   7.88
  703   3497     .019     8197697    438359904    430162207 524.55   3.38  21.14   4.46
  705   3109     .006     1033535    165989537    164956001 526.94   3.69  13.15   3.66
  710   8146     .005    75756385  14804099416  14728343031 233.79   9.19  18.87   6.69
  764   5732     .001     2435093   3919157529   3916722436 480.78   4.59   3.37   4.27
  788   5051     .036    22724779    623803249    601078470 417.02   3.33  25.40   3.42
  792   7841     .000     1463639   3937655073   3936191434 431.21   4.81   3.00   2.87
  807   4023     .000       37572    286035856    285998284 457.70   6.03    .77   4.59
  840   9072     .010   306803783  31684805640  31378001857 505.40   4.59  19.00   4.13
  926   2959     .031   143197546   4560223136   4417025590 521.97   6.20  32.15   6.63
  956   5259     .017     5520141    316472055    310951913 525.75   4.61  18.13  10.34

Number of cases read:  38    Number of cases listed:  38
```

In this example the variable REGSEX is created by subtracting one from the variable ITSEX. As a result the girls receive a code of 0 and the boys receive a code of 1 on this variable. In this particular model the variable REGSEX is used to predict the plausible values of science achievement by country. In the listing of the results we can see that there is one entry or line for each value of the variable IDCNTRY after selecting only those cases with IDGRADER=2. The first column has the country code and the second column has the number of students in each sample. Because of the way in which REGSEX is coded, the constant (B00) is the mean science achievement for the girls, the error of the constant (B00.SE) is the standard error of this estimate. The first regression coefficient (B01) is the difference in the average value in science achievement between the boys and the girls, and the standard error of this coefficient (B01.SE) is the combined jackknifed sampling error and imputation error of this difference.

For example, from the first line in the output we can say that in Australia (IDCNTRY = 36) data was available for 4032 cases. The squared of the multiple correlation between gender and science achievement was .011 with girls achieving on average 531.54 and boys scoring on average 17.80 points higher. This difference is not statistically significant (17.80 / 6.85 = 2.6).

9.6 Replicating Analyses from the TIMSS 1999 International Reports: Student Level

Many analyses of the TIMSS data can be undertaken using student-level data. We have already presented some examples in the previous sections when explaining how to use the macros provided with the data files. We now proceed to work through additional examples of actual analyses from the TIMSS 1999 international reports, where all the steps are undertaken, including the invocation of the corresponding SPSS macro.

Example Analysis with Student-Level Variables Not Using Plausible Values

In the first example, we want to replicate the analysis of eighth graders' reports on the number of hours studied each day. These results, originally appear in Exhibit R1.11 of the TIMSS 1999 International Science Report, are reproduced in Exhibit 9.7. Since the results in this exhibit are not based on plausible values and we want to report the average value of the variable BSDGSTDT, we need to use the macro JACKGEN.

Exhibit 9.7 Sample Exhibit for Student-Level Analysis NOT Using Plausible Values Taken From the TIMSS 1999 International Science Report

	Average Hours Spent Each Day Studying or Doing Homework[1]				Percentage of Students Reporting Spending Some Time Studying All Three Subjects: Science, Mathematics, and Other
	Science	Mathematics	Other School Subjects	Total	
Australia	0.6 (0.02)	0.7 (0.02)	0.8 (0.02)	2.0 (0.04)	74 (1.6)
Belgium (Flemish)	0.8 (0.03)	1.1 (0.03)	1.4 (0.04)	2.9 (0.05)	86 (1.2)
Bulgaria	1.1 (0.03)	1.1 (0.04)	1.3 (0.04)	3.0 (0.06)	74 (1.9)
Canada	0.6 (0.01)	0.8 (0.02)	1.0 (0.02)	2.2 (0.04)	78 (1.0)
Chile	0.9 (0.02)	0.9 (0.02)	1.2 (0.03)	2.4 (0.04)	75 (1.0)
Chinese Taipei	0.6 (0.02)	0.7 (0.02)	1.0 (0.02)	2.0 (0.05)	55 (1.3)
Cyprus	0.7 (0.02)	1.1 (0.03)	1.5 (0.03)	2.8 (0.04)	79 (0.8)
Czech Republic	0.6 (0.02)	0.7 (0.02)	0.7 (0.02)	1.9 (0.04)	74 (1.4)
England	– –	– –	– –	– –	– –
Finland	0.5 (0.01)	0.6 (0.01)	0.7 (0.01)	1.8 (0.02)	90 (0.8)
Hong Kong, SAR	0.5 (0.01)	0.7 (0.02)	0.7 (0.02)	1.6 (0.04)	53 (1.3)
Hungary	1.1 (0.02)	0.8 (0.02)	1.2 (0.03)	2.8 (0.04)	90 (0.8)
Indonesia	1.1 (0.02)	1.2 (0.03)	1.3 (0.02)	3.0 (0.05)	83 (1.0)
Iran, Islamic Rep.	1.6 (0.03)	1.9 (0.03)	2.0 (0.04)	r 4.0 (0.05)	92 (0.5)
Israel	0.8 (0.02)	1.1 (0.03)	1.4 (0.04)	2.7 (0.05)	79 (0.9)
Italy	1.0 (0.02)	1.3 (0.03)	1.9 (0.03)	3.6 (0.04)	91 (0.8)
Japan	0.4 (0.01)	0.6 (0.01)	0.8 (0.02)	1.7 (0.04)	59 (1.4)
Jordan	1.5 (0.03)	1.7 (0.03)	2.4 (0.05)	r 3.7 (0.06)	r 87 (0.9)
Korea, Rep. of	0.4 (0.01)	0.6 (0.02)	0.7 (0.02)	1.6 (0.03)	50 (0.9)
Latvia (LSS)	0.8 (0.02)	1.0 (0.02)	1.5 (0.03)	3.0 (0.04)	89 (0.7)
Lithuania ‡	0.8 (0.02)	0.9 (0.03)	1.5 (0.04)	2.8 (0.04)	89 (1.0)
Macedonia, Rep. of	2.0 (0.05)	1.2 (0.03)	1.5 (0.04)	r 3.4 (0.05)	90 (0.5)
Malaysia	1.3 (0.02)	1.6 (0.02)	1.8 (0.03)	3.8 (0.04)	94 (0.4)
Moldova	1.7 (0.04)	1.1 (0.03)	1.4 (0.04)	r 3.3 (0.05)	83 (0.8)
Morocco	r 1.5 (0.06)	r 1.7 (0.07)	r 1.8 (0.06)	s 3.1 (0.05)	s 77 (1.3)
Netherlands	0.6 (0.02)	0.6 (0.02)	1.0 (0.02)	2.2 (0.04)	89 (1.1)
New Zealand	0.6 (0.02)	0.7 (0.02)	0.9 (0.02)	2.0 (0.04)	76 (1.3)
Philippines	1.7 (0.04)	1.7 (0.04)	2.1 (0.04)	r 3.3 (0.04)	88 (0.7)
Romania	1.2 (0.03)	1.6 (0.05)	1.4 (0.04)	3.4 (0.06)	77 (1.2)
Russian Federation	1.5 (0.03)	1.1 (0.03)	1.2 (0.04)	3.1 (0.05)	89 (0.7)
Singapore	1.2 (0.02)	1.3 (0.02)	1.7 (0.03)	3.5 (0.04)	90 (0.8)
Slovak Republic	0.8 (0.02)	0.8 (0.02)	0.9 (0.02)	2.3 (0.03)	88 (0.8)
Slovenia	0.9 (0.02)	0.8 (0.02)	0.9 (0.02)	2.5 (0.03)	85 (1.0)
South Africa	1.5 (0.05)	1.8 (0.04)	2.0 (0.06)	r 3.1 (0.06)	71 (1.9)
Thailand	1.0 (0.02)	1.1 (0.02)	1.2 (0.02)	2.9 (0.04)	88 (0.6)
Tunisia	1.2 (0.03)	1.8 (0.03)	2.1 (0.03)	r 3.6 (0.04)	82 (0.8)
Turkey	1.2 (0.02)	1.2 (0.02)	1.9 (0.03)	3.5 (0.05)	90 (0.7)
United States	0.6 (0.01)	0.8 (0.02)	0.9 (0.02)	2.1 (0.04)	72 (1.6)
International Avg.	1.0 (0.00)	1.1 (0.00)	1.3 (0.01)	2.8 (0.01)	80 (0.2)

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1998-1999.

Background data provided by students.

1 Average hours based on: No time=0; less than 1 hour=.5; 1-2 hours=1.5; 3-5 hours=4; more than 5 hours=7.

‡ Lithuania tested the same cohort of students as other countries, but later in 1999, at the beginning of the next school year.

( ) Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

A dash (–) indicates data are not available.

An "r" indicates a 70-84% student response rate. An "s" indicates a 50-69% student response rate.

To replicate the results in this exhibit, we need to undertake several steps. After reviewing the codebooks and the questionnaire information we find out the students were asked three questions about the number of hours they spend studying mathematics, science, and other subjects. The data collected from these variables are summarized in the variable BSDGSTDT, and this variable is found in the Student Background data file. 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 changes were made we can continue with our analysis without any modifications.

We then proceed to read, from the SPSS system file that contains this variable, the sampling weight for the student (TOTWGT), the variables that contain the jackknife replication information (JKZONE and JKREP), the variable that will be used to select the eighth graders from the data file (IDGRADER), and the variable containing the country identification code (IDCNTRY). In this analysis we will use the data for all countries in the database, although the exact same steps need to be taken if you want to examine these variables within a single country or for a select group of countries.

The SPSS code is presented in Exhibit 9.8 and is included in the CD under the name EXAMPLE1.SPS. Selections of the results obtained from this program are displayed in Exhibit 9.9. We have included as part of the program the corresponding value labels and format statements so that the different categories or groups are labeled appropriately.

Notice that one of the steps in this program is to select only those students in the eighth grade who have non-missing data in the variable.

Note also in this analysis that we have used the data from all the countries, although we did select from the system file only those variables that are relevant to our analysis. In general, this type of analysis is quite feasible with a powerful desktop computer; however, you need to keep in mind that computing and storage requirements for these types of analysis are quite demanding.

In general, to perform analyses such as those using the Student Background data files, you need to do the following:

- Identify the variable or variables of interest in the student file and find out about any specific national adaptations to the variable.
- Retrieve the relevant variables from the data files, including the sampling weights, JRR replication information, and any other variables used in the selection of cases.
- Use the macro JACKGEN with the corresponding arguments and parameters.
- Print out the result file.

Exhibit 9.8 SPSS Control Statements for Performing Analyses with Student-Level Variables NOT Using Plausible Values (EXAMPLE1.SPS)

```
get file = "d:\timss\data\bm2\bsgallm2.sav"
   / keep=idcntry idstud intms99 idgrader jkrep jkzone totwgt
          bsdgstdt.

select if not(missing(bsdgstdt)) and intms99=1 and idgrader = 2.

value labels
 idcntry
   036 'AUSTRALIA         '   956 'BELGIUM (Flemish) '   100 'BULGARIA          '
   124 'CANADA            '   152 'CHILE             '   196 'CYPRUS            '
   203 'CZECH REP.        '   926 'ENGLAND           '   246 'FINLAND           '
   344 'HONG KONG         '   348 'HUNGARY           '   360 'INDONESIA         '
   364 'IRAN, ISLAMIC REP.'   376 'ISRAEL            '   380 'ITALY             '
   400 'JORDAN            '   392 'JAPAN             '   410 'KOREA, REP. OF    '
   440 'LITHUANIA         '   428 'LATVIA            '   504 'MOROCCO           '
   498 'MOLDOVA, REP. OF  '   807 'MACEDONIA, REP. OF'   458 'MALAYSIA          '
   528 'NETHERLANDS       '   554 'NEW ZEALAND       '   608 'PHILIPPINES       '
   642 'ROMANIA           '   643 'RUSSIAN FEDERATION'   702 'SINGAPORE         '
   703 'SLOVAK REP.       '   705 'SLOVENIA          '   764 'THAILAND          '
   788 'TUNISIA           '   792 'TURKEY            '   158 'CHINESE TAIPEI    '
   840 'UNITED STATES     '   710 'SOUTH AFRICA      ' .

save outfile = student.

* Now use the macro to get the results.
include " d:\timss\programs\jackgen.sps".

jackgen
      infile = student          /
      cvar   = idcntry          /
      dvar   = bsdgstdt         /
      njkz   = 75               /
      jkz    = jkzone           /
      jkr    = jkrep            /
      wgt    = totwgt    .

sort cases by idcntry .

print format n (f6.0) totwgt (f10.0) mnx mnx_se pct pct_se (f6.2).

report format=list automatic
 / var   = n totwgt mnx mnx_se pct pct_se
 / break = idcntry .
```

Exhibit 9.9 Extract of SPSS Computer Output for Performing Analyses with Student-Level Variables Not Using Plausible Values (EXAMPLE 1)

IDCNTRY	N	TOTWGT	MNX	MNX_SE	PCT	PCT_SE
AUSTRALIA	3841	247322	1.96	.04	1.56	.04
BULGARIA	2794	75854	3.00	.06	.48	.04
CANADA	8376	354966	2.22	.04	2.24	.05
CHILE	5317	187079	2.43	.04	1.18	.03
CHINESE TAIPEI	5489	294996	1.99	.05	1.86	.03
CYPRUS	2848	8956	2.84	.04	.06	.00
CZECH REP.	3382	116430	1.90	.04	.73	.03
FINLAND	2852	58237	1.79	.02	.37	.01
HONG KONG	4971	75898	1.64	.04	.48	.01
HUNGARY	3005	104755	2.82	.04	.66	.01
INDONESIA	5285	1762776	3.04	.05	11.13	.32
IRAN, ISLAMIC REP.	3802	1173708	4.08	.04	7.41	.27
ISRAEL	3660	70994	2.68	.05	.45	.01
ITALY	3029	498906	3.60	.04	3.15	.09
JAPAN	4640	1384517	1.73	.04	8.74	.13
JORDAN	3278	57966	3.84	.06	.37	.01
KOREA, REP. OF	5919	589907	1.62	.03	3.72	.05
LATVIA	2650	16684	3.02	.04	.11	.00
LITHUANIA	2186	37268	2.81	.04	.24	.00
MALAYSIA	4843	346308	3.84	.04	2.19	.05
MOLDOVA, REP. OF	2865	46358	3.35	.05	.29	.01
MOROCCO	3037	192250	3.31	.06	1.21	.04
NETHERLANDS	2841	189313	2.22	.04	1.20	.09
NEW ZEALAND	3425	48819	2.00	.04	.31	.01
PHILIPPINES	4528	733526	3.33	.04	4.63	.13
ROMANIA	2791	212858	3.41	.06	1.34	.05
RUSSIAN FEDERATION	3703	1752345	3.16	.05	11.06	.43
SINGAPORE	4505	37567	3.52	.04	.24	.00
SLOVAK REP.	3365	69454	2.30	.03	.44	.01
SLOVENIA	2984	22552	2.48	.03	.14	.00
SOUTH AFRICA	5537	583547	3.14	.06	3.68	.14
THAILAND	5296	670705	2.92	.04	4.23	.08
TUNISIA	3753	103689	3.74	.04	.65	.02
TURKEY	6286	498382	3.62	.05	3.15	.10
MACEDONIA, REP. OF	3125	23464	3.52	.05	.15	.00
UNITED STATES	8430	3131090	2.08	.04	19.76	.86
BELGIUM (Flemish)	4861	60145	2.95	.05	.38	.01

In this example, each country's mean value for BSDGSTDT is reported for the eighth grade. The results are presented by country after selecting only those cases with IDGRADER=2. The country is in the first column. The second column has the number of students sampled in the country. The third column corresponds to the weights given these sampled students in the whole population, followed by their mean for BSDGSTDT and the corresponding standard errors. The last two columns represent the percentages of students sampled responding and the corresponding standard error.

For example, from the first line of the report, we can say that in Australia data was available for 3841 cases. These students represent a population of 247322 students. Australian students spend, on average, 1.96 hours studying each day. The standard error of this mean is .04.

Example Analysis with Student-Level Variables Using Plausible Values

In this example, we want to replicate another one of the results presented in the international report. We are interested in looking at the eighth graders' reports on the number of books in their home and their achievement in science. These are the results that are presented in Exhibit 9.10 and appear in Exhibit R1.3 of the TIMSS 1999 International Science Report. Since the results in this exhibit are based on plausible values, we need to use the macro JACKPV.

Exhibit 9.10 Sample Exhibit for Student-Level Analysis Involving Plausible Values Taken From the TIMSS 1999 International Science Report

	Three or More Bookcases (More Than 200 Books)		About Two Bookcases (101-200 Books)		About One Bookcase (26-100 Books)		About One Shelf (11-25 Books)		None or Very Few (0-10 Books)	
	Percent of Students	Average Achievement	Percent of Students	Average Achievement	Percent of Students	Average Achievement	Percent of Students	Average Achievement	Percent of Students	Average Achievement
Australia	38 (1.2)	565 (4.7)	27 (0.9)	544 (4.8)	24 (0.9)	521 (6.0)	7 (0.6)	495 (9.2)	3 (0.4)	483 (15.0)
Belgium (Flemish)	14 (0.8)	561 (4.9)	14 (0.6)	557 (5.7)	31 (1.3)	542 (4.2)	21 (0.7)	522 (5.8)	19 (1.3)	504 (5.4)
Bulgaria	37 (2.2)	548 (7.6)	19 (0.9)	529 (5.6)	24 (1.1)	509 (5.2)	12 (1.0)	477 (8.1)	9 (1.3)	450 (9.1)
Canada	31 (0.9)	553 (4.0)	24 (0.8)	541 (3.3)	28 (0.7)	527 (3.4)	11 (0.5)	498 (5.1)	5 (0.3)	495 (12.4)
Chile	9 (0.6)	468 (10.4)	11 (0.6)	457 (8.1)	28 (0.9)	441 (5.3)	32 (0.9)	403 (4.0)	20 (1.1)	378 (4.7)
Chinese Taipei	16 (0.8)	616 (6.1)	12 (0.5)	603 (7.3)	31 (0.7)	579 (6.0)	23 (0.7)	554 (4.7)	17 (0.9)	507 (4.4)
Cyprus	17 (0.8)	479 (4.5)	23 (0.9)	476 (3.7)	36 (1.0)	465 (4.0)	19 (0.9)	437 (4.9)	5 (0.5)	402 (10.0)
Czech Republic	28 (1.4)	565 (5.5)	30 (1.4)	548 (5.8)	34 (1.1)	523 (4.8)	7 (0.8)	493 (7.6)	1 (0.2)	~ ~
England	26 (1.2)	593 (6.7)	23 (0.8)	550 (7.3)	32 (1.1)	526 (4.7)	13 (0.8)	483 (6.4)	7 (0.7)	472 (11.5)
Finland	22 (1.1)	552 (5.7)	22 (0.9)	544 (4.7)	39 (1.1)	536 (4.4)	14 (0.8)	506 (5.3)	4 (0.4)	495 (11.6)
Hong Kong, SAR	8 (0.5)	548 (6.4)	10 (0.5)	534 (6.0)	27 (0.7)	537 (4.9)	27 (0.7)	530 (4.8)	28 (0.9)	517 (4.2)
Hungary	38 (1.4)	581 (3.7)	21 (0.8)	561 (5.6)	25 (1.0)	539 (4.4)	12 (0.8)	502 (8.6)	3 (0.5)	460 (14.4)
Indonesia	5 (0.5)	437 (17.8)	5 (0.4)	451 (14.1)	26 (0.9)	446 (4.6)	39 (1.0)	429 (5.5)	26 (1.3)	438 (5.5)
Iran, Islamic Rep.	9 (0.7)	482 (8.3)	8 (0.7)	479 (8.8)	22 (0.6)	465 (6.1)	32 (0.8)	443 (4.7)	29 (1.4)	425 (4.9)
Israel	22 (1.3)	498 (5.3)	21 (0.8)	504 (5.7)	33 (1.0)	471 (6.1)	18 (0.9)	434 (7.0)	6 (0.5)	388 (15.8)
Italy	20 (0.9)	523 (7.5)	15 (0.7)	518 (5.9)	28 (0.9)	497 (4.4)	25 (0.9)	471 (5.6)	12 (0.8)	453 (7.2)
Japan	18 (0.7)	577 (5.3)	18 (0.6)	567 (5.0)	31 (0.7)	548 (2.6)	19 (0.6)	541 (4.6)	14 (0.6)	518 (5.1)
Jordan	10 (0.6)	485 (6.7)	10 (0.6)	493 (6.5)	28 (0.8)	471 (4.8)	31 (1.1)	447 (4.3)	21 (1.2)	416 (5.8)
Korea, Rep. of	20 (0.8)	589 (3.8)	23 (0.6)	562 (4.6)	36 (0.7)	544 (2.1)	10 (0.5)	510 (4.9)	10 (0.4)	490 (5.6)
Latvia (LSS)	47 (1.4)	522 (6.2)	25 (0.9)	499 (4.5)	21 (0.9)	480 (4.6)	6 (0.7)	465 (10.6)	1 (0.2)	~ ~
Lithuania ‡	17 (1.0)	528 (7.1)	21 (1.0)	517 (6.2)	36 (1.2)	482 (5.2)	20 (1.1)	454 (6.1)	7 (0.8)	433 (10.1)
Macedonia, Rep. of	7 (0.7)	492 (8.0)	9 (0.6)	498 (5.9)	30 (1.2)	487 (5.4)	38 (1.3)	449 (6.6)	15 (1.3)	408 (9.8)
Malaysia	9 (0.7)	538 (8.1)	12 (0.6)	521 (8.3)	32 (0.9)	505 (6.3)	34 (0.9)	472 (4.7)	13 (0.8)	460 (6.6)
Moldova	9 (0.7)	486 (7.8)	11 (0.8)	489 (7.5)	28 (1.0)	470 (4.9)	33 (1.2)	456 (5.1)	20 (1.1)	427 (7.4)
Morocco	3 (0.3)	343 (13.9)	5 (0.4)	372 (20.0)	20 (1.0)	335 (11.4)	35 (0.9)	323 (5.9)	37 (1.7)	320 (5.7)
Netherlands	24 (1.8)	575 (9.3)	23 (1.2)	554 (6.9)	31 (1.1)	546 (7.2)	15 (1.4)	508 (12.1)	8 (1.4)	499 (12.3)
New Zealand	33 (1.1)	541 (5.7)	24 (0.8)	529 (5.3)	27 (0.8)	499 (5.3)	10 (0.7)	460 (8.2)	6 (0.5)	407 (12.1)
Philippines	5 (0.4)	355 (16.5)	5 (0.4)	388 (17.3)	15 (0.8)	393 (11.9)	33 (0.9)	350 (8.0)	41 (1.4)	323 (7.0)
Romania	15 (1.3)	515 (8.1)	15 (0.9)	510 (8.3)	32 (1.1)	474 (6.0)	24 (1.6)	447 (6.8)	14 (1.1)	434 (9.3)
Russian Federation	23 (1.5)	555 (6.0)	29 (1.1)	541 (7.1)	31 (1.3)	521 (7.7)	13 (1.0)	495 (8.9)	4 (0.5)	470 (20.8)
Singapore	12 (0.6)	599 (11.4)	14 (0.7)	599 (10.3)	40 (1.1)	579 (7.2)	22 (1.0)	540 (8.8)	12 (0.8)	516 (8.8)
Slovak Republic	17 (0.9)	566 (6.6)	24 (1.0)	550 (5.6)	43 (1.1)	531 (3.9)	14 (1.0)	494 (6.4)	2 (0.4)	~ ~
Slovenia	14 (0.9)	564 (5.6)	20 (0.8)	556 (5.6)	46 (1.0)	532 (3.3)	16 (0.8)	498 (6.4)	4 (0.4)	483 (11.1)
South Africa	7 (0.5)	293 (25.6)	6 (0.5)	337 (18.9)	14 (0.8)	308 (15.3)	31 (0.8)	245 (8.7)	43 (1.6)	208 (5.9)
Thailand	6 (0.5)	509 (7.0)	8 (0.5)	502 (6.6)	27 (0.9)	493 (4.5)	37 (1.1)	477 (4.1)	22 (1.0)	465 (5.4)
Tunisia	9 (0.6)	438 (7.2)	9 (0.6)	444 (6.4)	25 (0.7)	438 (4.7)	36 (0.9)	424 (5.2)	21 (1.1)	422 (3.4)
Turkey	6 (0.4)	441 (11.1)	8 (0.5)	459 (7.4)	28 (0.8)	449 (5.3)	37 (0.8)	426 (5.0)	22 (1.1)	415 (5.2)
United States	28 (1.2)	557 (4.5)	22 (0.6)	538 (4.6)	29 (0.8)	508 (3.9)	14 (0.7)	468 (7.0)	8 (0.6)	442 (6.0)
International Avg.	18 (0.2)	517 (1.6)	16 (0.1)	511 (1.2)	29 (0.2)	493 (1.0)	22 (0.1)	464 (1.0)	14 (0.2)	441 (1.5)

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1998-1999.

Background data provided by students.

‡ Lithuania tested the same cohort of students as other countries, but later in 1999, at the beginning of the next school year.

( ) Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

A tilde (~) indicates insufficient data to report achievement.

To replicate the results in this exhibit, we need to undertake several steps. After reviewing the codebooks and the questionnaire information we find out if the students were asked a question about the number of books in the home (see Supplement 1 for a copy of the student questionnaires). The data collected from this variable are captured in the variable BSBGBOOK, and this variable is found in the Student Background data file. 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 changes were made we can continue with our analysis without any modifications.

We then proceed to read, from the SPSS system file that contains this variable, the international science achievement plausible values (BSSSCI01-BSSSCI05), the sampling weight for the students (TOTWGT), the variables that contain the jackknife replication information (JKZONE and JKREP), the variable that will be used to select the eighth graders from the data file (IDGRADER), and the variable containing the country identification code (IDCNTRY). In this analysis we will use the data for all countries in the database, although the exact same steps need to be taken if you want to examine these variables within a single country or for a select group of countries.

The SPSS code is presented in Exhibit 9.11. Selections of the results obtained from this program are displayed in Exhibit 9.12. We have included as part of the program the corresponding value labels and format statements so that the different categories or groups are labeled appropriately.

Notice that one of the steps in each of this program is to select only those students in the eighth grade who chose one of the five options presented in the question.

Note also in this analysis that we have used the data from all the countries, although we did select from the system file only those variables that are relevant to our analysis. In general, this type of analysis is quite feasible with a powerful desktop computer; however, you need to keep in mind that computing and storage requirements for these types of analysis are quite demanding.

In general, to perform analyses such as those using the Student Background data files, you need to do the following:

- Identify the variable or variables of interest in the student file and find out about any specific national adaptations to the variable.
- Retrieve the relevant variables from the data files, including the achievement score, sampling weights, JRR replication information, and any other variables used in the selection of cases.
- Use the macro JACKPV with the corresponding arguments and parameters.
- Print out the result file.

Exhibit 9.11 SPSS Control Statements for Performing Analyses with Student-Level Variables Involving Plausible Values (EXAMPLE2.SPS)

```
get file = "d:\timss\data\bm2\bsgallm2.sav"
   / keep=idcntry idstud idgrader jkrep jkzone totwgt
          bsbgbook bsssci01 to bsssci05 intms99.

select if not(missing(bsbgbook)) and intms99=1 and idgrader = 2.

value labels
  bsbgbook 1  "None or very few (0-10)"
           2  "For 1 shelf (11-25)"
           3  "For 1 bookcase (26-100)"
           4  "For 2 bookcases (101-200)"
           5  "For 3 or more bookcases (>200)" /
  idcntry
   036 'AUSTRALIA         '   956 'BELGIUM (Flemish) '   100 'BULGARIA          '
   124 'CANADA            '   152 'CHILE             '   196 'CYPRUS            '
   203 'CZECH REP.        '   926 'ENGLAND           '   246 'FINLAND           '
   344 'HONG KONG         '   348 'HUNGARY           '   360 'INDONESIA         '
   364 'IRAN, ISLAMIC REP.'   376 'ISRAEL            '   380 'ITALY             '
   400 'JORDAN            '   392 'JAPAN             '   410 'KOREA, REP. OF    '
   440 'LITHUANIA         '   428 'LATVIA            '   504 'MOROCCO           '
   498 'MOLDOVA, REP. OF  '   807 'MACEDONIA, REP. OF'   458 'MALAYSIA          '
   528 'NETHERLANDS       '   554 'NEW ZEALAND       '   608 'PHILIPPINES       '
   642 'ROMANIA           '   643 'RUSSIAN FEDERATION'   702 'SINGAPORE         '
   703 'SLOVAK REP.       '   705 'SLOVENIA          '   764 'THAILAND          '
   788 'TUNISIA           '   792 'TURKEY            '   158 'CHINESE TAIPEI    '
   840 'UNITED STATES     '   710 'SOUTH AFRICA      ' .

save outfile = student.

* Now use the macro to get the results.
include "d:\timss\programs\jackpv.sps".

jackpv infile= student  /
       cvar  = idcntry bsbgbook /
       pvs   = BSSSCI01 TO BSSSCI05 /
       npv   = 5        /
       jkz   = JKZONE   /
       jkr   = JKREP    /
       njkz  = 75       /
       WGT   = TOTWGT   .

sort cases by idcntry.

print formats n (f6.0) totwgt (f10.0) mnpv mnpv_se pct pct_se (f6.2).

report format=list automatic
  / var   = bsbgbook (label) n totwgt mnpv mnpv_se pct pct_se
  / break = idcntry .
```

Exhibit 9.12 Extract of SPSS Computer Output for Performing Analyses with Student-Level Variables Involving Plausible Values (EXAMPLE 2)

BSBGBOOK	N	TOTWGT	MNPV	MNPV_SE	PCT	PCT_SE
AUSTRALIA						
None or very few (0-10)	130	8731	482.87	15.00	3.47	.40
For 1 shelf (11-25)	266	18361	495.35	9.24	7.29	.61
For 1 bookcase (26-100)	927	61382	521.47	5.98	24.36	.95
For 2 bookcases (101-200)	1040	67799	544.46	4.81	26.91	.86
For 3 or more bookcases (>200)	1545	95691	565.26	4.74	37.98	1.22
BULGARIA						
None or very few (0-10)	279	7627	449.82	9.07	8.71	1.27
For 1 shelf (11-25)	371	10178	476.67	8.05	11.62	1.02
For 1 bookcase (26-100)	772	20888	509.11	5.25	23.84	1.12
For 2 bookcases (101-200)	636	16817	529.05	5.56	19.19	.87
For 3 or more bookcases (>200)	1184	32107	548.34	7.59	36.65	2.20
CANADA						
None or very few (0-10)	446	17772	495.07	12.42	4.84	.32
For 1 shelf (11-25)	1044	42011	497.64	5.10	11.45	.50
For 1 bookcase (26-100)	2537	103001	526.87	3.37	28.07	.71
For 2 bookcases (101-200)	2121	89829	540.97	3.34	24.48	.77
For 3 or more bookcases (>200)	2526	114301	552.68	4.04	31.15	.89
.						
.						
.						
.						
.						
UNITED STATES						
None or very few (0-10)	774	263683	441.63	6.01	8.11	.59
For 1 shelf (11-25)	1287	448473	467.99	7.05	13.79	.71
For 1 bookcase (26-100)	2502	927299	508.02	3.92	28.51	.83
For 2 bookcases (101-200)	1877	700350	538.25	4.58	21.54	.56
For 3 or more bookcases (>200)	2344	912338	556.69	4.53	28.05	1.21
ENGLAND						
None or very few (0-10)	182	34823	472.16	11.54	6.60	.71
For 1 shelf (11-25)	350	66603	483.40	6.43	12.63	.79
For 1 bookcase (26-100)	900	167841	525.71	4.69	31.82	1.07
For 2 bookcases (101-200)	641	120040	549.55	7.25	22.76	.82
For 3 or more bookcases (>200)	760	138215	593.14	6.72	26.20	1.20
BELGIUM (Flemish)						
None or very few (0-10)	813	12339	504.02	5.42	19.14	1.27
For 1 shelf (11-25)	1096	13823	522.35	5.78	21.44	.73
For 1 bookcase (26-100)	1676	20042	542.45	4.15	31.09	1.30
For 2 bookcases (101-200)	813	9063	557.00	5.65	14.06	.62
For 3 or more bookcases (>200)	798	9196	560.75	4.94	14.26	.79

In this example each country's mean plausible value for science achievement is reported for each response category in the variable BSBGBOOK. The results are presented by country for each value of the variable IDCNTRY after selecting only those cases with IDGRADER=2. The country and the five response options are presented in the first column. The second column has the number of students sampled in each category. The third column corresponds to the weights given these sampled students in the whole population, followed by their mean plausible values for science achievement and the corresponding standard errors. The last two columns represent the percentages of students sampled responding within each category and the corresponding standard errors.

For example, from the first line of the report, we can say that in Australia the 1545 students (38 percent of the sampled students) who reported having more than 200 books had a mean science achievement of 565.26 with a standard error of 4.74; while the 130 students (3.5 percent of the sampled students) who reported having fewer than 10 books had a mean science achievement of 482.87 with a standard error of 15.00.

9.7 Performing Analyses with Teacher-Level Variables

When analyzing the teacher data, it is first necessary to link the students with their corresponding teachers. Each student record in the Student Background data file can contain a link to as many as six different teachers in the Teacher Background data file. To facilitate the linking between students and their teachers in the teacher file, the Student-Teacher Linkage file was created and is part of the International Database. This file is called BST<COUNTRY>M1 and BST<COUNTRY>M2 in 1995 and 1999, respectively. The Student-Teacher Linkage file contains one record for each student-by-teacher combination, with the corresponding identification variables.

Each record also contains the number of mathematics and science teachers for the student and a set of weights that can be used when conducting analyses with these data. Student achievement plausible values, sampling weights, and JRR replication information have been added to the Student-Teacher Linkage file in order to simplify the merging process for analyses that link teacher variables to student achievement. For such analyses it is necessary to merge only the Teacher Background file with the Student-Teacher Linkage file. For analyses linking teacher variables to other student variables, it is necessary also to merge the Student Background files with the Teacher Background file after it has been combined with the Student-Teacher Linkage file.

Conducting analyses with the teacher data requires some extra steps that are not required when analyzing the student or school background data.

For our example, we want to find out about the age of the science teachers who teach the eighth-grade students in each of the TIMSS countries. In particular, we want to find out what percentage of eighth-grade students are taught by teachers of a specific group based on the teachers' ages (BTBGAGE), and the mean achievement in science of the students taught by those teachers. These results are located in the Science Teacher Background Data Almanac (bsalm7_m2.*) for questionnaire item TQS2A-01 as shown in Exhibit 9.13.

Exhibit 9.13 Sample Data Almanac Sheet for Teacher-Level Analysis Using Plausible Values Taken From the TIMSS 1999 International Science Report

Third International Mathematics and Science Study - 1999 Assessment 16:43 Sunday, February 18, 2001 4
Science Teacher Background Data Almanac - 8th Grade

Question : Teacher age
Location : TQS2A-01 (BTBGAGE)

Country	Sample	Valid N	1.UNDER 25 Mean	2.25-29 Mean	3.30-39 Mean	4.40-49 Mean	5.50-59 Mean	6.60 OR MORE Mean	NOT ADMINIS TERED Mean	OMIT Mean
Australia	543	465	536.9	533.6	538.9	533.4	553.1	548.2	548.1	597.0
Belgium (Flemish)	409	372	537.7	540.8	526.5	534.2	536.4	477.1	522.5	540.8
Bulgaria	593	571	520.0	505.5	516.8	505.5	518.3	488.2	557.9	484.6
Canada	474	439	533.7	534.4	532.2	538.2	527.1	543.2	525.1	529.1
Chile	185	180	444.8	461.5	438.6	421.4	397.1	407.9	391.4	402.2
Chinese Taipei	150	147	579.2	567.2	568.7	572.8	570.4	529.6	590.7	.
Cyprus	252	222	.	424.2	452.4	465.4	463.2	.	445.0	490.8
Czech Republic	568	546	515.6	551.2	535.9	545.5	537.2	547.5	517.9	625.6
England	537	359	512.9	543.6	546.7	545.6	561.0	425.5	525.2	565.8
Finland	434	418	538.1	532.6	536.8	537.0	531.9	542.5	544.8	543.7
Hong Kong, SAR	137	135	481.5	538.1	537.2	522.5	513.1	546.7	501.6	.
Hungary	592	581	561.1	554.9	553.8	552.2	548.4	533.6	583.5	.
Indonesia	302	297	398.0	436.0	430.2	442.8	474.1	.	387.9	432.7
Iran, Islamic Rep.	170	168	427.3	435.8	452.6	456.3	447.1	.	.	480.5
Israel	259	235	428.1	470.8	460.9	474.3	485.0	505.7	466.7	481.2
Italy	183	183	.	.	493.9	493.9	491.7	517.6	.	.
Japan	144	144	562.0	540.4	549.2	545.3	556.1	586.9	.	.
Jordan	147	146	413.9	445.8	454.1	465.9	434.4	.	449.6	.
Korea, Rep. of	186	186	528.3	556.4	547.7	552.0	541.4	539.8	.	.
Latvia (LSS)	433	402	507.2	493.4	502.7	502.8	507.3	508.6	487.6	.
Lithuania	450	442	474.2	484.2	488.0	494.1	480.1	497.4	489.2	.
Macedonia, Rep. of	596	593	468.6	461.3	448.8	442.0	469.8	479.9	481.0	478.5
Malaysia	150	150	475.9	472.8	494.9	509.0	505.0	.	.	.
Moldova	600	577	466.9	463.9	465.4	452.8	457.8	476.8	442.4	424.0
Morocco	687	606	326.0	317.4	325.3	321.3	.	.	324.7	306.0
Netherlands	351	305	535.3	546.1	544.7	550.5	538.1	625.7	540.4	.
New Zealand	250	234	492.6	500.4	503.3	525.2	502.3	546.8	472.9	489.2
Philippines	150	149	395.9	341.4	335.2	352.4	322.5	424.0	394.3	.
Romania	588	564	438.1	456.8	470.0	470.7	481.3	488.1	485.3	479.1
Russian Federation	756	755	513.8	532.8	530.4	526.8	534.4	524.1	.	563.2
Singapore	145	144	604.9	576.6	572.5	559.2	559.8	494.6	.	501.4
Slovak Republic	578	179	511.1	511.5	547.2	531.5	537.7	534.6	535.7	.
Slovenia	447	216	495.5	519.3	537.9	533.3	534.6	512.3	533.1	.
South Africa	190	185	274.4	259.3	220.8	292.3	445.7	177.8	207.0	.
Thailand	150	148	472.0	471.2	478.0	485.7	507.6	517.8	503.7	.
Tunisia	149	148	424.4	416.9	430.1	445.3	430.3	.	402.9	.
Turkey	204	203	414.2	417.9	424.3	448.8	439.0	.	.	400.4
United States	1079	904	536.3	528.6	501.2	516.8	524.3	532.1	491.7	485.8
International Avg.	374	334	481.9	485.0	486.7	491.2	499.0	502.7	478.3	490.5

The TIMSS teacher files do not contain representative samples of teachers within a country. Rather, these are the teachers for a representative sample of students within a country. Therefore, it is appropriate that statements about the teachers be made only in terms of how many students are taught by teachers of one kind or another, and not in terms of how many teachers in the country do one thing or another.

As before, we first proceed to identify the variables relevant to the analysis in the corresponding files, and review the documentation on the specific national adaptations to the questions of interest (Supplements 1 and 2). Since we are using teacher-level variables we need to look into the teacher file and the Student-Teacher Linkage files to find the variables. From the science teacher file we extract the variable that contains the information on the science teachers' age (BTBGAGE), the variable that identifies the country (IDCNTRY), and the two variables that will allow us to link the teacher information to the student data (IDTEACH and IDLINK).

There is one teacher file for the mathematics teachers and a second teacher file for the science teachers. If you want to look only at mathematics teachers, then you will need to use the mathematics teacher file (BTM<COUNTRY>M2); if the interest is in the science teachers then you will need to use the science teacher file (BTS<COUNTRY>M2); but if the interest is in the mathematics and science teachers combined, both these files need to be combined by appending or adding one file to the other. In doing so it is important to keep in mind that although there are variables in common between these two files, most of them are not.

In our example, our teacher variable of interest (age) is a categorical variable with six categories. However, we want to categorize the teachers into four groups (under 29 years old, 30-39 years old, 40-49 years old, and 50 years or over). While reading the teacher file we use commands in SPSS to collapse the different values into four categories and we label them accordingly. We then proceed to read the necessary information from the Student-Teacher Linkage file. From this file we keep the country identification (IDCNTRY) and the two variables that will allow us to link the student information to the teacher data (IDTEACH and IDLINK). We also keep the variable that indicates the grade for the student (IDGRADER), the science achievement plausible values (BSSSCI01-BSSSCI05), and the information necessary to compute the replicate weights for estimating the JRR standard error. We select the variable that will give us the correct weight for science teacher variable (SCIWGT). If you are interested in looking at the mathematics teachers, then the weight variable that should be selected is MATWGT; if you are interested in analyzing both mathematics and science teachers combined, the weight variable TCHWGT should be selected.

The two files are then merged or matched into one file that will then be used with the JACKPV macro. These two files will be merged using the variables IDCNTRY, IDTEACH, and IDLINK. The combination of values for these three variables is unique within the teacher data, but is repeated in the Student-Teacher Linkage file as many times as the specific teacher teaches students in a class. After the files are merged, the macro JACKPV is used and the results can be printed. The code in SPSS for this example is presented in Exhibits 9.14. Selections of the results obtained from this program are displayed in Exhibit 9.15.

Exhibit 9.14 SPSS Control Statement for Performing Teacher Level Analysis Using Plausible Values

```
* Get the teacher data, sort and save.
get file = "d:\timss\data\bm2\btsallm2.sav"
   / keep = idcntry idteach idlink btbgage.

recode btbgage (1,2 = 1) (  3 = 2) (  4 = 3) (5,6 = 4)
               (else = sysmis) into btdgage.

sort cases by idcntry idteach idlink.

save outfile = teacher.

* Get the student teacher link data and save.
get file = "d:\timss\data\bm2\bstallm2.sav"
   / keep = idcntry idteach idlink idgrader jkrep jkzone sciwgt
            bsssci01 to bsssci05 intms99.

select if sciwgt > 0 and idgrader=2.

sort cases by idcntry idteach idlink.

save outfile= studteac.

* Now merge the two files.
match files
  / file=studteac
  / table=teacher
  / by idcntry idteach idlink.

select if not(missing(btdgage)) and intms99=1.

* Define the format for the variables used.
value labels
 btdgage   1 "Under 29 y.o."
           2 "30-39 years old"
           3 "40-49 years old"
           4 "50 years or older"/
 idcntry
 036 'AUSTRALIA         '   956 'BELGIUM (Flemish) '   100 'BULGARIA          '
 124 'CANADA            '   152 'CHILE             '   196 'CYPRUS            '
 203 'CZECH REP.        '   926 'ENGLAND           '   246 'FINLAND           '
 344 'HONG KONG         '   348 'HUNGARY           '   360 'INDONESIA         '
 364 'IRAN, ISLAMIC REP.'   376 'ISRAEL            '   380 'ITALY             '
 400 'JORDAN            '   392 'JAPAN             '   410 'KOREA, REP. OF    '
 440 'LITHUANIA         '   428 'LATVIA            '   504 'MOROCCO           '
 498 'MOLDOVA, REP. OF  '   807 'MACEDONIA, REP. OF'   458 'MALAYSIA          '
 528 'NETHERLANDS       '   554 'NEW ZEALAND       '   608 'PHILIPPINES       '
 642 'ROMANIA           '   643 'RUSSIAN FEDERATION'   702 'SINGAPORE         '
 703 'SLOVAK REP.       '   705 'SLOVENIA          '   764 'THAILAND          '
 788 'TUNISIA           '   792 'TURKEY            '   158 'CHINESE TAIPEI    '
 840 'UNITED STATES     '   710 'SOUTH AFRICA      ' .

save outfile = merged.

* Now use the macro to get the results.

include "d:\timss\programs\jackpv.sps".

jackpv infile= merged /
       cvar  = idcntry btdgage /
       pvs   = BSSSCI01 TO BSSSCI05 /
       npv   = 5        /
       jkz   = JKZONE   /
       jkr   = JKREP    /
       njkz  = 75       /
       WGT   = sciwgt   / .

sort cases by idcntry.

print formats n (f6.0) sciwgt (f10.0) mnpv mnpv_se pct pct_se (F6.2).

report format=list automatic
 / var = btdgage (label) n sciwgt mnpv mnpv_se pct pct_se
 / break = idcntry.
```

Exhibit 9.15 Extract of SPSS Output for Performing Teacher Level Analysis Using Plausible Values (EXAMPLE 3)

```
BTDGAGE                N        SCIWGT       MNPV     MNPV_SE      PCT      PCT_SE
_________________   ______   __________   ______   _______    ______    ______

AUSTRALIA
Under 29 y.o.          627        36628   534.66       8.17     15.75       2.66
30-39 years old       1091        72944   538.88       6.56     31.37       3.39
40-49 years old       1287        79488   533.43       7.07     34.19       3.31
50 years or older      647        43444   552.88       9.95     18.69       2.71

BULGARIA
Under 29 y.o.         1197        10070   507.05       8.11     13.24       3.04
30-39 years old       3139        20554   516.75       9.12     27.03       2.80
40-49 years old       3643        24722   505.54       5.06     32.51       2.78
50 years or older     3182        20696   516.18       8.11     27.22       2.83

CANADA
Under 29 y.o.         1739        72340   534.36       6.10     20.95       3.06
30-39 years old       2754       105991   532.24       3.90     30.69       2.55
40-49 years old       2251       106509   538.22       3.70     30.84       2.95
50 years or older     1496        60506   527.21       6.69     17.52       2.59

.
.
.
.
.

UNITED STATES
Under 29 y.o.         1431       563546   530.29       8.06     20.03       2.60
30-39 years old       1484       527177   501.20      10.56     18.74       2.18
40-49 years old       2419       816588   516.78       8.31     29.03       2.82
50 years or older     2340       905665   525.27       6.76     32.20       2.68

ENGLAND
Under 29 y.o.          628        78473   536.69       9.92     23.79       4.02
30-39 years old        687        75922   546.72       9.66     23.02       3.56
40-49 years old       1101       102994   545.57       9.96     31.22       3.97
50 years or older      734        72492   559.60      13.22     21.98       3.36

BELGIUM (Flemish)
Under 29 y.o.         1962        14757   539.63       6.39     24.66       2.77
30-39 years old       1503        14267   526.51       9.64     23.84       3.09
40-49 years old       2347        20577   534.16       5.89     34.39       3.47
50 years or older     1144        10242   535.20      11.35     17.11       2.53
_
```

In this example the variable BTDAGE is created by collapsing the categorical variable BTBGAGE into only four categories, excluding those teachers missing this information. The results are then presented by country for each value of the variable IDCNTRY after selecting only those cases with IDGRADER=2. The country and the four age categories are presented in the first column. The second column has the number of students sampled in each category. The third column corresponds to the weight given these sampled students in the whole population, followed by the mean plausible value for science achievement and its corresponding standard error. The last two columns represent the percentage of students sampled within each category and its corresponding standard error.

For example, we can say that in Australia the 627 students (15.75 percent of the sampled students) with a teacher under 29 years old had a mean science achievement of 534.66 with a standard error of 8.17, while the 647 students (18.69 percent of the sampled students) with a teacher 50 years or older had a mean science achievement of 552.88 with a standard error of 9.95.

In summary, to perform analyses such as those using the Student and Teacher Background data files you need to do the following:

- Identify the variable or variables of interest in the corresponding teacher file and find out about any specific national adaptations to the variable.
- Retrieve the relevant variable or variables from the corresponding teacher data files. If you are interested in looking at both mathematics and science teachers combined, then the files for these teachers need to be added or appended to each other.
- Retrieve the relevant variables from the Student-Teacher Linkage file. This includes the identification information for the country and teacher (IDCNTRY, IDTEACH, and IDLINK), the achievement score, JRR replication information, and the sampling weight. If the analysis is to be based on mathematics teachers only, then the weight variable to use is MATWGT. If the analysis is to be based on the science teachers only, then the weight variable to be used is SCIWGT. If the analysis is to be based on the science and mathematics teachers combined, then the weight variable to be used is TCHWGT.
- Merge the variables from the teacher data files into the Student-Teacher Linkage files using the variables IDCNTRY, IDTEACH and IDLINK.
- Use the macro JACKPV or JACKGEN with the corresponding arguments and parameters.
- Print out the result file.

9.8 Performing Analyses with School-Level Variables

Although the students in the TIMSS samples were selected from within a sample of schools, the school sample was designed to optimize the resulting sample of students, rather than to give an optimal sample of schools.

For this reason, it is always preferable to analyze school-level variables as attributes of students, rather than as elements in their own right. Although the school samples were not optimized, it is still possible to compute weighted numbers of schools with particular characteristics for providing reasonable estimates of percentages and averages across primary or middle schools in each country. The following example, however, describe only analyses based on student-weighted data.

For student-weighted analyses, the school-level data are analyzed to make statements about the number of students attending schools with one characteristic or another rather than the number of schools with certain characteristics. When school-level variables are analyzed, we recommend that you merge the selected school-level variables with the student-level file, and then use the sampling and weight information contained in the student-level file to make the desired statements. The examples presented in this section describe how this can be accomplished using SPSS.

Let us say that we want to find out the percentage of eighth graders who attend schools located in a certain geographical area of the country (BCBGCOMM), and their average achievement in science. These results are located in the School Background Data Almanac by Science Achievement (bsalm4_m2.*) for questionnaire item SCQ2-01 as shown in Exhibit 9.16.

Exhibit 9.16 Sample Data Almanac Sheet for School-Level Analysis Using Plausible Values Taken From the TIMSS 1999 International Science Report

Third International Mathematics and Science Study - 1999 Assessment 16:43 Sunday, February 18, 2001 1
School Background Data Almanac by Science Achievement - 8th Grade

Question : In what type of community is your school located?
Location : SCQ2-01 (BCBGCOMM)

Country	Sample	Valid N	RAPHICA LLY ISOLATE D AREA %	AGE OR RURAL (FARM) AREA %	KIRTS OF TOWN OR CITY %	4.CLOS CENTER OF TOWN OR CIT %	ADMINIS TERED %	%	RAPHICA ISOLATE Mean	AGE OR (FARM) Mean	KIRTS TOWN Mean	CENTER TOWN Mean	ADMINIS Mean	Mean
Australia	169	142	2.0	8.5	33.3	56.3	10.2	5.2	490.2	540.4	531.7	545.7	539.6	554.2
Belgium (Flemish)	135	122	.	32.0	23.5	44.5	6.3	5.7	.	535.6	513.0	549.1	546.8	499.4
Bulgaria	163	160	0.7	24.1	8.6	66.7	1.5	0.0	527.2	491.9	520.7	527.7	487.9	.
Canada	385	364	3.9	15.2	18.2	62.8	2.7	2.7	529.4	526.9	538.3	533.8	520.7	534.8
Chile	185	182	3.0	16.4	26.9	53.8	0.6	1.1	405.9	382.3	400.0	443.1	404.6	420.4
Chinese Taipei	150	150	2.0	10.0	38.8	49.1	0.0	0.0	539.5	533.3	570.3	576.7	.	.
Cyprus	61	55	0.7	26.6	29.5	43.2	9.5	0.0	491.2	432.8	470.1	471.8	452.6	.
Czech Republic	142	139	.	20.9	41.5	37.6	1.7	0.2	.	532.7	538.4	545.6	504.9	606.3
England	128	107	0.9	14.2	40.2	44.7	17.6	0.0	510.7	590.8	531.2	537.0	521.4	.
Finland	159	157	1.4	35.8	30.0	32.9	0.8	1.3	545.9	536.1	531.0	537.3	561.3	527.3
Hong Kong, SAR	137	128	1.7	2.5	26.4	69.3	6.1	0.7	511.5	523.8	529.3	533.0	508.1	471.4
Hungary	147	142	0.6	29.8	19.2	50.4	1.1	2.0	512.2	520.2	562.0	567.1	558.4	575.3
Indonesia	150	146	1.5	40.3	29.1	29.0	0.0	2.4	366.2	418.2	430.6	467.4	.	443.2
Iran, Islamic Rep.	170	169	1.0	24.3	38.6	36.1	0.0	0.5	402.5	423.7	452.4	461.2	.	426.4
Israel	139	123	7.2	28.5	17.8	46.4	8.6	1.9	496.8	451.8	442.3	477.5	487.2	511.1
Italy	180	178	35.5	4.0	25.9	34.7	0.0	1.0	499.2	512.2	475.9	498.3	.	483.1
Japan	140	139	2.1	15.6	47.1	35.2	0.0	0.7	565.5	546.8	552.1	546.9	.	540.9
Jordan	147	146	1.8	26.3	26.6	45.3	0.0	0.8	463.5	422.9	451.1	464.7	.	484.4
Korea, Rep. of	150	149	.	11.8	35.7	52.5	0.0	0.6	.	524.2	544.3	557.0	.	557.4
Latvia (LSS)	145	128	13.3	23.4	28.3	35.0	8.8	1.4	492.3	487.1	507.3	517.5	494.1	463.3
Lithuania	150	149	2.4	27.9	8.6	61.1	0.6	0.0	442.1	476.1	496.5	494.0	511.7	.
Macedonia, Rep. of	145	143	1.6	31.6	19.5	47.3	0.0	1.3	428.1	419.2	461.2	484.1	.	447.9
Malaysia	150	150	4.7	37.6	16.6	41.2	0.0	0.0	470.0	473.9	482.8	515.8	.	.
Moldova	150	149	2.1	59.6	4.2	34.1	0.4	0.0	455.4	453.9	484.6	466.2	374.4	.
Morocco	173	168	0.7	19.1	15.7	64.5	2.6	0.0	337.2	302.6	318.0	329.5	329.6	.
Netherlands	126	98	.	27.5	44.6	27.9	16.8	0.6	.	544.0	552.0	537.3	540.2	542.4
New Zealand	152	139	4.4	12.6	25.0	58.1	6.8	1.9	512.6	494.1	513.8	509.5	522.7	504.0
Philippines	150	149	4.2	17.2	12.9	65.8	0.0	0.7	317.6	322.1	348.0	353.5	.	239.6
Romania	147	142	1.4	37.6	18.4	42.6	3.6	0.0	506.5	448.2	466.4	493.7	475.0	.
Russian Federation	189	189	0.2	35.8	.	64.0	0.0	0.0	564.4	504.8	.	542.8	.	.
Singapore	145	145	.	.	.	100.0	0.0	0.0	.	.	.	567.9	.	.
Slovak Republic	145	138	.	34.0	37.4	28.6	2.7	1.5	.	516.9	538.8	543.9	582.6	587.9
Slovenia	149	145	2.6	35.5	22.5	39.5	2.2	0.0	546.7	532.6	533.6	532.6	536.8	.
South Africa	194	186	2.7	47.2	20.0	30.0	2.1	1.3	182.6	190.1	257.5	323.7	196.7	250.9
Thailand	150	149	17.3	35.2	23.0	24.4	0.7	0.0	459.3	473.1	479.4	515.3	462.2	.
Tunisia	149	144	7.2	18.0	30.7	44.1	1.3	1.7	419.7	405.7	430.8	440.5	402.8	435.9
Turkey	204	196	2.8	3.1	15.3	78.8	0.9	2.3	452.1	422.2	424.5	434.1	376.4	463.6
United States	221	182	4.7	19.2	25.4	50.7	18.0	0.3	525.2	505.2	537.7	509.6	507.2	499.1
International Avg.	160	152	4.3	24.6	25.7	48.1	3.5	1.0	467.8	470.8	483.8	498.7	477.2	482.8

As in the previous example, the first step in our analysis is to locate the variables of interest in the specific codebook and file. We find the variable BCBGCOMM in the School Background file, and the student weights and plausible values in the Student Background file. We then proceed to review the documentation on national adaptations and discover that Australia has modified this variable slightly to fit their particular context. At this time we could proceed in one of two ways: we could exclude Australia from our analysis or we could label the variable accordingly so that we will not be making incorrect inferences about the specific groups. In the latter case, since we want to also explore the results for Australia we take the precaution of labeling the values for variable BCBGCOMM in a generic way before we proceed with the analysis.

After these considerations, we then proceed to read the School Background file and keep only the variables that are relevant to our analysis. In this case we keep the country identification (IDCNTRY) and school identification (IDSCHOOL). We keep these variables because these are the variables that will allow us to merge the school data to the student data. We also keep from the School Background file the variable of interest, in this case BCBGCOMM. We then read the variables of interest from the student data file. First we read the identification of the country and the school (IDCNTRY and IDSCHOOL) which will allow us to merge the student data to the school data. We also select from this variable the international science achievement plausible values (BSSSCI01-BSSSCI05), the sampling weight for the student (TOTWGT), the variables that contain the jackknife replication information (JKZONE and JKREP), and the variable that will be used to select the eighth graders from the data file (IDGRADER).

We then proceed to merge the school information with the student information using the variables IDCNTRY and IDSCHOOL as merge variables, and then use the macro JACKPV to obtain the corresponding percentages of students within each group, and their mean achievement scores in science. The computer code used to run this analysis in SPSS can be found in Exhibit 9.17 and an extract of the results is shown in Exhibit 9.18.

Exhibit 9.17 SPSS Control Statements for Performing Student-Weighted Analyses with School-Level Variables (EXAMPLE4.SPS)

```
* Get the school data, sort and save it.
get file = "d:\timss\data\bm2\bcgallm2.sav"
   / keep = idcntry idschool bcbgcomm.

sort cases by idcntry idschool.

save outfile = school.

* Get the student data, sort and save it.
get file = "d:\timss\data\bm2\bsgallm2.sav"
  / keep = idcntry idschool idstud idgrader jkrep jkzone
           totwgt bsssci01 to bsssci05 .

select if idgrader = 2 .

sort cases by idcntry idschool.

save outfile = student.

* Now merge the two files.
match files
  / file=student
  / table=school
  / by idcntry idschool.

select if not(missing(bcbgcomm)).

execute.

* Define the format for the variables used.
value labels
 idgrader   1  "Seventh Grade   "  2  "Eighth Grade"      /
 bcbgcomm   1  "Community Type 1"  2  "Community Type 2"
            3  "Community Type 3"  4  "Community Type 4"  /
 idcntry
 036 'AUSTRALIA         '   956 'BELGIUM (Flemish) '   100 'BULGARIA          '
 124 'CANADA            '   152 'CHILE             '   196 'CYPRUS            '
 203 'CZECH REP.        '   926 'ENGLAND           '   246 'FINLAND           '
 344 'HONG KONG         '   348 'HUNGARY           '   360 'INDONESIA         '
 364 'IRAN, ISLAMIC REP.'   376 'ISRAEL            '   380 'ITALY             '
 400 'JORDAN            '   392 'JAPAN             '   410 'KOREA, REP. OF    '
 440 'LITHUANIA         '   428 'LATVIA            '   504 'MOROCCO           '
 498 'MOLDOVA, REP. OF  '   807 'MACEDONIA, REP. OF'   458 'MALAYSIA          '
 528 'NETHERLANDS       '   554 'NEW ZEALAND       '   608 'PHILIPPINES       '
 642 'ROMANIA           '   643 'RUSSIAN FEDERATION'   702 'SINGAPORE         '
 703 'SLOVAK REP.       '   705 'SLOVENIA          '   764 'THAILAND          '
 788 'TUNISIA           '   792 'TURKEY            '   158 'CHINESE TAIPEI    '
 840 'UNITED STATES     '   710 'SOUTH AFRICA      ' .

save outfile = merged.

* Now use the macro to get the results.
include "d:\timss\programs\jackpv.sps".

jackpv infile= merged /
       cvar  = idcntry bcbgcomm /
       pvs   = BSSSCI01 TO BSSSCI05 /
       npv   = 5         /
       jkz   = JKZONE    /
       jkr   = JKREP     /
       njkz  = 75        /
       WGT   = TOTWGT    / .

sort cases by idcntry bcbgcomm.

print formats n (f6.0) totwgt (f10.0) mnpv mnpv_se pct pct_se (F6.2).

report format=list automatic
 / var = bcbgcomm (label) n totwgt mnpv mnpv_se pct pct_se
 / break = idcntry .
```

Exhibit 9.18 Extract of SPSS Computer Output for Performing Student-Weighted Analyses with School-Level Variables(EXAMPLE 4)

BCBGCOMM	N	TOTWGT	MNPV	MNPV_SE	PCT	PCT_SE
AUSTRALIA						
Community Type 1	70	4369	490.23	4.25	1.99	1.16
Community Type 2	279	18620	540.38	17.94	8.46	2.57
Community Type 3	1143	73238	531.72	8.26	33.28	3.93
Community Type 4	1892	123818	545.68	7.66	56.27	4.15
BULGARIA						
Community Type 1	16	604	527.16	10.36	.69	.04
Community Type 2	716	20942	491.94	11.43	24.06	3.18
Community Type 3	314	7455	520.70	17.88	8.57	2.25
Community Type 4	2162	58025	527.69	6.30	66.67	4.36
CANADA						
Community Type 1	492	13556	529.41	9.37	3.86	1.00
Community Type 2	1527	53203	526.94	5.25	15.15	2.12
Community Type 3	1567	63860	538.34	7.02	18.19	2.50
Community Type 4	4684	220446	533.75	2.38	62.79	2.97
.						
.						
.						
.						
.						
UNITED STATES						
Community Type 1	277	128961	525.24	15.72	4.73	2.27
Community Type 2	1247	522877	505.19	7.92	19.18	2.49
Community Type 3	1651	693146	537.74	7.80	25.42	3.54
Community Type 4	4313	1381329	509.65	6.63	50.67	4.71
ENGLAND						
Community Type 1	21	3911	510.71	10.92	.86	.86
Community Type 2	327	64843	590.84	10.62	14.24	3.39
Community Type 3	985	182882	531.16	6.88	40.17	4.62
Community Type 4	1130	203619	537.02	8.59	44.73	4.47
BELGIUM (Flemish)						
Community Type 2	1528	18438	535.60	5.71	31.99	3.48
Community Type 3	932	13534	513.03	9.73	23.48	3.40
Community Type 4	2306	25659	549.12	5.33	44.52	3.96

In this example the variable BCBGCOMM is relabeled using generic labels due to national adaptations that could affect interpretation. The results are then presented by country for each value of the variable IDCNTRY after selecting only those cases with IDGRADER=2. The country and the four community types are presented in the first column. The second column has the number of students sampled in each category. The third column corresponds to the weight given these sampled students in the whole population, followed by the mean plausible value for science achievement and its corresponding standard error. The last two columns represent the percentage of students sampled within each category and its corresponding standard error.

For example, we can say that in Australia, 70 students who represent 4369 students in the population responded that they live in community type 1. These students had a mean science achievement of 490.23 with a standard error of 4.25. These students represent 1.99 percent of the sampled students and this percentage has a standard error of 1.16.

In summary, to perform analyses such as those using the Student and School Background files, you need to do the following:

- Identify the variable or variables of interest in the student file and find out about any specific national adaptations to the variable.
- Retrieve the relevant variables from the student files, including the achievement score, sampling weights, JRR replication information, and any other variables used in the selection of cases.
- Retrieve the relevant classification variable or variables from the school database.
- Merge the variables from the school database onto the student database using the variables IDCNTRY and IDSCHOOL.
- Use the macro JACKGEN or JACKPV with the corresponding arguments and parameters.
- Print out the result file.

10 Performing Analyses with the TIMSS Data Using SAS

10.1 Overview

This chapter presents some basic examples of analyses that can be performed with the TIMSS International Database using the sampling weights and scores discussed in previous chapters. It also provides details on some SAS progrmas to conduct such analyses, and the results of these analyses. The analyses presented here are simple in nature, andare designed primarily to familiarize you with the different files and their structure, as well as the relevant variables that need to be included in most analyses. The programs compute the percent of students in specified subgroups, the mean achievement on science in those groups, and the corresponding standard errors for the percent and mean statistics. Additionally some examples of regression statistics are presented. These analyses, based on student, teacher and school data, replicate analyses that are included in the TIMSS 1999 Science International Report (Martin et. al, 2000). You are invited to compare the results from these analyses to the tables in the reports, and are encouraged to practice analyzing the TIMSS data by trying to replicate the tables that are presented in the international reports.[1]

In our examples we use macros written for SAS that can be used to perform any of the analyses that are described below. These are general procedures that can be used for many purposes, provided you have some basic knowledge of the SAS macro language. If you have some programming experience in this statistical package, then you will be able to make the necessary modifications to the macros to obtain the desired results. Before using these macros, you should create a dataset in SAS that contains the variables necessary for the analysis. As part of this chapter we describe the control files included in the CD that can be used to do this.

10.2 Contents of the CDs

There are two CDs that accompany this User Guide – one CD containing the TIMSS 1999 data and one containing the TIMSS 1995 data. Each CD has the following internal file structure:

- A main directory identifying the year (TIMSS95 or TIMSS99).
- Within each main directory, there are six sub-directories.

DATA:	Contains data files in ASCII format
PROGRAMS:	Contains SPSS and SAS programs
CODEBOOK:	Contains codebook files
ALMANACS:	Contains data almanacs

▼ ▼ ▼

[1] Documentation regarding the computational methods used to obtain any derived variables included in the international reports is presented in Supplement 3.

TCMA:	Contains Test–Curriculum Matching Analysis Data
ITEMS:	Contains the Item Information files

The directory names within each CD and the file names generally follow the DOS 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 in the files contained in the CDs are indicated the following.

.SAS	SAS Control file or program
.SPS	SPSS Control file or program
.DAT	ASCII Data file
.LST	Almanac
.CDT	Codebook in Printout format
.CDF	Codebook in Machine readable format
.CSV	Test Curriculum Matching Analysis
.TXT	Item Information Tables

The DATA sub-directory contains the TIMSS data files in ASCII format. The data files that are in this directory are described in an earlier chapter of this guide. Each of these files has two corresponding control files in the PROGRAMS sub-directory. One of these reads the ASCII data and creates a SAS data set, the other reads the ASCII data and creates an SPSS system file. This chapter will focus on the files that can be used with SAS.

The following programs also can be found in this directory:

BSASCRM1.SAS and BSASCRM2.SAS

These files contain SAS programs that can be used to convert the response codes to the cognitive items to their corresponding correctness score levels. The use of these programs is described in this chapter.

JACKPV.SAS

This macro program in SAS can be used to compute weighted percentages of students within defined groups, and their mean achievement scores on an achievement scale using the available plausible values. This macro makes use of the plausible values provided for each subject in computing the mean achievement scores. This macro also generates replicate weights and computes the JRR sampling variances for the percentages of students within the group, and the JRR and imputation variances for the mean achievement scores. This macro should only be used when multiple plausible values are used in the analysis.

JACKGEN.SAS

This macro program in SAS can be used to compute weighted percentages of students within defined groups, and their means on a specified continuous variable. This macro also generates replicate weights and computes jackknife repeated replication (JRR) sampling variances for the percentages and mean estimates. The variable can be any continuous variable in the file. How to use each of these macro programs is described later in this chapter. When computing the average of the achievement scores or when computing plausible values, you will need to use the macro JACKPV.SAS

JACKREG.SAS

This macro program in SAS can be used to compute weighted regression coefficients and their corresponding standard errors within defined groups. It also computes descriptive statistics on the variables. This macro can be used with any variable in the analysis but it does not make use of the 5 plausible values.

JACKREGP.SAS

This macro program in SAS can be used to compute weighted regression coefficients and their corresponding standard errors when using plausible values as the predicted scores within defined groups. It also computes descriptive statistics on the variables.

Each of the four macros above has a corresponding sample program that calls these and prints out the results. These are discussed later in the chapter.

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

These are the programs used in the examples presented later in this chapter. These programs are included only in the CD with the 1999 data, although the same examples can be easily adapted to perform the same analyses with the 1995 data.

10.3 Creating SAS Datasets

The CD contains SAS control code to read each one of the ASCII data files and create a SAS dataset. Each of these control files contain information on the location of each variable in the file, its format, a descriptive label for each variable and their categories (in the case of categorical variables), and code for handling missing data. The control and data files have been created to facilitate access of the data on a country by country basis. The command lines in the control files should be edited to produce programs that will create SAS files for any specified countries. While most of the program code is functional as provided, you will need to edit input and output directories. Performing analyses that require the data from more than one country will necessitate merging or appending the respective data files into a larger one. Alternatively, you can access the data and compute the necessary statistics on a country by country basis by reading one file at a time, computing the necessary statistics, and then moving on to the next country's data. The method chosen by you will depend greatly on the storage and processing capacity of the computer system that is used. For the examples that we present in this User Guide we have combined the data files of individual countries into one larger data file that contains the data for all participating countries. The 3-character identifier for this file is "ALL".

When creating a dataset using SAS, you will need to do the following:

1. Open the corresponding control file, for example, BSGCTRM2.SAS .

2. In the line where it reads "%LET CTY = …" select the three character identification code for each country whose data you want to create. If you leave this line unedited the data for all countries listed will be created.

3. In the line where it reads "LIBNAME…" you will need to enter the path where you want to store the datasets that you will be creating.

4. In the line where it reads "FILENAME rawinp…" enter the path where the raw data is located.

5. Submit the code for processing. After processing is complete you will find the corresponding datasets in the location you specified in step 3 above.

For example, in the extract presented in exhibit 10.1 the student background data for Australia, Belgium Flemish, Bulgaria, Canada, England and the United States will be read from the directory "/usr/isc/timss/idb/data/raw/." Separate SPSS system files will be stored in the directory "/usr/isc/timss/idb/data" under the name BSG<3-character country code>M2. Only the sections that need to be modified are presented in this exhibit.

Exhibit 10.1 Extract from SAS Control Code for Creating a Student Background SAS Data Set

```
%let cty = AUS BFL BGR CAN ENG USA;


Options nosource nocenter;

LIBNAME libdat  "/usr/isc/timss/idb/data" ;
LIBNAME library "/usr/isc/timss/idb/data" ;

%macro doit;
  %let i = 1;
     %do %while(%length(%scan(&cty,&i)));
       %let ccode = %lowcase(%scan(&cty,&i));

%let fname= bsg&ccode.m2 ;

FILENAME rawinp "/usr/isc/timss/idb/data/raw/&fname..dat" ;


%PUT INFO: data set definition ;
%PUT INFO: data set definitions may need to be modified ;

*Output data set ;
data libdat.&fname;
.
.


RUN ;

       %let i = %eval(&i + 1);
     %end;

%mend doit;

%doit;
```

10.4 Scoring the Items

There were several types of items administered as part of the TIMSS tests. There were multiple-choice items, in which the student was asked to select one of four or five options as the correct response. These were administered as part of the achievement test. The responses to these items are coded with one digit. The codes used to represent the responses to these items are as follows.

Code 1	Option A
Code 2	Option B
Code 3	Option C
Code 4	Option D
Code 5	Option E
Code 6	Not reached
Code 7	Invalid response (chose more than one of the options available)
Code 8	Not administered
Code 9	No response although the item was administered and was reached (i.e. item waas omiited)

There were also free-response items where the students were asked to construct a response to a question, rather than choosing an answer from a list of options. The answers to these questions were scored by scorers trained to use the two-digit scoring rubrics described in an earlier chapter. The first digit of the two-digit code indicates the score given to the question, and the second digit in conjunction with the first provides diagnostic information on the specific answer given by the student. These types of response codes were used for the free-response items administered as part of the achievement test. The codes used to represent the responses to these items are the following:

Code 30 to 39	Three-point answer. Second digit provides diagnostic information.
Code 20 to 29	Two-point answer. Second digit provides diagnostic information.
Code 10 to 19	One-point answer. Second digit provides diagnostic information.
Code 70 to 79	Zero-point answer. Second digit provides diagnostic information.
Code 90	Uninterpretable.[2]
Code 96	Not reached.
Code 98	Not administered.

The achievement data files contained in the CD include information about the answer given to each item administered to a student. You might want to work with these item data after they are recoded to the right-wrong format, in the case of multiple-choice items, or to the level of correctness in the case of the free-response items. To this effect, we have included in the CD a set of programs in SAS that will allow you to recode the items from the achievement test to their right-wrong or correctness-level format. These programs contain a macro called SCOREIT and the necessary call to this macro so that all the items in the corresponding file are scored. This program will convert the response option codes for multiple-choice items to dichotomous score levels (0 or 1) based on scoring keys. For the free-response items the two-digit codes will be converted to the corresponding correctness score level (3, 2, 1, 0) based on the value of the first digit, as described in Chapter 2.

Two files are included to provide control code to perform the recodes of the test items in the achievement test file:

BSASCRM1 Written Assessment Files in 1995

BSASCRM2 Written Assessment Files in 1999

▼ ▼ ▼

[2] This code was used only in 1995.

When using these programs, you must first consider the recoding scheme that is desired. For example, under certain circumstances you might want to recode the not reached responses as incorrect (codes 6 and 96), whereas under other circumstances you might want to recode these responses as not administered or invalid. In the case of TIMSS, not reached responses were recoded as not administered (and effectively as missing responses) for the purpose of calibrating the items. But the not-reached responses were then recoded as incorrect when scoring the item for the individual countries, and for the purpose of calculating the scale scores for the individuals. By default, the scoring program provided with the database recodes the items coded as not reached and those left blank as incorrect responses.

To use the SCOREIT macro you need to include it as part of the SAS programs used for the analysis. This is done by using the INCLUDE statement in the corresponding program. When using SAS, the scoring program code should be included when the dataset containing the item responses is created or read into memory and becomes the working file. Both of these programs recode the items onto themselves, so if you want to preserve the original answers and codes assigned to the questions, then the file with the recoded item variables needs to be saved under a different file name. A copy of the macro that scores the items in SAS is presented in Exhibit 10.2.

Exhibit 10.2 Extracted Sections of SAS Macro SCOREIT Used to Convert Cognitive Item Response Codes to Correctness-Score Levels

```
%MACRO SCOREIT(ITEM,TYPE,RIGHT,NR,NA,OM,OTHER) ;
 %IF %upcase(&TYPE) = "MC" %THEN %DO ;
   SCORE = 0 ;
     IF &ITEM = &RIGHT THEN SCORE = 1 ;
     IF &ITEM = &NR    THEN SCORE = 0 ;
     IF &ITEM = &NA    THEN SCORE = . ;
     IF &ITEM = &OM    THEN SCORE = 0 ;
     IF &ITEM = &OTHER THEN SCORE = 0 ;
   &ITEM = SCORE;
 %END;
 %IF %upcase(&TYPE = "OE") %THEN %DO;
   SCORE = 0 ;
     IF &ITEM >= 30 AND &ITEM < 40 THEN SCORE = 3 ;
     IF &ITEM >= 20 AND &ITEM < 30 THEN SCORE = 2 ;
     IF &ITEM >= 10 AND &ITEM < 20 THEN SCORE = 1 ;
     IF &ITEM >= 70 AND &ITEM < 80 THEN SCORE = 0 ;
     IF &ITEM = &NR    THEN SCORE = 0 ;
     IF &ITEM = &NA    THEN SCORE = . ;
     IF &ITEM = &OM    THEN SCORE = 0 ;
     IF &ITEM = &OTHER THEN SCORE = 0 ;
   &ITEM = SCORE;
 %END;
%MEND SCOREIT;

%LET ARIGHT  = <list items where option A is the correct one>;
%LET BRIGHT  = <list items where option B is the correct one>;
%LET CRIGHT  = <list items where option C is the correct one>;
%LET DRIGHT  = <list items where option D is the correct one>;
%LET ERIGHT  = <list items where option E is the correct one>;
%LET OPENEND = <list open-ended items>;

ARRAY ARIGHT  &ARIGHT;
ARRAY BRIGHT  &BRIGHT;
ARRAY CRIGHT  &CRIGHT;
ARRAY DRIGHT  &DRIGHT;
ARRAY ERIGHT  &ERIGHT;
ARRAY OPENEND &OPENEND;

DO OVER ARIGHT ; %SCOREIT(ARIGHT,"MC",1, 6, 8, 9, 7); END;
DO OVER BRIGHT ; %SCOREIT(BRIGHT,"MC",2, 6, 8, 9, 7); END;
DO OVER CRIGHT ; %SCOREIT(CRIGHT,"MC",3, 6, 8, 9, 7); END;
DO OVER DRIGHT ; %SCOREIT(DRIGHT,"MC",4, 6, 8, 9, 7); END;
DO OVER ERIGHT ; %SCOREIT(ERIGHT,"MC",5, 6, 8, 9, 7); END;
DO OVER OPENEND; %SCOREIT(OPENEND,"OE", ,96,98,99,90); END;
```

10.5 Basic Analyses with the TIMSS Data: Means, Percentages, Regression Coefficients and Their JRR Standard Errors

In this section four macros that can be used to compute the correct standard errors of sampling and imputation are described, including examples in which these macros are used to replicate tables in the TIMSS 1999 International Reports and Almanacs.

Computing Sampling and Imputation Variance for Plausible Values Using SAS (JACKPV.SAS)

This section presents example SAS code that can be used to compute the JRR standard errors for mean plausible values and percentages. This code is provided in the form of a SAS macro thatcomputes the percentages of students within subgroups defined by a set of classification variables, the JRR standard errors of these percentages, the means for the groups on one of the achievement scales using plausible values, and the standard errors of these means including the sampling and imputation variance components.

The JACKPV.SAS macro operates as follows:

1. Computes a set of replicate weights specified using the parameters NJKZ, JKZ, and JKR.

2. Aggregates or summarizes the file data by computing the sum of the weights for each subgroup, the sum of the weights overall, and the sum of a weighted set of plausible values.

3. Computes the percentages of people within each group, their means on the plausible values, and their corresponding standard errors. The resulting working file contains the corresponding statistics. The data set FINAL in the default directory contains the corresponding statistics.

When using this macro, you need to specify a set of classification variables, the name of the plausible values and how many there are, the number of replicate weights to be generated, the variables that contains the sampling information such as JKZONE and JKREP, and the sampling weight that is to be used for the analysis. You will also need to specify the data file that contains the data that is to be processed.

You need to know some basic SAS macro language in order to use JACKPV.SAS. The macro should be included in the program file where it is going to be used. If you are operating in batch mode, then the macro should be called in every batch. If you are using SAS interactively then the macro should be called once at the beginning of the session and it will remain active throughout the session. If the session is terminated and restarted at a later time the macro should be called once again. Once the macro is included in a specific session the word "JACKPV" should not be used within that session because doing so will invoke the macro.

The macro is included in the program file where it will be used by issuing the following command under SAS:

```
%include '<path>jackpv.SAS'.
```

where <path> points to the specific drive and directory where the macro JACKPV.SAS can be found. The macro requires that several parameters be specified when it is invoked. These parameters are:

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

CVAR	This lists the variables that are to be used to classify the students in the data file. This can be a single variable, or a list of variables. The maximum number of variables will depend mostly on the computer resources available to you at the time. It is recommended to always include the variable that identifies the country. At least one variable has to be specified, usually IDCNTRY.
ROOTPV	This is the prefix used to identify the plausible values for the achievement scale of interest. This corresponds to the first 7 characters of the plausible value variables. For example, the root of the overall science plausible value is "BSSSCI0," the root of the geometry plausible values is "BSMGEO0."
NPV	This is the number of plausible values that will be used in the analysis. Generally you will want to use all five plausible values for the analysis although under some circumstances less can be used. You should always use at least two plausible value for any analysis.
JKZ	The variable that captures the assignment of the student to a particular sampling zone. The name of this variable in all TIMSS files is JKZONE.
JKR	The variable that captures whether the case is to be dropped or have its weight doubled for the corresponding replicate weight. The name of this variable in all TIMSS files is JKREP.
NJKZ	This indicates the number of replicate weights to be generated when computing the JRR error estimates. When conducting analyses using the data from all countries, the value of NJKZ should be set to 75, the maximum possible value. When you are working with the data for only one country, you should set the NJKZ argument to as many replicates as are needed in the country. If the data from two or more countries is being used for an analysis, then the largest number of jackknife zones should be used. When in doubt about what number to set the NJKZ parameter, it should be set to 75. The error variance will always be estimated correctly if more replicate weights than necessary are computed, but will be underestimated if you specify less replicate weights than necessary.

WGT The sampling weight to be used in the analysis, generally TOTWGT when using the student files, or MATWGT, SCIWGT, or TCHWGT when using the teacher files.

The simplest and most straightforward way to invoke the macro is by using the conventional SAS notation for invoking macros. This involves listing the macro name followed by the corresponding list of arguments for the analysis in parenthesis, each separated by a comma. For example, if the macro is invoked using

```
%jackpv(totwgt , jkzone, jkrep, 75, idcntry itsex, bsssci0, 5, bsgallm2);
```

it will compute the mean science achievement scores using five plausible values, and their corresponding standard error, for boys and girls within each country, using the variable TOTWGT as the sampling weight. It will also compute the percentages of boys and girls within the country, and their corresponding standard errors. The data will be read from the data set BSGALLM2, and the standard error of the statistics will be computed based on 75 replicate weights.

The file that contains these results is then called FINAL and is saved to the default directory being used by SAS. The variables that are contained in this file are:

Classification Variables

Each of the classification variables is kept in the resulting file. In the above example there are two classification variables in the FINAL data set, IDCNTRY and ITSEX. There is one unique occurrence for each combination of the categories for these variables.

Weight Variable

Contains the estimate in the population that belongs to the groups defined by each specific combination of the classification variable categories. In our example this variable is called TOTWGT.

N

Contains the number of cases in the groups defined by each specific combination of categories for the classification variables. In the example this is the number of boys and girls in the sample for each country.

PCT

Contains the weighted percentages of people in the groups for the classification variable listed last, within the specific combination of the categories defined by the classification variables. In the example, it is the percentage of boys and girls within each country.

PCT_SE

Contains the standard errors of PCT computed using the JRR method for computing the standard error.

MNX

Contains the weighted means for the first plausible value for the groups defined by the corresponding combinations of classification variable categories.

MNX_SE

Contains the JRR standard errors of the means for the first plausible value for the groups computed using the JRR method for computing the standard error. This does not include the imputation error that should be computed when using plausible values.

MNPV

Contains the means of the plausible values used in the analysis.

MNPV_SE

Contains the standard errors for the means of the plausible values. These standard errors contain the sampling and the imputation components of the errors of the estimates.

The file resulting from using this macro can be printed using a SAS procedure of choice. An example call to this macro, and a printout of the resulting file is presented in Exhibit 10.3 below. The code is included in the file SampleJackPV.SAS.

Exhibit 10.3 SAS Control Code and Extract of Output File for Using the Macro JACKPV.SAS

```
libname bm2 "d:\timss\data\bm2";

data student;
 set bm2.bsgallm2;
  where itsex in(1,2) and intms99 = 1 and idgrader = 2;

%include "d:\timss\programs\jackpv.sas";
%jackpv (totwgt , jkzone, jkrep, 75, idcntry itsex, bsssci0, 5, student);

proc print data=final noobs;
     var idcntry itsex N totwgt mnpv mnpv_se mnx mnx_se pct pct_se;
     format idcntry itsex N 6.0 totwgt 10.0 mnpv mnpv_se mnx mnx_se pct pct_se 6.2;

run;

IDCNTRY  ITSEX     N     TOTWGT     mnpv    mnpv_se      mnx    mnx_se      pct    pct_se

  36       1     2033    132651    531.54     5.06     531.89     4.63     50.99      2.44
  36       2     1999    127478    549.33     5.96     548.72     5.27     49.01      2.44
 100       1     1676     45121    511.16     5.83     509.95     5.62     51.12      1.45
 100       2     1593     43146    525.04     6.51     526.72     6.25     48.88      1.45
 124       1     4392    185897    526.22     3.21     527.82     2.60     50.16      0.98
 124       2     4373    184743    540.26     2.42     539.56     2.33     49.84      0.98
 152       1     2981    104646    408.90     4.27     409.75     3.70     50.09      1.88
 152       2     2926    104263    431.89     5.14     431.30     4.81     49.91      1.88
 158       1     2874    155846    560.66     3.94     563.86     3.28     50.20      1.13
 158       2     2898    154582    577.56     5.66     580.82     4.80     49.80      1.13
 196       1     1528      4774    454.92     3.09     456.74     2.41     48.80      0.58
 196       2     1587      5009    465.36     3.02     466.58     2.65     51.20      0.58
.
.
.
.
.
 764       1     3138    395196    480.78     4.59     480.39     4.37     54.35      1.45
 764       2     2594    331890    484.15     4.44     483.84     4.25     45.65      1.45
 788       1     2561     71001    417.02     3.33     416.08     2.09     50.85      0.76
 788       2     2490     68638    442.43     4.33     439.53     2.30     49.15      0.76
 792       1     3301    260041    431.21     4.81     430.79     3.93     42.07      1.52
 792       2     4540    358016    434.21     4.26     433.24     3.63     57.93      1.52
 807       1     1988     14928    457.70     6.03     455.85     5.02     49.30      0.84
 807       2     2035     15352    458.47     5.42     460.49     4.43     50.70      0.84
 840       1     4617   1666060    505.40     4.59     506.79     4.17     49.94      0.75
 840       2     4455   1670235    524.40     5.47     522.55     5.12     50.06      0.75
 926       1     1455    268833    521.97     6.20     526.33     5.42     48.70      1.76
 926       2     1504    283133    554.12     5.25     555.33     5.00     51.30      1.76
 956       1     2644     32632    525.75     4.61     525.98     4.42     49.79      3.15
 956       2     2615     32907    543.89     7.19     542.03     6.50     50.21      3.15
```

In this example, the mean of all five plausible values and the mean of the first plausible value and their corresponding standard errors for science achievement are calculated separately for boys and girls by country. In the listing of the results we can see that there are entries for each country corresponding to the results for females (ITSEX = 1) and males (ITSEX = 2) after selecting only those cases with IDGRADER=2. The first column has the country code, the second column indicates the gender of the students, the third column has the number of students in each sample, and the fourth column has the total weight of this sample in the population. This is followed by the mean of all five plausible values in science achievement and the corresponding standard error. Following this is the mean of the first plausible value in science achievement and its corresponding standard error. The last two columns indicate the percentage of girls and boys in each country's sample and their corresponding standard errors.

For example, Australia (idcntry 36) sampled 2033 girls representing 132651 students in the whole population. The mean of the five plausible values for these girls is 531.54 with a standard error of 5.06. Girls made up 50.99% of Australia's sampled students. The mean of the first plausible value for the girls is 531.89 with a standard error of 4.63. Additionally, Australia sampled 1999 boys representing 127478 students in the whole population. The mean of the five plausible values for these boys is 549.33 with a standard error of 5.96. Boys made up 49.01% of Australia's sampled students. The mean of the first plausible value for he boys is 548.72 with a standard error of 5.27.

Computing Sampling Variance for Variables Other than Plausible Values (JACKGEN.SAS)

In this section example SAS code that can be used to compute the JRR standard errors for means and percentages of variables other than plausible values is described. This code is provided in the form of an SAS macro called JACKGEN.SAS thatcomputes the percentages of students within subgroups defined by a set of classification variables, the JRR standard errors of these percentages, the means for the groups on a variable of choice, and the JRR standard errors of these means. Although you can compute weighted percent and mean estimates using other basic SAS commands, the macro JACKGEN.SAS also computes the JRR error estimate for these means and percentages.

This macro operates as follows:

1. Computes a set of replicate weights, usually 75, for each record using the procedure described in the previous chapter.

2. Aggregates or summarizes the file data by computing the sum of the weights for each subgroup, the sum of the weights overall, and the sum of a weighted set of analysis variable values.

3. Computes the percentages of people within each group, their means on the analysis variable, and their corresponding standard errors. The resulting dataset, called FINAL, contains the corresponding statistics.

When using this macro, you need to specify a set of classification variables, one analysis variable, the number of replicate weights to be generated, the variables that contain the sampling information such as JKZONE and JKREP, and the sampling weight that is to be used for the analysis. You will also need to specify the data file that contains the data that is to be processed.

You need to know some basic SAS macro language in order to use the macro. It needs to be first included in the program file where it is going to be used. If you are operating in batch mode, then the macro needs to be called in every batch. If you are using SAS interactively then the macro needs to be called once at the beginning of the session and it will remain active throughout the session. If the session is terminated and restarted at a later time the macro needs to be called once again. Once the macro is included in a specific session the word "JACKGEN" should not be used within that program because doing so will invoke the macro.

The macro is included in the program file where it will be used by issuing the following command under SAS:

```
%include '<path>jackgen.SAS'.
```

where <path> points to the specific drive and directory where the macro JACKGEN.SAS can be found. The macro requires that several argument be specified when it is invoked. These parameters are:

INFILE	The name of the data file that contains the variables necessary for the analysis. If the path location is included as part of the file name the name of the file has to be enclosed in quotes. It is important to emphasize that this dataset must include only those cases that are of interest in the analysis. If you want to have specific cases excluded from the analysis, as for example students with missing variables or selected students from a specific grade, this should be done prior to invoking the macro.
CVAR	This lists the variables that are to be used to classify the students in the data file. This can be a single variable, or a list of variables. The maximum number of variables will depend mostly on the computer resources available to you at the time. It is recommended to always include the variable that identifies the country. At least one variable has to be specified, usually IDCNTRY.
DVAR	This is the variable for which means are to be computed. Only one variable can be listed here. If you want to examine, for example, results in two different variables, then the macro needs to be invoked separately to generate each table.
NJKZ	This indicates the number of replicate weights to be generated when computing the JRR error estimates. When conducting analyses using the data from all countries, the value of NJKZ should be set to 75. When you are working with the data for only one country, you should set the NJKZ argument to as many replicates as there were in the country. If the data from two or more countries is being used for an analysis, then the larger number of jackknife zones should be used. When in doubt about what number to set the NJKZ parameter, it should be set to 75. The error variance will always be estimated correctly if more replicate weights than necessary are computed,

but will be underestimated if you specify less replicate weights than necessary.

JKZ — The variable that captures the assignment of the student to a particular sampling zone. The name of this variable in all TIMSS files is JKZONE.

JKR — The variable that captures whether the case is to be dropped or have its weight doubled for the corresponding replicate weight. The name of this variable in all TIMSS files is JKREP.

WGT — The sampling weight to be used in the analysis, generally TOTWGT when using the student files, or MATWGT, SCIWGT, or TCHWGT when using the teacher files.

The simplest and most straightforward way to invoke the macro is by using the conventional SAS notation for invoking macros. This involves listing the macro name followed by the corresponding list of arguments for the analysis in parentheses, each separated by a comma. For example, if the macro is invoked using:

```
%jackgen(totwgt , jkzone, jkrep, 75, idcntry idgrader, bsdage, bsgallm2);
```

it will compute means of age and their corresponding standard errors, by grade, within each country, using the variable TOTWGT as the sampling weight. The data will be read from the dataset BSGALLM2.

The file that contains these results is then called FINAL and is saved to the default directory being used by SAS. The variables that are contained in this file are:

Classification Variables

Each of the classification variables is kept in the resulting file. In our example above there are two variables in the resulting dataset. These are IDCNTRY and IDGRADER. There is one unique occurrence for each combination of the categories for these variables.

Weight Variable

Contains the estimates in the population that belong to the group defined by each specific combination of the classification variable categories.

MNX

Contains the weighted means of the variable DVAR for the groups defined by the corresponding combinations of classification variable categories.

MNX_SE

Contains the standard errors of the MNX values computed using the JRR method for computing the standard error.

PCT

Contains the weighted percentages of people in the groups for the classification variable listed last, within the specific combination of the categories defined by the groups initially. In our example, we would obtain the percentage of students by grade for each country.

PCT_SE

Contains the standard errors of PCT computed using the JRR method for computing the standard error.

The file resulting from using this macro can then be printed using a SAS procedure of choice. An example call to this macro, and a subset of the resulting file is presented in Exhibit 10.4. In this example the macro will compute the percentages in each grade, by country, and the mean achievement scores in science. This code is included in the file SampleJACKGEN.SAS.

Exhibit 10.4 SAS Control Code and Extract of Output File for Using the Macro JACKGEN.SAS

```
libname bm2 "d:\timss\data\bm2";

data student;
 set bm2.bsgallm2;
  where intms99=1 and bsdage > 0 and idgrader = 2;

%include "d:\timss\programs\jackgen.sas";
%jackgen (totwgt , jkzone, jkrep, 75, idcntry idgrader, bsdage, student);

proc print data=final noobs;
     var idcntry idgrader N totwgt mnx mnx_se pct pct_se;
     format idcntry idgrader n 6.0 totwgt 10.0 mnx mnx_se pct pct_se 6.2;

run;

IDCNTRY    IDGRADER         N        TOTWGT       mnx    mnx_se       pct    pct_se

    36            2      3652        242266     14.26      0.02    100.00      0.00
   100            2      3267         88189     14.79      0.04    100.00      0.00
   124            2      8711        368224     14.03      0.01    100.00      0.00
   152            2      5898        208619     14.35      0.02    100.00      0.00
   158            2      5772        310429     14.20      0.01    100.00      0.00
   196            2      3015          9458     13.80      0.01    100.00      0.00
   203            2      3448        119207     14.40      0.01    100.00      0.00
   246            2      2906         59415     13.84      0.01    100.00      0.00
   344            2      5156         78764     14.24      0.03    100.00      0.00
   348            2      3149        110043     14.44      0.02    100.00      0.00
   360            2      5842       1953607     14.57      0.03    100.00      0.00
   364            2      5290       1651634     14.56      0.03    100.00      0.00
   376            2      4122         80028     14.05      0.03    100.00      0.00
   380            2      3328        548711     13.96      0.01    100.00      0.00
   392            2      4684       1397517     14.38      0.00    100.00      0.00
   400            2      5047         89069     14.00      0.01    100.00      0.00
   410            2      6113        609393     14.43      0.01    100.00      0.00
   428            2      2821         17845     14.50      0.02    100.00      0.00
   440            2      2346         40155     15.19      0.02    100.00      0.00
   458            2      5577        397761     14.36      0.02    100.00      0.00
   498            2      3643         58888     14.41      0.02    100.00      0.00
   504            2      5099        328317     14.19      0.02    100.00      0.00
   528            2      2890        193356     14.21      0.03    100.00      0.00
   554            2      3584         51110     14.03      0.01    100.00      0.00
   608            2      6593       1077070     14.07      0.04    100.00      0.00
   642            2      3419        259136     14.81      0.02    100.00      0.00
   643            2      4328       2055768     14.10      0.02    100.00      0.00
   702            2      4966         41346     14.35      0.01    100.00      0.00
   703            2      3475         72053     14.27      0.01    100.00      0.00
   705            2      3092         23387     14.76      0.01    100.00      0.00
   710            2      7601        776084     15.48      0.10    100.00      0.00
   764            2      5727        726455     14.46      0.02    100.00      0.00
   788            2      5042        139393     14.78      0.03    100.00      0.00
   792            2      7836        617822     14.20      0.02    100.00      0.00
   807            2      4018         30229     14.62      0.01    100.00      0.00
   840            2      8776       3234450     14.18      0.02    100.00      0.00
   926            2      2842        529162     14.20      0.01    100.00      0.00
   956            2      5221         64979     14.12      0.02    100.00      0.00
```

In this example BSDAGE is used to calculate the mean age of the sampled students in each country. In the listing of the results we can see that there is one entry or line for each of the values of the variable IDCNTRY after selecting only those cases with IDGRADER=2. The first column has the country code, the second column has the grades, the third column has the number of students in each sample after selecting only those cases where BSDAGE is not missing, and the fourth column is the total weight this sample represents in the whole population. This is followed by the mean age of the sampled students and the corresponding standard error.

For example, from line one in the above results, we can determine that Australia sampled 3652 students from the population 2 grade representing 242266 students in the whole population, with a mean age of 14.26 and a standard error of .02. Since there is only one grade in the category, each set represents 100% with a zero standard error.

Computing Regression Coefficients and Their JRR Standard Errors for Variables Other Than Plausible Values (JACKREG.SAS)

In this section, example SAS code that may be used to compute regression coefficients and their JRR standard errors is described. The CD containing the TIMSS International Database contains the SAS macro program called JACKREG.SAS. The macro computes the multiple correlation between the specified dependent and independent variables within a subgroup defined by a set of classification variables, as well as the regression coefficients, and the JRR standard error of the regression coefficients.

If you wish to conduct regression analyses using plausible values as the dependent variables, please refer to JACKREGP.SAS, which is described in the next section.

This macro operates as follows:

1. Computes a set of replicate weights, usually 75, for each record using the procedure described in the previous chapter.

2. Aggregates or summarizes the file data by computing the sum of the weights for each category, the sum of the weights overall, and the sum of a weighted set of analysis variable values.

3. Computes the regression coefficients within each group and their corresponding standard errors. The resulting dataset, FINAL, contains the corresponding statistics.

When using this macro, you need to specify a set of classification variables, the analysis variable, the number of replicate weights to be generated, the variables that contain the sampling information such as JKZONE and JKREP, and the sampling weight that is to be used for the analysis. You will also need to specify the data file that contains the data that is to be processed.

You need to know some basic SAS macro language in order to use the macro. It needs to be first included in the program file where it is going to be used. If you are operating in batch mode, then the macro needs to be called in every batch. If you are using SAS interactively then the macro needs to be called once at the beginning of the session and it will remain active throughout the session. If the session is terminated and restarted at a later time the macro needs to be called once again.

Once the macro is included in a specific session the word “JACKREG” should not be used within that program because doing so will invoke the macro.

The macro is included in the program file where it will be used by issuing the following command under SAS:

```
%include '<path>jackreg.SAS'.
```

where <path> points to the specific drive and directory where the macro JACKREG.SAS can be found. The macro requires that several parameters be specified when it is invoked. These parameters are:

INFILE	The name of the data file that contains the variables necessary for the analysis. If the path location is included as part of the file name, the name of the file has to be enclosed in quotes. It is important to emphasize that this dataset must include only those cases that are of interest in the analysis. If you want to have specific cases excluded from the analysis, as for example students with missing variables or selected students from a specific grade, this should be done prior to invoking the macro.
CVAR	This lists the variables that are to be used to classify the students in the data file. This can be a single variable, or a list of variables. The maximum number of variables will depend mostly on the computer resources available to you at the time. It is recommended to always include the variable that identifies the country. At least one variable has to be specified, usually IDCNTRY.
XVAR	This is a list of independent variables, at least one, that under the regression model will be used as predictors of the dependent variable specified in DVAR. These independent variables can be continuous or categorical, or any other type of coded variable. For example, it could be the variable ITSEX as originally coded in the data files, or dummy coded as 1 or 0.
DVAR	This is the dependent variable that under the regression model is predicted by the variable or variables specified by the XVAR parameter. Only one variable can be listed here. If you want to use the same set of predictor variables to predict two different variables, then the macro needs to be invoked separately to generate each set of results.

NJKZ — This indicates the number of replicate weights to be generated when computing the JRR error estimates. When conducting analyses using the data from all countries, the value of NJKZ should be set to 75. When you are working with the data for only one country, you should set the NJKZ argument to as many replicates as there were in the country. If the data from two or more countries is being used for an analysis, then the larger number of jackknife zones should be used. When in doubt about what number to set the NJKZ parameter, it should be set to 75. The error variance will always be estimated correctly if more replicate weights than necessary are computed, but will be underestimated if you specify less replicate weights than necessary.

JKZ — The variable that captures the assignment of the student to a particular sampling zone. The name of this variable in all TIMSS files is JKZONE.

JKR — The variable that captures whether the case is to be dropped or have its weight doubled for the corresponding replicate weight. The name of this variable in all TIMSS files is JKREP.

WGT — The sampling weight to be used in the analysis, generally TOTWGT when using the student files, or MATWGT, SCIWGT, or TCHWGT when using the teacher files.

The simplest and most straightforward way to invoke the macro is by using the conventional SAS notation for invoking macros. This involves listing the macro name followed by the corresponding list of arguments for the analysis, each separated by a comma. For example, if the macro is invoked using:

```
%jackreg(totwgt , jkzone, jkrep, 75, idcntry idgrader, regsex, bsdgstdt, bsgallm2);
```

it will compute the regression equation for the variable REGSEX as a predictor of the number of hours spent studying. The data will be read from the data set BSGALLM2 and the standard error of the statistics will be computed based on 75 replicate weights.

The file that contains these results is called REG and can be found in the default directory. The variables that are contained in this file are:

Classification Variables

Each of the classification variables is kept in the resulting file. In our example above there would be two variables in the resulting dataset, IDCNTRY and IDGRADER. There is one unique occurrence for each combination of the categories for these variables.

Mult_RSQ

The squared multiple correlation coefficient for the model.

SS_Res, SS_Reg, SS_Total

The residual, regression, and total sum of squares for the model within each group as defined by the classification variables.

Regression Coefficients and Standard Errors (B## and B##_SE)

These are the regression coefficients for each of the predictor variables in the model and their corresponding jackknifed standard errors. The coefficient zero (B00) is the intercept for the model. The other coefficients receive a sequential number starting with 01. This sequential number corresponds to the order of the variables in the list of variables specified in the parameter XVAR.

The file resulting from using this macro can then be printed using a SAS procedure of choice. An example call to this macro, and a subset of the resulting file is presented in Exhibit 10.5. This code is included in the file SampleJACKREG.SAS

Exhibit 10.5 SAS Control Code and Extract of Output File for Using the Macro JACKREG.SAS

```
libname bm2 "d:\timss\data\bm2";

data student;
 set bm2.bsgallm2;
  where itsex in(1,2) and intms99=1;
   regsex = itsex - 1;

%include "d:\timss\programs\jackreg.sas";
%jackreg (totwgt , jkzone, jkrep, 75, idcntry idgrader, regsex, bsdgstdt, student);

proc print data=reg noobs;
     where idgrader = 2;
     var idcntry N Mult_RSQ SS_Res SS_Reg SS_Total b00 b00_se b01 b01_se;
     format idcntry n 6.0 Mult_RSQ 5.3
            SS_Total SS_Reg SS_Res 10.0 B00 B00_SE B01 B01_SE 6.2;
run;

IDCNTRY    N     Mult_RSQ    SS_Res   SS_Reg  SS_Total     b00   b00_se    b01  b01_se

 36     3841      0.009     449814     4186    442443    2.08    0.05    -0.25    0.06
100     2792      0.049     235956    12204    248160    3.39    0.08    -0.80    0.09
124     8371      0.019     760287    14716    775003    2.43    0.05    -0.41    0.05
152     5317      0.015     517818     7775    525593    2.63    0.05    -0.41    0.06
158     5489      0.016     823214    13569    836783    2.20    0.06    -0.43    0.05
196     2848      0.043      26456     1187     27643    3.22    0.06    -0.73    0.06
203     3382      0.042     158369     6927    165295    2.13    0.05    -0.49    0.05
246     2852      0.012      48049      581     48630    1.88    0.03    -0.20    0.04
344     4970      0.003     176151      554    176706    1.73    0.06    -0.17    0.07
348     3001      0.051     244410    13006    257416    3.17    0.05    -0.71    0.06
360     5285      0.021    5262695   115350   5378045    3.29    0.05    -0.51    0.06
364     3802      0.010    3943955    41262   3985216    4.32    0.07    -0.39    0.09
376     3660      0.020     180099     3694    183793    2.90    0.06    -0.46    0.07
380     3029      0.055    1560554    91651   1652205    4.03    0.05    -0.86    0.07
392     4640      0.004    2712643    10930   2723573    1.82    0.05    -0.18    0.04
400     3278      0.029     240479     7177    247656    4.21    0.07    -0.71    0.10
410     5919      0.004    1226404     4348   1230753    1.71    0.04    -0.17    0.06
428     2650      0.036      41833     1566     43399    3.32    0.05    -0.61    0.07
440     2186      0.083      83853     7595     91448    3.25    0.06    -0.90    0.07
458     4843      0.024    1060761    26575   1087336    4.10    0.06    -0.56    0.06
498     2865      0.023     156736     3742    160478    3.61    0.05    -0.57    0.08
504     3035      0.006     771961     4858    776820    3.51    0.07    -0.33    0.08
528     2841      0.015     268735     4205    272939    2.36    0.05    -0.30    0.06
554     3425      0.022      89525     2038     91563    2.20    0.04    -0.41    0.06
608     4527      0.002    2596503     4320   2600824    3.41    0.05    -0.15    0.07
642     2791      0.039     759805    30525    790330    3.80    0.08    -0.76    0.09
643     3703      0.020    4885180   101575   4986756    3.39    0.05    -0.48    0.05
702     4505      0.024     115797     2799    118597    3.81    0.05    -0.55    0.06
703     3365      0.051     117520     6348    123868    2.60    0.05    -0.60    0.07
705     2984      0.039      49707     2009     51716    2.77    0.05    -0.60    0.06
710     5537      0.001    2224060     1689   2225748    3.19    0.07    -0.11    0.06
764     5296      0.026    1667980    44822   1712802    3.15    0.04    -0.52    0.05
788     3753      0.010     393483     3855    397338    3.94    0.06    -0.39    0.07
792     6286      0.021    1691707    37148   1728855    3.94    0.05    -0.55    0.05
807     3125      0.019      75946     1437     77383    3.77    0.05    -0.50    0.07
840     8430      0.031    6638022   214619   6852641    2.34    0.04    -0.52    0.03
956     4861      0.029     143824     4274    148098    3.22    0.05    -0.53    0.07
```

In this example, the variable REGSEX is created by subtracting one from the variable ITSEX. As a result the girls receive a code of 0 and the boys receive a code of 1 on this variable. In this particular model the variable REGSEX is used to predict the values of the variable BSDGSTDT by country and by grade. In the listing of the results we can see that there is one entry or line for each value of the variable IDCNTRY after selecting only those cases with IDGRADER=2. The first column has the country code and the second column has the number of students in each sample. Keep in mind that this number corresponds only to those students who had valid data for the variables in the regression model. Then the multiple R and R squared are listed for each country, followed by the sums of squares for the model and corresponding regression coefficients. Because of the way in which the variable REGSEX is coded, the constant (B00) is the mean value for the variable BSDSTDT for the girls and the error of the constant (B00_SE) is standard error of the estimate. The first regression coefficient (B01) is the difference in the average value of the variable BSDSTDT between the boys and the girls, and the standard error of this coefficient (B01_SE) is the jackknifed standard error of this difference.

For example from the first line in the output we can say that in Australia (IDCNTRY = 36) data was available for 3841 cases. The squared multile correlation between gender and hours of study is .009 with girls studying, on average about 2.08 hours per day and boys studying 15 minutes (0.26) less than girls on average. This difference is statistically significant (0.26 / .06 = 4.3).

Overall we can say that on average, and across all countries, girls report studying more hours than boys. This difference is significant in most cases, which can be determined by dividing the value of B01 by its standard error and comparing it to the appropriate critical value.

Computing Regression Coefficients and Their JRR Standard Errors with Plausible Values (JACKREGP.SAS)

In this section example SAS code that can be used to compute the JRR standard errors for regression coefficients using plausible values as the dependent variable is described. This code is provided in the form of an SAS macro called JACKREGP.SAS. This macro computes the multiple correlation coefficient between the specified plausible values and independent variables within subgroups defined by a set of classification variables, as well as the regression coefficients and the JRR standard errors.

Th JACKREGP.SAS macro operates as follows:

1. Computes a set of replicate weights, usually 75, for each record using the procedure described in the previous chapter.

2. Aggregates or summarizes the file data by computing the corresponding sum of the weights for each category, the sum of the weights overall, and the sum of a weighted set of analysis variable values.

3. Computes the regression coefficients within each group and their corresponding standard errors. The resulting working file, called FINAL, contains the corresponding statistics.

When using this macro, you need to specify a set of classification variables, the analysis variable, the number of replicate weights to be generated, the variables that contain the sampling information such as JKZONE and JKREP, and the sampling weight that is to be used for the analysis. You will also need to specify the data file that contains the data that is to be processed.

You need to know some basic SAS macro language in order to use the macro. It needs to be first included in the program file where it is going to be used. If you are operating in batch mode, then the macro needs to be called in every batch. If you are using SAS interactively then the macro needs to be called once at the beginning of the session and it will remain active throughout the session. If the session is terminated and restarted at a later time the macro needs to be called once again. Once the macro is included in a specific session the word "JACKREGP" should not be used within that program because doing so will invoke the macro.

The macro is included in the program file where it will be used by issuing the following command under SAS:

```
include '<path>jackregp.SAS'.
```

where <path> points to the specific drive and directory where the macro JACKREGP.SAS can be found. The macro requires that several parameters be specified when it is invoked. These parameters are:

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

CVAR — This lists the variables that are to be used to classify the students in the data file. This can be a single variable, or a list of variables. The maximum number of variables will depend mostly on the computer resources available to you at the time. It is recommended to always include the variable that identifies the country. At least one variable has to be specified, usually IDCNTRY.

XVAR — This is a list of independent variables, possibly one, which under the regression model will be used as predictors of the dependent variables specified by the plausible values. These independent variables can be continuous or categorical, or any other type of coded variable. For example, it could be the variable ITSEX as originally coded in the data files, or dummy coded as 1 or 0.

ROOTPV — This is the prefix used to identify the plausible values for the achievement scale of interest. This corresponds to the first 7 characters of the plausible value variables. For example, the root of the overall science plausible value is "BSSSCI0," the root of the geometry plausible values is "BSMGEO0."

NPV — This is the number of plausible values that will be used in the analysis. Generally you will want to use all five plausible values for the analysis although under some circumstances less can be used. You should always use at least two plausible value for any analysis.

NJKZ — This indicates the number of replicate weights to be generated when computing the JRR error estimates. When conducting analyses using the data from all countries, the value of NJKZ should be set to 75. When you are working with the data for only one country, you should set the NJKZ argument to as many replicates as there were in the country. If the data from two or more countries is being used for an analysis, then the larger number of jackknife zones should be used. When in doubt about what number to set the NJKZ parameter, it should be set to 75. The error variance will always be estimated correctly if more replicate weights than necessary are computed, but will be underestimated if you specify less replicate weights than necessary.

JKZ — The variable that captures the assignment of the student to a particular sampling zone. The name of this variable in all TIMSS files is JKZONE.

JKR — The variable that captures whether the case is to be dropped or have its weight doubled for the corresponding replicate weight. The name of this variable in all TIMSS files is JKREP.

WGT — The sampling weight to be used in the analysis, generally TOTWGT when using the student files, or MATWGT, SCIWGT, or TCHWGT when using the teacher files.

The simplest and most straightforward way to invoke the macro is by using the conventional SAS notation for invoking macros. This involves listing the macro name followed by the corresponding list of arguments for the analysis, each separated by a slash. For example, if the macro is invoked using:

```
%jackregF(totwgt,jkzone,jkrep,75,idcntry idgrader,regsex,bsssci0,5,bsgallm2);
```

it will compute the regression equation for the variable REGSEX as a predictor of the plausible values in science. The data will be read from the data set BSGALLM2 and the standard error of the statistics will be computed based on 75 replicate weights.

The file that contains these results is then called REG and is saved to the default directory being used by SAS. The variables that are contained in this file are:

Classification Variables

Each of the classification variables is kept in the resulting file. In our example above there would be two variables in the resulting dataset, IDCNTRY and IDGRADER. There is one unique occurrence for each combination of the categories for these variables.

Mult_RSQ

The squared multiple correlation coefficient for the model.

SS_Res, SS_Reg, SS_Total

The residual, regression, and total sum of squares for the model within each group as defined by the classification variables.

Regression Coefficients and Standard Errors (B## and B##_SE)

These are the regression coefficients for each of the predictor variables in the model and their corresponding jackknifed standard errors combined with the imputation error. The coefficient zero (B00) is the intercept for the model. The other coefficients receive a sequential number starting with 01. This sequential number corresponds to the order of the variables in the list of variables specified in the parameter XVAR.

The file resulting from using this macro can then be printed using a SAS procedure of choice. An example call to this macro, and a subset of the resulting file is presented in Exhibit 10.6. This code is included in the file SampleJACKREGP.SAS

Exhibit 10.6 SAS Control Code and Extract of Output File for Using the Macro JACKREGP.SAS

```
libname bm2 "d:\timss\data\bm2";

data student;
 set bm2.bsgallm2;
  where itsex in(1,2) and intms99=1;
   regsex = itsex - 1;

%include "d:\timss\programs\jackregP.sas";
%jackregP (totwgt , jkzone, jkrep, 75, idcntry idgrader, regsex , bsssci0,5, student);

proc print data=reg noobs;
    where idgrader = 2;
    var idcntry N Mult_RSQ SS_Res SS_Reg SS_Total b00 b00_se b01 b01_se;
    format idcntry n 6.0 Mult_RSQ 5.3
   SS_Total SS_Reg SS_Res 12.0 B00 B00_SE B01 B01_SE 6.2;
run;


IDCNTRY  n    Mult_RSQ    SS_Res     SS_Reg     SS_Total     b00   b00_se   b01  b01_se

 36   4032     0.011  1965803828   21533145   1987336973   531.54   5.06   17.80   6.85
100   3269     0.006   762406599    4404020    766810619   511.16   5.83   13.88   6.15
124   8765     0.008  2228824268   18552616   2247376884   526.22   3.21   14.04   3.87
152   5907     0.017  1582230836   28127046   1610357882   408.90   4.27   23.00   6.18
158   5772     0.009  2432306921   22268514   2454575435   560.66   3.94   16.90   4.21
196   3115     0.004    68814332     270839     69085171   454.92   3.09   10.44   3.90
203   3453     0.044   735402325   33353184    768755509   523.26   4.83   33.33   4.76
246   2920     0.004   363829635    1622161    365451796   530.13   4.00   10.23   4.97
344   5178     0.011   378820418    4226667    383047085   522.21   4.38   14.43   6.12
348   3166     0.023   767203483   17660238    784863720   540.08   4.05   25.20   4.20
360   5848     0.011 13785164882  157243045  13942407927   426.97   6.46   17.20   6.78
364   5301     0.033 11319666222  386210898  11705877120   429.58   5.66   31.00   7.59
376   4192     0.005   892821127    4418074    897239201   461.15   6.02   14.39   6.09
380   3328     0.011  4151341445   47595532   4198936976   484.39   4.13   18.17   5.83
392   4745     0.009  8093677754   69695742   8163373496   542.75   2.79   13.63   4.59
400   5052     0.007   934581924    7002933    941584856   459.73   5.03  -17.66   8.22
410   6114     0.015  4316627249   67765990   4384393239   538.11   3.98   20.87   5.07
428   2872     0.010   109153583    1064028    110217611   495.29   5.57   15.20   3.96
440   2361     0.017   274084673    4611645    278696318   477.85   4.44   21.29   4.57
458   5577     0.004  2655473156   10038330   2665511485   488.11   5.47    9.47   7.02
498   3711     0.003   541417440    1845789    543263230   454.21   4.37   10.63   5.44
504   5392     0.008  3610433250   30338714   3640771963   312.26   5.94   17.95   8.31
528   2958     0.014  1146056617   16409182   1162465799   536.21   7.13   18.07   4.06
554   3613     0.002   445520073     677691    446197763   506.06   5.45    7.03   7.82
608   6599     0.003 15652101806   46484473  15698586279   350.90   8.25  -12.14   8.37
642   3424     0.002  2449211708    3821308   2453033016   468.20   6.43    7.21   5.44
643   4332     0.012 17733265787  213644280  17946910067   519.46   7.09   20.35   3.94
702   4966     0.011   386283454    4392472    390675927   557.36   7.93   20.46   7.88
703   3497     0.019   430162207    8197697    438359904   524.55   3.38   21.14   4.46
705   3109     0.006   164956001    1033535    165989537   526.94   3.69   13.15   3.66
710   8146     0.005 14728343031   75756385  14804099416   233.79   9.19   18.87   6.69
764   5732     0.001  3916722436    2435093   3919157529   480.78   4.59    3.37   4.27
788   5051     0.036   601078470   22724779    623803249   417.02   3.33   25.40   3.42
792   7841     0.000  3936191434    1463639   3937655073   431.21   4.81    3.00   2.87
807   4023     0.000   285998284      37572    286035856   457.70   6.03    0.77   4.59
840   9072     0.010 31378001857  306803783  31684805640   505.40   4.59   19.00   4.13
926   2959     0.031  4417025590  143197546   4560223136   521.97   6.20   32.15   6.63
956   5259     0.017   310951913    5520141    316472055   525.75   4.61   18.13  10.34
```

In this example the variable REGSEX is created by subtracting one from the variable ITSEX. As a result the girls receive a code of 0 and the boys receive a code of 1 on this variable. In this particular model the variable REGSEX is used to predict the plausible values of science achievement by country. In the listing of the results we can see that there is one entry or line for each value of the variable IDCNTRY after selecting only those cases with IDGRADER=2. The first column has the country code and the second column has the number of students in each sample. Because of the way in which REGSEX is coded, the constant (B00) is the mean science achievement for the girls, the error of the constant (B00.SE) is the standard error of this estimate. The first regression coefficient (B01) is the difference in the average value in science achievement between the boys and the girls, and the standard error of this coefficient (B01.SE) is the combined jackknifed sampling error and imputation error of this difference.

For example from the first line in the output we can say that Australia (IDCNTRY = 36) data was available for 4032 cases. The squared multiple correlation between gender and science achievement was 0.011 with girls achieving on average 531.54 and boys scoring on average 17.80 points higher. This difference is not statistically significant (17.80 / 6.85 = 2.6).

10.6 Replicating Analyses from the TIMSS 1999 International Database Reports: Student Level

Many analyses of the TIMSS data can be undertaken using student-level data. We have already presented some examples in the previous sections when explaining how to use the macros provided with the data files. We now proceed to work through additional examples of actual analyses from the TIMSS 1999 international reports, where all the steps are undertaken, including the invocation of the corresponding SAS macro.

Example Analysis with Student-Level Variables Not Using Plausible Values

In the first example, we want to replicate the analysis of eighth graders' reports on the number of hours studied each day. These results , originally presented in Exhibit R1.11 of the TIMSS 1999 International Science Report, are reproduced in Exhibit 9.7. Since the results in this table are not based on plausible values and we want to report the average value of the variable BSDGSTDT, we need to use the macro JACKGEN.

Exhibit 10.7 Sample Exhibit for Student-Level Analysis NOT using Plausible Values Taken From the TIMSS 1999 International Science Report

Country	Average Hours Spent Each Day Studying or Doing Homework[1]				Percentage of Students Reporting Spending Some Time Studying All Three Subjects: Science, Mathematics, and Other
	Science	Mathematics	Other School Subjects	Total	
Australia	0.6 (0.02)	0.7 (0.02)	0.8 (0.02)	2.0 (0.04)	74 (1.6)
Belgium (Flemish)	0.8 (0.03)	1.1 (0.03)	1.4 (0.04)	2.9 (0.05)	86 (1.2)
Bulgaria	1.1 (0.03)	1.1 (0.04)	1.3 (0.04)	3.0 (0.06)	74 (1.9)
Canada	0.6 (0.01)	0.8 (0.02)	1.0 (0.02)	2.2 (0.04)	78 (1.0)
Chile	0.9 (0.02)	0.9 (0.02)	1.2 (0.03)	2.4 (0.04)	75 (1.0)
Chinese Taipei	0.6 (0.02)	0.7 (0.02)	1.0 (0.02)	2.0 (0.05)	55 (1.3)
Cyprus	0.7 (0.02)	1.1 (0.03)	1.5 (0.03)	2.8 (0.04)	79 (0.8)
Czech Republic	0.6 (0.02)	0.7 (0.02)	0.7 (0.02)	1.9 (0.04)	74 (1.4)
England	– –	– –	– –	– –	– –
Finland	0.5 (0.01)	0.6 (0.01)	0.7 (0.01)	1.8 (0.02)	90 (0.8)
Hong Kong, SAR	0.5 (0.01)	0.7 (0.02)	0.7 (0.02)	1.6 (0.04)	53 (1.3)
Hungary	1.1 (0.02)	0.8 (0.02)	1.2 (0.03)	2.8 (0.04)	90 (0.8)
Indonesia	1.1 (0.02)	1.2 (0.03)	1.3 (0.02)	3.0 (0.05)	83 (1.0)
Iran, Islamic Rep.	1.6 (0.03)	1.9 (0.03)	2.0 (0.04)	r 4.0 (0.05)	92 (0.5)
Israel	0.8 (0.02)	1.1 (0.03)	1.4 (0.04)	2.7 (0.05)	79 (0.9)
Italy	1.0 (0.02)	1.3 (0.03)	1.9 (0.03)	3.6 (0.04)	91 (0.8)
Japan	0.4 (0.01)	0.6 (0.01)	0.8 (0.02)	1.7 (0.04)	59 (1.4)
Jordan	1.5 (0.03)	1.7 (0.03)	2.4 (0.05)	r 3.7 (0.06)	r 87 (0.9)
Korea, Rep. of	0.4 (0.01)	0.6 (0.02)	0.7 (0.02)	1.6 (0.03)	50 (0.9)
Latvia (LSS)	0.8 (0.02)	1.0 (0.02)	1.5 (0.03)	3.0 (0.04)	89 (0.7)
Lithuania ‡	0.8 (0.02)	0.9 (0.03)	1.5 (0.04)	2.8 (0.04)	89 (1.0)
Macedonia, Rep. of	2.0 (0.05)	1.2 (0.03)	1.5 (0.04)	r 3.4 (0.05)	90 (0.5)
Malaysia	1.3 (0.02)	1.6 (0.02)	1.8 (0.03)	3.8 (0.04)	94 (0.4)
Moldova	1.7 (0.04)	1.1 (0.03)	1.4 (0.04)	r 3.3 (0.05)	83 (0.8)
Morocco	r 1.5 (0.06)	r 1.7 (0.07)	r 1.8 (0.06)	s 3.1 (0.05)	s 77 (1.3)
Netherlands	0.6 (0.02)	0.6 (0.02)	1.0 (0.02)	2.2 (0.04)	89 (1.1)
New Zealand	0.6 (0.02)	0.7 (0.02)	0.9 (0.02)	2.0 (0.04)	76 (1.3)
Philippines	1.7 (0.04)	1.7 (0.04)	2.1 (0.04)	r 3.3 (0.04)	88 (0.7)
Romania	1.2 (0.03)	1.6 (0.05)	1.4 (0.04)	3.4 (0.06)	77 (1.2)
Russian Federation	1.5 (0.03)	1.1 (0.03)	1.2 (0.04)	3.1 (0.05)	89 (0.7)
Singapore	1.2 (0.02)	1.3 (0.02)	1.7 (0.03)	3.5 (0.04)	90 (0.8)
Slovak Republic	0.8 (0.02)	0.8 (0.02)	0.9 (0.02)	2.3 (0.03)	88 (0.8)
Slovenia	0.9 (0.02)	0.8 (0.02)	0.9 (0.02)	2.5 (0.03)	85 (1.0)
South Africa	1.5 (0.05)	1.8 (0.04)	2.0 (0.06)	r 3.1 (0.06)	71 (1.9)
Thailand	1.0 (0.02)	1.1 (0.02)	1.2 (0.02)	2.9 (0.04)	88 (0.6)
Tunisia	1.2 (0.03)	1.8 (0.03)	2.1 (0.03)	r 3.6 (0.04)	82 (0.8)
Turkey	1.2 (0.02)	1.2 (0.02)	1.9 (0.03)	3.5 (0.05)	90 (0.7)
United States	0.6 (0.01)	0.8 (0.02)	0.9 (0.02)	2.1 (0.04)	72 (1.6)
International Avg.	1.0 (0.00)	1.1 (0.00)	1.3 (0.01)	2.8 (0.01)	80 (0.2)

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1998-1999.

Background data provided by students.

1 Average hours based on: No time=0; less than 1 hour=.5; 1-2 hours=1.5; 3-5 hours=4; more than 5 hours=7.

‡ Lithuania tested the same cohort of students as other countries, but later in 1999, at the beginning of the next school year.

( ) Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

A dash (–) indicates data are not available.

An "r" indicates a 70-84% student response rate. An "s" indicates a 50-69% student response rate.

To replicate the results in this table, we need to undertake several steps. After reviewing the codebooks and the questionnaire information we find out the students were asked three questions about the number of hours they spend studying mathematics, science, and other subjects. The data collected from these variables are summarized in the variable BSDGSTDT, and this variable is found in the Student Background data file. 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 changes were made we can continue with our analysis without any modifications.

We then proceed to read, from the SAS dataset that contains this variable, the sampling weight for the student (TOTWGT), the variables that contain the jackknife replication information (JKZONE and JKREP), the variable that will be used to select the eighth graders from the data file (IDGRADER), and the variable containing the country identification code (IDCNTRY). In this analysis we will use the data for all countries in the database, although the exact same steps need to be taken if you want to examine these variables within a single country or for a select group of countries.

The SAS code is presented in Exhibit 10.8 and is included in the CD under the name EXAMPLE1.SAS. Selections of the results obtained from this program are displayed in Exhibit 10.9. We have included as part of the program the corresponding value labels and format statements so that the different categories or groups are labeled appropriately.

Notice that one of the steps in this program is to select only those students in the eighth grade who have non-missing data in the variable.

Note also in this analysis that we have used the data from all the countries, although we did select from the dataset only those variables that are relevant to our analysis. In general, this type of analysis is quite feasible with a powerful desktop computer, however, you need to keep in mind that computing and storage requirements for these types of analysis are quite demanding.

In general, to perform analyses such as using the Student Background data files, you need to do the following:

- Identify the variable or variables of interest in the student file and find out about any specific national adaptations to the variable.
- Retrieve the relevant variables from the data files, including the sampling weights, JRR replication information, and any other variables used in the selection of cases.
- Use the macro JACKGEN with the corresponding arguments and parameters.
- Print out the result file.

Exhibit 10.8 SAS Control Statements for Performing Analyses with Student-Level Variables NOT Using Plausible Values (EXAMPLE1.SAS)

```
libname bm2 "d:\timss\data\bm2";

%include "d:\Timss\Programs\jackgen.sas";

data student;
 set bm2.bsgallm2;
  where nmiss(bsdgstdt) = 0 and intms99 = 1 and idgrader = 2;

* Define the format for the variables used;
proc format library=work;
 value country
 036= 'AUSTRALIA         '   956= 'BELGIUM (Flemish) '   100= 'BULGARIA          '
 124= 'CANADA            '   152= 'CHILE             '   196= 'CYPRUS            '
 203= 'CZECH REP.        '   926= 'ENGLAND           '   246= 'FINLAND           '
 344= 'HONG KONG         '   348= 'HUNGARY           '   360= 'INDONESIA         '
 364= 'IRAN, ISLAMIC REP.'   376= 'ISRAEL            '   380= 'ITALY             '
 400= 'JORDAN            '   392= 'JAPAN             '   410= 'KOREA, REP. OF    '
 440= 'LITHUANIA         '   428= 'LATVIA            '   504= 'MOROCCO           '
 498= 'MOLDOVA, REP. OF  '   807= 'MACEDONIA, REP. OF'   458= 'MALAYSIA          '
 528= 'NETHERLANDS       '   554= 'NEW ZEALAND       '   608= 'PHILIPPINES       '
 642= 'ROMANIA           '   643= 'RUSSIAN FEDERATION'   702= 'SINGAPORE         '
 703= 'SLOVAK REP.       '   705= 'SLOVENIA          '   764= 'THAILAND          '
 788= 'TUNISIA           '   792= 'TURKEY            '   158= 'CHINESE TAIPEI    '
 840= 'UNITED STATES     '   710= 'SOUTH AFRICA      ';

%jackgen (totwgt , jkzone, jkrep, 75, idcntry, bsdgstdt, student);

proc print data=final noobs;
     by idcntry ;
     var N totwgt mnx mnx_se pct pct_se;
     format idcntry country.
            n 6.0 totwgt 10.0 mnx pct mnx_se pct_se 6.2;

run;
```

Exhibit 10.9 Extract of SAS Computer Output for Performing Analyses with Student-Level Variables Not Using Plausible Values (EXAMPLE 1)

```
*COUNTRY ID*=AUSTRALIA
     N        TOTWGT       mnx    mnx_se       pct    pct_se
  3841        247322      1.96      0.04      1.56      0.04

*COUNTRY ID*=BULGARIA
     N        TOTWGT       mnx    mnx_se       pct    pct_se
  2794         75854      3.00      0.06      0.48      0.04

*COUNTRY ID*=CANADA
     N        TOTWGT       mnx    mnx_se       pct    pct_se
  8376        354966      2.22      0.04      2.24      0.05
.
.
.
*COUNTRY ID*=UNITED STATES
     N        TOTWGT       mnx    mnx_se       pct    pct_se
  8430       3131090      2.08      0.04     19.76      0.86

*COUNTRY ID*=BELGIUM (Flemish)
     N        TOTWGT       mnx    mnx_se       pct    pct_se
  4861         60145      2.95      0.05      0.38      0.01
```

In this example, each country's mean value for BSDGSTDT is reported for the eighth grade. The results are presented by country after selecting only those cases with IDGRADER=2. The country is in the first column. The second column has the number of students sampled in the country. The third column corresponds to the weights given these sampled students in the whole population, followed by their mean for BSDGSTDT and the corresponding standard errors. The last two columns represent the percentages of students sampled responding and the corresponding standard error.

For example, from the first line of the report, we can say that in Australia data was available for 3841 cases. These students represent a population of 247322 students. Australian students spend, on average, 1.96 hours studying each day. The standard error of this mean is .04.

Example Analysis with Student-Level Variables Using Plausible Values

In this example, we want to replicate another one of the results presented in the international report. We are interested in looking at the eighth graders' reports on the number of books in their home and their achievement in science. These are the results that are presented in Exhibit 10.10 and appear in Exhibit R1.3 of the TIMSS 1999 International Science Report. Since the results in this table are based on plausible values, we need to use the macro JACKPV.

Exhibit 10.10 Sample Exhibit for Student-Level Analysis Involving Plausible Values Taken From the TIMSS 1999 International Science Report

	Three or More Bookcases (More Than 200 Books)		About Two Bookcases (101-200 Books)		About One Bookcase (26-100 Books)		About One Shelf (11-25 Books)		None or Very Few (0-10 Books)	
	Percent of Students	Average Achievement	Percent of Students	Average Achievement	Percent of Students	Average Achievement	Percent of Students	Average Achievement	Percent of Students	Average Achievement
Australia	38 (1.2)	565 (4.7)	27 (0.9)	544 (4.8)	24 (0.9)	521 (6.0)	7 (0.6)	495 (9.2)	3 (0.4)	483 (15.0)
Belgium (Flemish)	14 (0.8)	561 (4.9)	14 (0.6)	557 (5.7)	31 (1.3)	542 (4.2)	21 (0.7)	522 (5.8)	19 (1.3)	504 (5.4)
Bulgaria	37 (2.2)	548 (7.6)	19 (0.9)	529 (5.6)	24 (1.1)	509 (5.2)	12 (1.0)	477 (8.1)	9 (1.3)	450 (9.1)
Canada	31 (0.9)	553 (4.0)	24 (0.8)	541 (3.3)	28 (0.7)	527 (3.4)	11 (0.5)	498 (5.1)	5 (0.3)	495 (12.4)
Chile	9 (0.6)	468 (10.4)	11 (0.6)	457 (8.1)	28 (0.9)	441 (5.3)	32 (0.9)	403 (4.0)	20 (1.1)	378 (4.7)
Chinese Taipei	16 (0.8)	616 (6.1)	12 (0.5)	603 (7.3)	31 (0.7)	579 (6.0)	23 (0.7)	554 (4.7)	17 (0.9)	507 (4.4)
Cyprus	17 (0.8)	479 (4.5)	23 (0.9)	476 (3.7)	36 (1.0)	465 (4.0)	19 (0.9)	437 (4.9)	5 (0.5)	402 (10.0)
Czech Republic	28 (1.4)	565 (5.5)	30 (1.4)	548 (5.8)	34 (1.1)	523 (4.8)	7 (0.8)	493 (7.6)	1 (0.2)	~ ~
England	26 (1.2)	593 (6.7)	23 (0.8)	550 (7.3)	32 (1.1)	526 (4.7)	13 (0.8)	483 (6.4)	7 (0.7)	472 (11.5)
Finland	22 (1.1)	552 (5.7)	22 (0.9)	544 (4.7)	39 (1.1)	536 (4.4)	14 (0.8)	506 (5.3)	4 (0.4)	495 (11.6)
Hong Kong, SAR	8 (0.5)	548 (6.4)	10 (0.5)	534 (6.0)	27 (0.7)	537 (4.9)	27 (0.7)	530 (4.8)	28 (0.9)	517 (4.2)
Hungary	38 (1.4)	581 (3.7)	21 (0.8)	561 (5.6)	25 (1.0)	539 (4.4)	12 (0.8)	502 (8.6)	3 (0.5)	460 (14.4)
Indonesia	5 (0.5)	437 (17.8)	5 (0.4)	451 (14.1)	26 (0.9)	446 (4.6)	39 (1.0)	429 (5.5)	26 (1.3)	438 (5.5)
Iran, Islamic Rep.	9 (0.7)	482 (8.3)	8 (0.7)	479 (8.8)	22 (0.6)	465 (6.1)	32 (0.8)	443 (4.7)	29 (1.4)	425 (4.9)
Israel	22 (1.3)	498 (5.3)	21 (0.8)	504 (5.7)	33 (1.0)	471 (6.1)	18 (0.9)	434 (7.0)	6 (0.5)	388 (15.8)
Italy	20 (0.9)	523 (7.5)	15 (0.7)	518 (5.9)	28 (0.9)	497 (4.4)	25 (0.9)	471 (5.6)	12 (0.8)	453 (7.2)
Japan	18 (0.7)	577 (5.3)	18 (0.6)	567 (5.0)	31 (0.7)	548 (2.6)	19 (0.6)	541 (4.6)	14 (0.6)	518 (5.1)
Jordan	10 (0.6)	485 (6.7)	10 (0.6)	493 (6.5)	28 (0.8)	471 (4.8)	31 (1.1)	447 (4.3)	21 (1.2)	416 (5.8)
Korea, Rep. of	20 (0.8)	589 (3.8)	23 (0.6)	562 (4.6)	36 (0.7)	544 (2.1)	10 (0.5)	510 (4.9)	10 (0.4)	490 (5.6)
Latvia (LSS)	47 (1.4)	522 (6.2)	25 (0.9)	499 (4.5)	21 (0.9)	480 (4.6)	6 (0.7)	465 (10.6)	1 (0.2)	~ ~
Lithuania ‡	17 (1.0)	528 (7.1)	21 (1.0)	517 (6.2)	36 (1.2)	482 (5.2)	20 (1.1)	454 (6.1)	7 (0.8)	433 (10.1)
Macedonia, Rep. of	7 (0.7)	492 (8.0)	9 (0.6)	498 (5.9)	30 (1.2)	487 (5.4)	38 (1.3)	449 (6.6)	15 (1.3)	408 (9.8)
Malaysia	9 (0.7)	538 (8.1)	12 (0.6)	521 (8.3)	32 (0.9)	505 (6.3)	34 (0.9)	472 (4.7)	13 (0.8)	460 (6.6)
Moldova	9 (0.7)	486 (7.8)	11 (0.8)	489 (7.5)	28 (1.0)	470 (4.9)	33 (1.2)	456 (5.1)	20 (1.1)	427 (7.4)
Morocco	3 (0.3)	343 (13.9)	5 (0.4)	372 (20.0)	20 (1.0)	335 (11.4)	35 (0.9)	323 (5.9)	37 (1.7)	320 (5.7)
Netherlands	24 (1.8)	575 (9.3)	23 (1.2)	554 (6.9)	31 (1.1)	546 (7.2)	15 (1.4)	508 (12.1)	8 (1.4)	499 (12.3)
New Zealand	33 (1.1)	541 (5.7)	24 (0.8)	529 (5.3)	27 (0.8)	499 (5.3)	10 (0.7)	460 (8.2)	6 (0.5)	407 (12.1)
Philippines	5 (0.4)	355 (16.5)	5 (0.4)	388 (17.3)	15 (0.8)	393 (11.9)	33 (0.9)	350 (8.0)	41 (1.4)	323 (7.0)
Romania	15 (1.3)	515 (8.1)	15 (0.9)	510 (8.3)	32 (1.1)	474 (6.0)	24 (1.6)	447 (6.8)	14 (1.1)	434 (9.3)
Russian Federation	23 (1.5)	555 (6.0)	29 (1.1)	541 (7.1)	31 (1.3)	521 (7.7)	13 (1.0)	495 (8.9)	4 (0.5)	470 (20.8)
Singapore	12 (0.6)	599 (11.4)	14 (0.7)	599 (10.3)	40 (1.1)	579 (7.2)	22 (1.0)	540 (8.8)	12 (0.8)	516 (8.8)
Slovak Republic	17 (0.9)	566 (6.6)	24 (1.0)	550 (5.6)	43 (1.1)	531 (3.9)	14 (1.0)	494 (6.4)	2 (0.4)	~ ~
Slovenia	14 (0.9)	564 (5.6)	20 (0.8)	556 (5.6)	46 (1.0)	532 (3.3)	16 (0.8)	498 (6.4)	4 (0.4)	483 (11.1)
South Africa	7 (0.5)	293 (25.6)	6 (0.5)	337 (18.9)	14 (0.8)	308 (15.3)	31 (0.8)	245 (8.7)	43 (1.6)	208 (5.9)
Thailand	6 (0.5)	509 (7.0)	8 (0.5)	502 (6.6)	27 (0.9)	493 (4.5)	37 (1.1)	477 (4.1)	22 (1.0)	465 (5.4)
Tunisia	9 (0.6)	438 (7.2)	9 (0.6)	444 (6.4)	25 (0.7)	438 (4.7)	36 (0.9)	424 (5.2)	21 (1.1)	422 (3.4)
Turkey	6 (0.4)	441 (11.1)	8 (0.5)	459 (7.4)	28 (0.8)	449 (5.3)	37 (0.8)	426 (5.0)	22 (1.1)	415 (5.2)
United States	28 (1.2)	557 (4.5)	22 (0.6)	538 (4.6)	29 (0.8)	508 (3.9)	14 (0.7)	468 (7.0)	8 (0.6)	442 (6.0)
International Avg.	18 (0.2)	517 (1.6)	16 (0.1)	511 (1.2)	29 (0.2)	493 (1.0)	22 (0.1)	464 (1.0)	14 (0.2)	441 (1.5)

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1998-1999.

Background data provided by students.

‡ Lithuania tested the same cohort of students as other countries, but later in 1999, at the beginning of the next school year.

( ) Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

A tilde (~) indicates insufficient data to report achievement.

To replicate the results in this table, we need to undertake several steps. After reviewing the codebooks and the questionnaire information we find out if the students were asked a question about the number of books in the home (see Supplement 1 for a copy of the student questionnaires). The data collected from this variable are captured in the variable BSBGBOOK, and this variable is found in the Student Background data file. 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 changes were made we can continue with our analysis without any modifications.

We then proceed to read, from the SAS dataset that contains this variable, the international science achievement plausible values (BSSSCI01-BSSSCI05), the sampling weight for the students (TOTWGT), the variables that contain the jackknife replication information (JKZONE and JKREP), the variable that will be used to select the eighth graders from the data file (IDGRADER), and the variable containing the country identification code (IDCNTRY). In this analysis we will use the data for all countries in the database, although the exact same steps need to be taken if you want to examine these variables within a single country or for a select group of countries.

The SAS code is presented in Exhibit 10.11 Selections of the results obtained from this program are displayed in Exhibit 10.12 We have included as part of the program the corresponding value labels and format statements so that the different categories or groups are labeled appropriately.

Notice that one of the steps in each of this program is to select only those students in the eighth grade who chose one of the five options presented in the question.

Note also in this analysis that we have used the data from all the countries, although we did select from the dataset only those variables that are relevant to our analysis. In general, this type of analysis is quite feasible with a powerful desktop computer, however you need to keep in mind that computing and storage requirements for these types of analysis are quite demanding.

In general, to perform analyses such as those using the Student Background data files, you need to do the following:

- Identify the variable or variables of interest in the student file and find out about any specific national adaptations to the variable.
- Retrieve the relevant variables from the data files, including the achievement score, sampling weights, JRR replication information, and any other variables used in the selection of cases.
- Use the macro JACKPV with the corresponding arguments and parameters.
- Print out the result file.

Exhibit 10.11 SAS Control Statements for Performing Analyses with Student-Level Variables Involving Plausible Values (EXAMPLE2.SAS)

```
libname bm2 "d:\timss\data\bm2";

%include "d:\UserGuide\Programs\jackpv.sas";

data student;
 set bm2.bsgallm2;
  where bsbgbook in(1,2,3,4,5) and intms99 = 1 and idgrader = 2;

* Define the format for the variables used;
proc format library=work;
 value books 1 = "None or very few (0-10)"
             2 = "For 1 shelf (11-25)"
             3 = "For 1 bookcase (26-100)"
             4 = "For 2 bookcases (101-200)"
             5 = "For 3 or more bookcases (>200)";

 value country
 036= 'AUSTRALIA         '   956= 'BELGIUM (Flemish) '   100= 'BULGARIA          '
 124= 'CANADA            '   152= 'CHILE             '   196= 'CYPRUS            '
 203= 'CZECH REP.        '   926= 'ENGLAND           '   246= 'FINLAND           '
 344= 'HONG KONG         '   348= 'HUNGARY           '   360= 'INDONESIA         '
 364= 'IRAN, ISLAMIC REP.'   376= 'ISRAEL            '   380= 'ITALY             '
 400= 'JORDAN            '   392= 'JAPAN             '   410= 'KOREA, REP. OF    '
 440= 'LITHUANIA         '   428= 'LATVIA            '   504= 'MOROCCO           '
 498= 'MOLDOVA, REP. OF  '   807= 'MACEDONIA, REP. OF'   458= 'MALAYSIA          '
 528= 'NETHERLANDS       '   554= 'NEW ZEALAND       '   608= 'PHILIPPINES       '
 642= 'ROMANIA           '   643= 'RUSSIAN FEDERATION'   702= 'SINGAPORE         '
 703= 'SLOVAK REP.       '   705= 'SLOVENIA          '   764= 'THAILAND          '
 788= 'TUNISIA           '   792= 'TURKEY            '   158= 'CHINESE TAIPEI    '
 840= 'UNITED STATES     '   710= 'SOUTH AFRICA      ';

%jackpv (totwgt , jkzone, jkrep, 75, idcntry bsbgbook, bsssci0, 5, student);

proc print data=final noobs;
     by idcntry ;
     var bsbgbook N totwgt mnpv mnpv_se pct pct_se;
     format idcntry country. bsbgbook books.
            N 6.0 totwgt 10.0 mnpv mnpv_se pct pct_se 6.2;

run;
```

Exhibit 10.12 Extract of SAS Computer Output for Performing Analyses with Student-Level Variables Involving Plausible Values (EXAMPLE 2)

```
*COUNTRY ID*=AUSTRALIA
BSBGBOOK                              N        TOTWGT      mnpv    mnpv_se       pct     pct_se
None or very few (0-10)             130          8731    482.87      15.00      3.47       0.40
For 1 shelf (11-25)                 266         18361    495.35       9.24      7.29       0.61
For 1 bookcase (26-100)             927         61382    521.47       5.98     24.36       0.95
For 2 bookcases (101-200)          1040         67799    544.46       4.81     26.91       0.86
For 3 or more bookcases (>200)     1545         95691    565.26       4.74     37.98       1.22

*COUNTRY ID*=BULGARIA
BSBGBOOK                              N        TOTWGT      mnpv    mnpv_se       pct     pct_se
None or very few (0-10)             279          7627    449.82       9.07      8.71       1.27
For 1 shelf (11-25)                 371         10178    476.67       8.05     11.62       1.02
For 1 bookcase (26-100)             772         20888    509.11       5.25     23.84       1.12
For 2 bookcases (101-200)           636         16817    529.05       5.56     19.19       0.87
For 3 or more bookcases (>200)     1184         32107    548.34       7.59     36.65       2.20

*COUNTRY ID*=CANADA
BSBGBOOK                              N        TOTWGT      mnpv    mnpv_se       pct     pct_se
None or very few (0-10)             446         17772    495.07      12.42      4.84       0.32
For 1 shelf (11-25)                1044         42011    497.64       5.10     11.45       0.50
For 1 bookcase (26-100)            2537        103001    526.87       3.37     28.07       0.71
For 2 bookcases (101-200)          2121         89829    540.97       3.34     24.48       0.77
For 3 or more bookcases (>200)     2526        114301    552.68       4.04     31.15       0.89

.
.
.
.
.

*COUNTRY ID*=UNITED STATES
BSBGBOOK                              N        TOTWGT      mnpv    mnpv_se       pct     pct_se
None or very few (0-10)             774        263683    441.63       6.01      8.11       0.59
For 1 shelf (11-25)                1287        448473    467.99       7.05     13.79       0.71
For 1 bookcase (26-100)            2502        927299    508.02       3.92     28.51       0.83
For 2 bookcases (101-200)          1877        700350    538.25       4.58     21.54       0.56
For 3 or more bookcases (>200)     2344        912338    556.69       4.53     28.05       1.21

*COUNTRY ID*=ENGLAND
BSBGBOOK                              N        TOTWGT      mnpv    mnpv_se       pct     pct_se
None or very few (0-10)             182         34823    472.16      11.54      6.60       0.71
For 1 shelf (11-25)                 350         66603    483.40       6.43     12.63       0.79
For 1 bookcase (26-100)             900        167841    525.71       4.69     31.82       1.07
For 2 bookcases (101-200)           641        120040    549.55       7.25     22.76       0.82
For 3 or more bookcases (>200)      760        138215    593.14       6.72     26.20       1.20

*COUNTRY ID*=BELGIUM (Flemish)
BSBGBOOK                              N        TOTWGT      mnpv    mnpv_se       pct     pct_se
None or very few (0-10)             813         12339    504.02       5.42     19.14       1.27
For 1 shelf (11-25)                1096         13823    522.35       5.78     21.44       0.73
For 1 bookcase (26-100)            1676         20042    542.45       4.15     31.09       1.30
For 2 bookcases (101-200)           813          9063    557.00       5.65     14.06       0.62
For 3 or more bookcases (>200)      798          9196    560.75       4.94     14.26       0.79
```

In this example, each country's mean plausible value for science achievement is reported for each response category in the variable BSBGBOOK. The results are presented by country for each value of the variable IDCNTRY after selecting only those cases with IDGRADER=2. The country and the five response options are presented in the first column. The second column has the number of students sampled in each category. The third column corresponds to the weights given these sampled students in the whole population, followed by their mean plausible values for science achievement and the corresponding standard errors. The last two columns represent the percentages of students sampled responding within each category and the corresponding standard errors.

For example, from the first line of the report, we can say that in Australia the 1545 students (38% of the sampled students) who reported having more than 200 books had a mean science achievement of 565.26 with a standard error of 4.74; while the 130 students (3.5% of the sampled students) who reported having fewer than 10 books had a mean science achievement of 482.87 with a standard error of 15.00.

10.7 Performing Analyses with Teacher-Level Variables

When analyzing the teacher data, it is first necessary to link the students with their corresponding teachers. Each student record in the Student Background data file can contain a link to as many as six different teachers in the Teacher Background data file. To facilitate the linking between students and their teachers in the teacher file, the Student-Teacher Linkage file was created and is part of the International Database. This file is called BST<COUNTRY>M1 and BST<COUNTRY>M2 in 1995 ad 1999, respectively. The Student-Teacher Linkage file contains one record for each student-by-teacher combination, with the corresponding identification variables.

Each record also contains the number of mathematics and science teachers for the student and a set of weights that can be used when conducting analyses with these data. Student achievement plausible values, sampling weights, and JRR replication information have been added to the Student-Teacher Linkage file in order to simplify the merging process for analyses that link teacher variables to student achievement. For such analyses it is necessary to merge only the Teacher Background file with the Student-Teacher Linkage file. For analyses linking teacher variables to other student variables, it is necessary also to merge the Student Background files with the Teacher Background file after it has been combined with the Student-Teacher Linkage file.

Conducting analyses with the teacher data requires some extra steps that are not required when analyzing the student or school background data.

For our example, we want to find out about the age of the science teachers that teach the eighth-grade students in each of the TIMSS countries. In particular, we want to find out what percentage of eighth-grade students are taught by teachers of a specific group based on the teachers' ages (BTBGAGE), and the mean achievement in science of the students taught by those teachers. These results are located in the Science Teacher Background Data Almanac (bsalm7_m2.*) for questionnaire item TQS2A-01 as shown in Exhibit 10.13.

Exhibit 10.13 Sample Data Almanac Sheet for Teacher-Level Analysis using Plausible Values Taken From the TIMSS 1999 International Science Report

Third International Mathematics and Science Study - 1999 Assessment 16:43 Sunday, February 18, 2001 4
Science Teacher Background Data Almanac - 8th Grade

Question : Teacher age
Location : TQS2A-01 (BTBGAGE)

Country	Sample	Valid N	1.UNDER 25 Mean	2.25-29 Mean	3.30-39 Mean	4.40-49 Mean	5.50-59 Mean	6.60 OR MORE Mean	NOT ADMINIS TERED Mean	OMIT Mean
Australia	543	465	536.9	533.6	538.9	533.4	553.1	548.2	548.1	597.0
Belgium (Flemish)	409	372	537.7	540.8	526.5	534.2	536.4	477.1	522.5	540.8
Bulgaria	593	571	520.0	505.5	516.8	505.5	518.3	488.2	557.9	484.6
Canada	474	439	533.7	534.4	532.2	538.2	527.1	543.2	525.1	529.1
Chile	185	180	444.8	461.5	438.6	421.4	397.1	407.9	391.4	402.2
Chinese Taipei	150	147	579.2	567.2	568.7	572.8	570.4	529.6	590.7	.
Cyprus	252	222	.	424.2	452.4	465.4	463.2	.	445.0	490.8
Czech Republic	568	546	515.6	551.2	535.9	545.5	537.2	547.5	517.9	625.6
England	537	359	512.9	543.6	546.7	545.6	561.0	425.5	525.2	565.8
Finland	434	418	538.1	532.6	536.8	537.0	531.9	542.5	544.8	543.7
Hong Kong, SAR	137	135	481.5	538.1	537.2	522.5	513.1	546.7	501.6	.
Hungary	592	581	561.1	554.9	553.8	552.2	548.4	533.6	583.5	.
Indonesia	302	297	398.0	436.0	430.2	442.8	474.1	.	387.9	432.7
Iran, Islamic Rep.	170	168	427.3	435.8	452.6	456.3	447.1	.	.	480.5
Israel	259	235	428.1	470.8	460.9	474.3	485.0	505.7	466.7	481.2
Italy	183	183	.	.	493.9	493.9	491.7	517.6	.	.
Japan	144	144	562.0	540.4	549.2	545.3	556.1	586.9	.	.
Jordan	147	146	413.9	445.8	454.1	465.9	434.4	.	449.6	.
Korea, Rep. of	186	186	528.3	556.4	547.7	552.0	541.4	539.8	.	.
Latvia (LSS)	433	402	507.2	493.4	502.7	502.8	507.3	508.6	487.6	.
Lithuania	450	442	474.2	484.2	488.0	494.1	480.1	497.4	489.2	.
Macedonia, Rep. of	596	593	468.6	461.3	448.8	442.0	469.8	479.9	481.0	478.5
Malaysia	150	150	475.9	472.8	494.9	509.0	505.0	.	.	.
Moldova	600	577	466.9	463.9	465.4	452.8	457.8	476.8	442.4	424.0
Morocco	687	606	326.0	317.4	325.3	321.3	.	.	324.7	306.0
Netherlands	351	305	535.3	546.1	544.7	550.5	538.1	625.7	540.4	.
New Zealand	250	234	492.6	500.4	503.3	525.2	502.3	546.8	472.9	489.2
Philippines	150	149	395.9	341.4	335.2	352.4	322.5	424.0	394.3	.
Romania	588	564	438.1	456.8	470.0	470.7	481.3	488.1	485.3	479.1
Russian Federation	756	755	513.8	532.8	530.4	526.8	534.4	524.1	.	563.2
Singapore	145	144	604.9	576.6	572.5	559.2	559.8	494.6	.	501.4
Slovak Republic	578	179	511.1	511.5	547.2	531.5	537.7	534.6	535.7	.
Slovenia	447	216	495.5	519.3	537.9	533.3	534.6	512.3	533.1	.
South Africa	190	185	274.4	259.3	220.8	292.3	445.7	177.8	207.0	.
Thailand	150	148	472.0	471.2	478.0	485.7	507.6	517.8	503.7	.
Tunisia	149	148	424.4	416.9	430.1	445.3	430.3	.	402.9	.
Turkey	204	203	414.2	417.9	424.3	448.8	439.0	.	.	400.4
United States	1079	904	536.3	528.6	501.2	516.8	524.3	532.1	491.7	485.8
International Avg.	374	334	481.9	485.0	486.7	491.2	499.0	502.7	478.3	490.5

The TIMSS teacher files do not contain representative samples of teachers within a country. Rather, these are the teachers for a representative sample of students within a country. Therefore, it is appropriate that statements about the teachers be made only in terms of how many students are taught by teachers of one kind or another, and not in terms of how many teachers in the country do one thing or another.

As before, we first proceed to identify the variables relevant to the analysis in the corresponding files, and review the documentation on the specific national adaptations to the questions of interest (Supplements 1 and 2). Since we are using teacher-level variables we need to look into the teacher file and the Student-Teacher Linkage files to find the variables. From the science teacher file we extract the variable that contains the information on the science teachers' age (BTBGAGE), the variable that identifies the country (IDCNTRY) and the two variables that will allow us to link the teacher information to the student data (IDTEACH and IDLINK).

There is one teacher file for the mathematics teachers and a second teacher file for the science teachers. If you want to look only at mathematics teachers, then you will need to use the mathematics teacher file (BTM<COUNTRY>M2); if the interest is in the science teachers then you will need to use the science teacher file

(BTS<COUNTRY>M2); but if the interest is in the mathematics and science teachers combined, both these files need to be combined by appending or adding one file to the other. In doing so it is important to keep in mind that although there are variables in common between these two files, most of them are not.

In our example, our teacher variable of interest (age) is a categorical variable with six categories. However, we want to categorize the teachers into 4 groups (under 29 years old, 30-39 years old, 40-49 years old, and 50 years or over). While reading the teacher file we use commands in SAS to collapse the different values into four categories and we label them accordingly. We then proceed to read the necessary information from the Student-Teacher Linkage file. From this file we keep the country identification (IDCNTRY) and the two variables that will allow us to link the student information to the teacher data (IDTEACH and IDLINK). We also keep the variable that indicates the grade for the student (IDGRADER), the science achievement plausible values (BSSSCI01-BSSSCI05), and the information necessary to compute the replicate weights for estimating the JRR standard error. We select the variable that will give us the correct weight for science teacher variable (SCIWGT). If you are interested in looking at the mathematics teachers, then the weight variable that should be selected is MATWGT; if you are interested in analyzing both mathematics and science teachers combined, the weight variable TCHWGT should be selected.

The two files are then merged or matched into one file that will then be used with the JACKPV macro. These two files will be merged using the variables IDCNTRY, IDTEACH, and IDLINK. The combination of values for these three variables is unique within the teacher data, but is repeated in the Student-Teacher Linkage file as many times as the specific teacher teaches students in a class. After the files are merged, the macro JACKPV is used and the results can be printed. The code in SAS for this example is presented in Exhibits 10.14. Selections of the results obtained from this program are displayed in Exhibit 10.15.

Exhibit 10.14 SAS Control Statements for Performing Teacher-Level Analysis using Plausible Values (EXAMPLE3.SAS)

```
libname bm2 "d:\timss\data\bm2";

%include "d:\timss\Programs\jackpv.sas";

* Read the variables from the Science Teacher file and sort by merge variables;
data teacher;
     set bm2.btsallm2 (keep=idcntry idteach idlink btbgage);
    select (Btbgage);
     when(1,2)   btdgage = 1;    /*under 29 yrs. old            */
     when(3)     btdgage = 2;    /*30-39                        */
     when(4)     btdgage = 3;    /*40-49                        */
     when(5,6)   btdgage = 4;    /*50 or over                   */
     otherwise;
    end;

proc sort data=teacher;
     by idcntry idteach idlink;

* Read the variables from the student teacher link file and sort by merge variables;
data studteac;
     set bm2.bstallm2
       (keep=idcntry idteach idlink idgrader jkrep jkzone sciwgt bsssci01-bsssci05 intms99);
proc sort data=studteac;
     by idcntry idteach idlink;

* Now merge the two files;
data merged;
     merge studteac teacher;
     by idcntry idteach idlink;
      if sciwgt > 0 and nmiss(btdgage) = 0 and intms99 = 1 and idgrader = 2;

* Define the format for the variables used;
proc format library=work;
 value age   1 = "Under 29 y.o."
             2 = "30-39 years old"
             3 = "40-49 years old"
             4 = "50 years or older";

 value country
 036= 'AUSTRALIA         '   956= 'BELGIUM (Flemish) '   100= 'BULGARIA          '
 124= 'CANADA            '   152= 'CHILE             '   196= 'CYPRUS            '
 203= 'CZECH REP.        '   926= 'ENGLAND           '   246= 'FINLAND           '
 344= 'HONG KONG         '   348= 'HUNGARY           '   360= 'INDONESIA         '
 364= 'IRAN, ISLAMIC REP.'   376= 'ISRAEL            '   380= 'ITALY             '
 400= 'JORDAN            '   392= 'JAPAN             '   410= 'KOREA, REP. OF    '
 440= 'LITHUANIA         '   428= 'LATVIA            '   504= 'MOROCCO           '
 498= 'MOLDOVA, REP. OF  '   807= 'MACEDONIA, REP. OF'   458= 'MALAYSIA          '
 528= 'NETHERLANDS       '   554= 'NEW ZEALAND       '   608= 'PHILIPPINES       '
 642= 'ROMANIA           '   643= 'RUSSIAN FEDERATION'   702= 'SINGAPORE         '
 703= 'SLOVAK REP.       '   705= 'SLOVENIA          '   764= 'THAILAND          '
 788= 'TUNISIA           '   792= 'TURKEY            '   158= 'CHINESE TAIPEI    '
 840= 'UNITED STATES     '   710= 'SOUTH AFRICA      ';

%jackpv(sciwgt , jkzone, jkrep, 75, idcntry btdgage, bsssci0, 5, merged);

proc print data=final noobs;
     by idcntry ;
     var btdgage N sciwgt mnpv mnpv_se pct pct_se;
     format idcntry country. btdgage age.
            N 6.0 sciwgt 10.0 mnpv mnpv_se pct pct_se 6.2;

run;
```

Exhibit 10.15 Extract of SAS Computer Output Performing Teacher-Level Analysis using Plausible Values (EXAMPLE3)

```
*COUNTRY ID*=AUSTRALIA

     btdgage             N        sciwgt      mnpv    mnpv_se        pct     pct_se
Under 29 y.o.          627         36628    534.66       8.17      15.75       2.66
30-39 years old       1091         72944    538.88       6.56      31.37       3.39
40-49 years old       1287         79488    533.43       7.07      34.19       3.31
50 years or older      647         43444    552.88       9.95      18.69       2.71

*COUNTRY ID*=BULGARIA

     btdgage             N        sciwgt      mnpv    mnpv_se        pct     pct_se
Under 29 y.o.         1197         10070    507.05       8.11      13.24       3.04
30-39 years old       3139         20554    516.75       9.12      27.03       2.80
40-49 years old       3643         24722    505.54       5.06      32.51       2.78
50 years or older     3182         20696    516.18       8.11      27.22       2.83

*COUNTRY ID*=CANADA

     btdgage             N        sciwgt      mnpv    mnpv_se        pct     pct_se
Under 29 y.o.         1739         72340    534.36       6.10      20.95       3.06
30-39 years old       2754        105991    532.24       3.90      30.69       2.55
40-49 years old       2251        106509    538.22       3.70      30.84       2.95
50 years or older     1496         60506    527.21       6.69      17.52       2.59

.
.
.
.
.

*COUNTRY ID*=UNITED STATES

     btdgage             N        sciwgt      mnpv    mnpv_se        pct     pct_se
Under 29 y.o.         1431        563546    530.29       8.06      20.03       2.60
30-39 years old       1484        527177    501.20      10.56      18.74       2.18
40-49 years old       2419        816588    516.78       8.31      29.03       2.82
50 years or older     2340        905665    525.27       6.76      32.20       2.68

*COUNTRY ID*=ENGLAND

     btdgage             N        sciwgt      mnpv    mnpv_se        pct     pct_se
Under 29 y.o.          628         78473    536.69       9.92      23.79       4.02
30-39 years old        687         75922    546.72       9.66      23.02       3.56
40-49 years old       1101        102994    545.57       9.96      31.22       3.97
50 years or older      734         72492    559.60      13.22      21.98       3.36

*COUNTRY ID*=BELGIUM (Flemish)

     btdgage             N        sciwgt      mnpv    mnpv_se        pct     pct_se
Under 29 y.o.         1962         14757    539.63       6.39      24.66       2.77
30-39 years old       1503         14267    526.51       9.64      23.84       3.09
40-49 years old       2347         20577    534.16       5.89      34.39       3.47
50 years or older     1144         10242    535.20      11.35      17.11       2.53
```

In this example the variable BTDAGE is created by collapsing the categorical variable BTBGAGE into only four categories, excluding those teachers missing this information. The results are then presented by country for each value of the variable IDCNTRY after selecting only those cases with IDGRADER=2. The country and the four age categories are presented in the first column. The second column has the number of students sampled in each category. The third column corresponds to the weight given these sampled students in the whole population, followed by the mean plausible value for science achievement and its corresponding standard errors. The last two columns represent the percentage of students sampled within each category and its corresponding standard error.

For example, we can say that in Australia the 627 students (15.75% of the sampled students) with a teacher under 29 years old had a mean science achievement of 534.66 with a standard error of 8.17, while the 647 students (18.69% of the sampled students) with a teacher 50 years or older had a mean science achievement of 552.88 with a standard error of 9.95.

In summary, to perform analyses such as those using the Student and Teacher Background data files you need to do the following:

- Identify the variable or variables of interest in the corresponding teacher file and find out about any specific national adaptations to the variable.
- Retrieve the relevant variable or variables from the corresponding teacher data files. If you are interested in looking at both mathematics and science teachers combined, then the files for these teachers need to be added or appended to each other.
- Retrieve the relevant variables from the Student-Teacher Linkage file. This includes the identification information for the country and teacher (IDCNTRY, IDTEACH, and IDLINK), the achievement score, JRR replication information, and the sampling weight. If the analysis is to be based on mathematics teachers only, then the weight variable to use is MATWGT. If the analysis is to be based on the science teachers only, then the weight variable to be used is SCIWGT. If the analysis is to be based on the science and mathematics teachers combined, then the weight variable to be used is TCHWGT.
- Merge the variables from the teacher data files into the Student-Teacher Linkage files using the variables IDCNTRY, IDTEACH and IDLINK.
- Use the macro JACKPV or JACKGEN with the corresponding arguments and parameters.
- Print out the result file.

10.8 Performing Analyses with School-Level Variables

Although the students in the TIMSS samples were selected from within a sample of schools, the school sample was designed to optimize the resulting sample of students, rather than to give an optimal sample of schools.

For this reason, it is always preferable to analyze school-level variables as attributes of students, rather than as elements in their own right. Although the school samples were not optimized, it is still possible to compute weighted numbers of schools with particular characteristics for providing reasonable estimates of percentages and averages across primary or middle schools in each country. The following example, however, describe only analyses based on student-weighted data.

For student-weighted analyses, the school-level data are analyzed to make statements about the number of students attending schools with one characteristic or another rather than the number of schools with certain characteristics. When school-level variables are analyzed, we recommend that you merge the selected school-level variables with the student-level file, and then use the sampling and weight information contained in the student-level file to make the desired statements. The examples presented in this section describe how this can be accomplished using SAS.

Let us say that we want to find out the percentage of eighth graders that attend schools located in a certain geographical area of the country (BCBGCOMM), and their average achievement in science. These results are located in the School Background Data Almanac by Science Achievement (bsalm4_m2.*) for questionnaire item SCQ2-01 as shown in Exhibit 10.16.

Exhibit 10.16 Sample Data Almanac Sheet for School-Level Analysis using Plausible Values Taken From the TIMSS 1999 International Science Report

Third International Mathematics and Science Study - 1999 Assessment 16:43 Sunday, February 18, 2001 1
School Background Data Almanac by Science Achievement - 8th Grade

Question : In what type of community is your school located?
Location : SCQ2-01 (BCBGCOMM)

Country	Sample	Valid N	RAPHICA LLY ISOLATE D AREA %	AGE OR RURAL (FARM) AREA %	KIRTS OF TOWN OR CITY %	4.CLOS CENTER OF TOWN OR CIT %	ADMINIS TERED %	%	RAPHICA ISOLATE Mean	AGE OR (FARM) Mean	KIRTS TOWN Mean	CENTER TOWN Mean	ADMINIS Mean	Mean
Australia	169	142	2.0	8.5	33.3	56.3	10.2	5.2	490.2	540.4	531.7	545.7	539.6	554.2
Belgium (Flemish)	135	122	.	32.0	23.5	44.5	6.3	5.7	.	535.6	513.0	549.1	546.8	499.4
Bulgaria	163	160	0.7	24.1	8.6	66.7	1.5	0.0	527.2	491.9	520.7	527.7	487.9	.
Canada	385	364	3.9	15.2	18.2	62.8	2.7	2.7	529.4	526.9	538.3	533.8	520.7	534.8
Chile	185	182	3.0	16.4	26.9	53.8	0.6	1.1	405.9	382.3	400.0	443.1	404.6	420.4
Chinese Taipei	150	150	2.0	10.0	38.8	49.1	0.0	0.0	539.5	533.3	570.3	576.7	.	.
Cyprus	61	55	0.7	26.6	29.5	43.2	9.5	0.0	491.2	432.8	470.1	471.8	452.6	.
Czech Republic	142	139	.	20.9	41.5	37.6	1.7	0.2	.	532.7	538.4	545.6	504.9	606.3
England	128	107	0.9	14.2	40.2	44.7	17.6	0.0	510.7	590.8	531.2	537.0	521.4	.
Finland	159	157	1.4	35.8	30.0	32.9	0.8	1.3	545.9	536.1	531.0	537.3	561.3	527.3
Hong Kong, SAR	137	128	1.7	2.5	26.4	69.3	6.1	0.7	511.5	523.8	529.3	533.0	508.1	471.4
Hungary	147	142	0.6	29.8	19.2	50.4	1.1	2.0	512.2	520.2	562.0	567.1	558.4	575.3
Indonesia	150	146	1.5	40.3	29.1	29.0	0.0	2.4	366.2	418.2	430.6	467.4	.	443.2
Iran, Islamic Rep.	170	169	1.0	24.3	38.6	36.1	0.0	0.5	402.5	423.7	452.4	461.2	.	426.4
Israel	139	123	7.2	28.5	17.8	46.4	8.6	1.9	496.8	451.8	442.3	477.5	487.2	511.1
Italy	180	178	35.5	4.0	25.9	34.7	0.0	1.0	499.2	512.2	475.9	498.3	.	483.1
Japan	140	139	2.1	15.6	47.1	35.2	0.0	0.7	565.5	546.8	552.1	546.9	.	540.9
Jordan	147	146	1.8	26.3	26.6	45.3	0.0	0.8	463.5	422.9	451.1	464.7	.	484.4
Korea, Rep. of	150	149	.	11.8	35.7	52.5	0.0	0.6	.	524.2	544.3	557.0	.	557.4
Latvia (LSS)	145	128	13.3	23.4	28.3	35.0	8.8	1.4	492.3	487.1	507.3	517.5	494.1	463.3
Lithuania	150	149	2.4	27.9	8.6	61.1	0.6	0.0	442.1	476.1	496.5	494.0	511.7	.
Macedonia, Rep. of	145	143	1.6	31.6	19.5	47.3	0.0	1.3	428.1	419.2	461.2	484.1	.	447.9
Malaysia	150	150	4.7	37.6	16.6	41.2	0.0	0.0	470.0	473.9	482.8	515.8	.	.
Moldova	150	149	2.1	59.6	4.2	34.1	0.4	0.0	455.4	453.9	484.6	466.2	374.4	.
Morocco	173	168	0.7	19.1	15.7	64.5	2.6	0.0	337.2	302.6	318.0	329.5	329.6	.
Netherlands	126	98	.	27.5	44.6	27.9	16.8	0.6	.	544.0	552.0	537.3	540.2	542.4
New Zealand	152	139	4.4	12.6	25.0	58.1	6.8	1.9	512.6	494.1	513.8	509.5	522.7	504.0
Philippines	150	149	4.2	17.2	12.9	65.8	0.0	0.7	317.6	322.1	348.0	353.5	.	239.6
Romania	147	142	1.4	37.6	18.4	42.6	3.6	0.0	506.5	448.2	466.4	493.7	475.0	.
Russian Federation	189	189	0.2	35.8	.	64.0	0.0	0.0	564.4	504.8	.	542.8	.	.
Singapore	145	145	.	.	.	100.0	0.0	0.0	.	.	.	567.9	.	.
Slovak Republic	145	138	.	34.0	37.4	28.6	2.7	1.5	.	516.9	538.8	543.9	582.6	587.9
Slovenia	149	145	2.6	35.5	22.5	39.5	2.2	0.0	546.7	532.6	533.6	532.6	536.8	.
South Africa	194	186	2.7	47.2	20.0	30.0	2.1	1.3	182.6	190.1	257.5	323.7	196.7	250.9
Thailand	150	149	17.3	35.2	23.0	24.4	0.7	0.0	459.3	473.1	479.4	515.3	462.2	.
Tunisia	149	144	7.2	18.0	30.7	44.1	1.3	1.7	419.7	405.7	430.8	440.5	402.8	435.9
Turkey	204	196	2.8	3.1	15.3	78.8	0.9	2.3	452.1	422.2	424.5	434.1	376.4	463.6
United States	221	182	4.7	19.2	25.4	50.7	18.0	0.3	525.2	505.2	537.7	509.6	507.2	499.1
International Avg.	160	152	4.3	24.6	25.7	48.1	3.5	1.0	467.8	470.8	483.8	498.7	477.2	482.8

As in the previous example, the first step in our analysis is to locate the variables of interest in the specific codebook and file. We find the variable BCBGCOMM in the School Background file, and the student weights and plausible values in the Student Background file. We then proceed to review the documentation on national adaptations and discover that Australia has modified this variable slightly to fit their particular context. At this time we could proceed in one of two ways: we could exclude Australia from our analysis or we could label the variable accordingly so that we will not be making incorrect inferences about the specific groups. In the latter case, since we want to also explore the results for Australia we take the precaution of labeling the values for variable BCBGCOMM in a generic way before we proceed with the analysis.

After these considerations, we then proceed to read the School Background file and keep only the variables that are relevant to our analysis. In this case we keep the country identification (IDCNTRY) and school identification (IDSCHOOL). We keep these variables because these are the variables that will allow us to merge the school data to the student data. We also keep from the School Background file the variable of interest, in this case BCBGCOMM. We then read the variables of interest from the student data file. First we read the identification of the country and the school (IDCNTRY and IDSCHOOL) which will allow us to merge the student data to the school data. We also select from this variable the international science achievement plausible values (BSSSCI01-BSSSCI05), the sampling weight for the student (TOTWGT), the variables that contain the jackknife replication information (JKZONE and JKREP), and the variable that will be used to select the eighth graders from the data file (IDGRADER).

We then proceed to merge the school information with the student information using the variables IDCNTRY and IDSCHOOL as merge variables, and then use the macro JACKPV to obtain the corresponding percentages of students within each group, and their mean achievement scores in science. The computer code used to run this analysis in SAS can be found in Exhibit 10.17 and an extract of the results are shown in Exhibit 10.18.

Exhibit 10.17 SAS Control Statements for Performing Student-Weighted Analyses with School-Level Variables (EXAMPLE4.SAS)

```
libname bm2 "d:\timss\data\bm2";

%include "d:\Timss\Programs\jackpv.sas";

* Read the variables from the school file and sort them by the merge variables;
data school;
     set bm2.bcgallm2
       (keep=idcntry idschool bcbgcomm);

proc sort data=school;
     by idcntry idschool;

* Read the variables from the student file and sort them by the merge variables;
data student;
     set bm2.bsgallm2
       (keep=idcntry idschool idgrader jkrep jkzone totwgt bsssci01-bsssci05 intms99);

proc sort data=student;
     by idcntry idschool;

* Merge the tudent and school files together;
data merged;
     merge student school;
     by idcntry idschool;
      if nmiss(bcbgcomm)=0 and intms99=1 and idgrader = 2;

* Define the format for the variables used;
proc format library=work;

 value comm  1 = "Community Type 1"
             2 = "Community Type 2"
             3 = "Community Type 3"
             4 = "Community Type 4";

 value country
 036= 'AUSTRALIA         '   956= 'BELGIUM (Flemish) '   100= 'BULGARIA          '
 124= 'CANADA            '   152= 'CHILE             '   196= 'CYPRUS            '
 203= 'CZECH REP.        '   926= 'ENGLAND           '   246= 'FINLAND           '
 344= 'HONG KONG         '   348= 'HUNGARY           '   360= 'INDONESIA         '
 364= 'IRAN, ISLAMIC REP.'   376= 'ISRAEL            '   380= 'ITALY             '
 400= 'JORDAN            '   392= 'JAPAN             '   410= 'KOREA, REP. OF    '
 440= 'LITHUANIA         '   428= 'LATVIA            '   504= 'MOROCCO           '
 498= 'MOLDOVA, REP. OF  '   807= 'MACEDONIA, REP. OF'   458= 'MALAYSIA          '
 528= 'NETHERLANDS       '   554= 'NEW ZEALAND       '   608= 'PHILIPPINES       '
 642= 'ROMANIA           '   643= 'RUSSIAN FEDERATION'   702= 'SINGAPORE         '
 703= 'SLOVAK REP.       '   705= 'SLOVENIA          '   764= 'THAILAND          '
 788= 'TUNISIA           '   792= 'TURKEY            '   158= 'CHINESE TAIPEI    '
 840= 'UNITED STATES     '   710= 'SOUTH AFRICA      ';

%jackpv(totwgt , jkzone, jkrep, 75, idcntry bcbgcomm, bsssci0, 5, merged);

proc print data=final noobs;
     by idcntry ;
     var bcbgcomm N totwgt mnpv mnpv_se pct pct_se;
     format idcntry country. bcbgcomm comm.
            N 6.0 totwgt 10.0 mnpv mnpv_se pct pct_se 6.2;

run;
```

Exhibit 10.18 Extract of SAS Computer Output for Performing Student-Weighted Analyses with School-Level Variables (EXAMPLE 4)

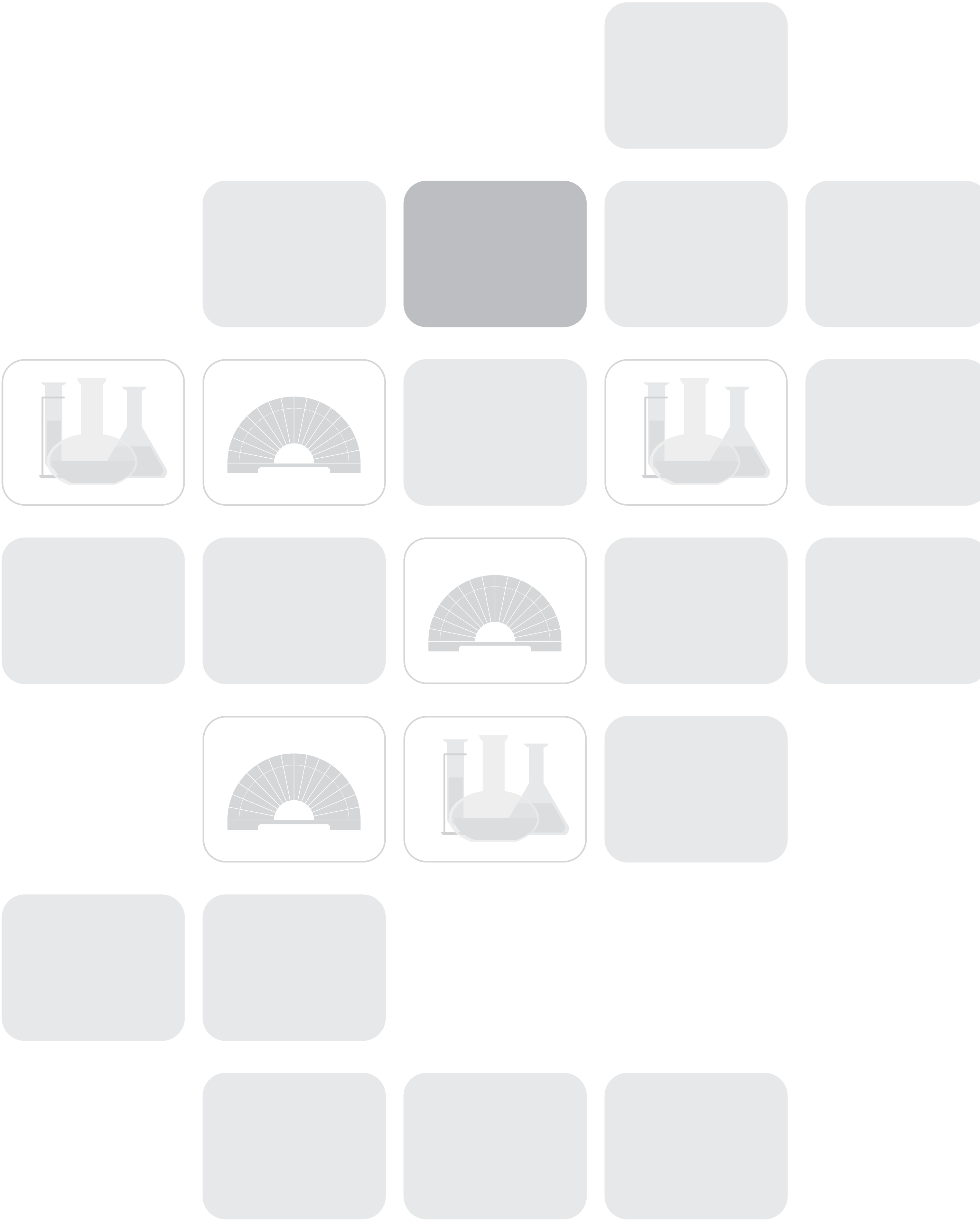
```
*COUNTRY ID*=AUSTRALIA

    BCBGCOMM            N         TOTWGT      mnpv    mnpv_se        pct     pct_se
Community Type 1       70           4369    490.23       4.25       1.99       1.16
Community Type 2      279          18620    540.38      17.94       8.46       2.57
Community Type 3     1143          73238    531.72       8.26      33.28       3.93
Community Type 4     1892         123818    545.68       7.66      56.27       4.15

*COUNTRY ID*=BULGARIA

    BCBGCOMM            N         TOTWGT      mnpv    mnpv_se        pct     pct_se
Community Type 1       16            604    527.16      10.36       0.69       0.70
Community Type 2      716          20942    491.94      11.43      24.06       3.18
Community Type 3      314           7455    520.70      17.88       8.57       2.25
Community Type 4     2162          58025    527.69       6.30      66.67       4.36

*COUNTRY ID*=CANADA

    BCBGCOMM            N         TOTWGT      mnpv    mnpv_se        pct     pct_se
Community Type 1      492          13556    529.41       9.37       3.86       1.00
Community Type 2     1527          53203    526.94       5.25      15.15       2.12
Community Type 3     1567          63860    538.34       7.02      18.19       2.50
Community Type 4     4684         220446    533.75       2.38      62.79       2.97

.
.
.
.
.

*COUNTRY ID*=UNITED STATES

    BCBGCOMM            N         TOTWGT      mnpv    mnpv_se        pct     pct_se
Community Type 1      277         128961    525.24      15.72       4.73       2.27
Community Type 2     1247         522877    505.19       7.92      19.18       2.49
Community Type 3     1651         693146    537.74       7.80      25.42       3.54
Community Type 4     4313        1381329    509.65       6.63      50.67       4.71

*COUNTRY ID*=ENGLAND

    BCBGCOMM            N         TOTWGT      mnpv    mnpv_se        pct     pct_se
Community Type 1       21           3911    510.71      10.92       0.86       0.86
Community Type 2      327          64843    590.84      10.62      14.24       3.39
Community Type 3      985         182882    531.16       6.88      40.17       4.62
Community Type 4     1130         203619    537.02       8.59      44.73       4.47

*COUNTRY ID*=BELGIUM (Flemish)

    BCBGCOMM            N         TOTWGT      mnpv    mnpv_se        pct     pct_se
Community Type 2     1528          18438    535.60       5.71      31.99       3.48
Community Type 3      932          13534    513.03       9.73      23.48       3.40
Community Type 4     2306          25659    549.12       5.33      44.52       3.96
```

In this example the variable BCBGCOMM is relabeled using generic labels due to national adaptations that could affect interpretation. The results are then presented by country for each value of the variable IDCNTRY after selecting only those cases with IDGRADER=2. The country and the four community types are presented in the first column. The second column has the number of students sampled in each category. The third column corresponds to the weight given these sampled students in the whole population, followed by the mean plausible value for science achievement and its corresponding standard error. The last two columns represent the percentage of students sampled within each category and its corresponding standard error.

For example, we can say that in Australia, 70 students who represent 4369 students in the population responded that they live in community type 1. These students had a mean science achievement of 490.23 with a standard error of 4.25. These students represent 1.99% of the sampled students and this percentage has a standard error of 1.16.

In summary, to perform analyses such as those using the Student and School Background files, you need to do the following:

- Identify the variable or variables of interest in the student file and find out about any specific national adaptations to the variable.
- Retrieve the relevant variables from the student files, including the achievement score, sampling weights, JRR replication information and any other variables used in the selection of cases.
- Retrieve the relevant classification variable or variables from the school database.
- Merge the variables from the school database onto the student database using the variables IDCNTRY and IDSCHOOL.
- Use the macro JACKGEN or JACKPV with the corresponding arguments and parameters.
- Print out the result file.



References

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

Beaton, A. E. & Gonzalez, E. J. (1997). "TIMSS Test-Curriculum Matching Analysis" in Martin, M.O. and Kelly, D.L. (Eds.), *TIMSS technical report, volume II: Implementation and Analysis*. Chestnut Hill, MA: Boston College.

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

Foy, P. (2000). "Sampling Weights" in Martin, M.O., Gregory, K.D. and Stemler, S.E. (Eds.), *TIMSS 1999 technical report: IEA's repeat of the Third International Mathematics and Science Study at the eighth grade*. Chestnut Hill, MA: Boston College.

Foy, P. and Joncas, M. (2000). "Implementation of the Sample Design" in Martin, M.O., Gregory, K.D. and Stemler, S.E. (Eds.), *TIMSS 1999 technical report: IEA's repeat of the Third International Mathematics and Science Study at the eighth grade*. Chestnut Hill, MA: Boston College.

Foy, P., Rust, K., and Schleicher, A. (1996). "Sample design" in M.O. Martin and D.L. Kelly (Eds.), *TIMSS technical report, volume I: Design and development*. Chestnut Hill, MA: Boston College.

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

Lie, S., Taylor, A., & Harmon, M. (1996). Scoring Techniques and Criteria, in M.O. Martin & D.L. Kelly (Eds.), *Third International Mathematics and Science Study technical report volume I: Design and development*. Chestnut Hill, MA: Boston College.

Martin, M.O., Gregory, K.D., and Stemler, S.E. (2000). *TIMSS 1999 technical report: IEA's repeat of the Third International Mathematics and Science Study at the eighth grade*. Chestnut Hill, MA: Boston College.

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

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

Martin, M.O and Kelly, D.L., Eds. (1997). *TIMSS technical report, volume III: Implementation and analysis, Final year of secondary school.* Chestnut Hill, MA: Boston College.

Martin, M.O. and Mullis, I.V.S., Eds. (1996). *TIMSS: Quality assurance in data collection*. Chestnut Hill, MA: Boston College.

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

Martin, M.O., Mullis, I.V.S., Gonzalez, E. J., Gregory, K.D., Smith, T.A., Chrostowski, S.J., Garden, R.A., & O'Connor, K.M. (2000). *TIMSS 1999 international science report: Findings from IEA's repeat of the Third International Mathematics and Science Study at the eighth grade*. Chestnut Hill, MA: Boston College.

Mislevy, R.J. & Bock, R.D. (1982). *BILOG: Item analysis and test scoring with binary logistic models* [Computer program]. Morresville, IN: Scientific Software.

Mislevy, R.J. (1985). Estimation of latent group effects. *Journal of the American Statistical Association, 80,* 993-97.

Mislevy, R.J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika, 56*, 177- 196.

Mullis, I.V.S and Smith, T.S. (1996). "Quality control steps for free-response scoring" in M.O. Martin and I.V.S. Mullis (Eds.), *TIMSS: Quality assurance in data collection*. Chestnut Hill, MA: Boston College.

Mullis, I.V.S. and Martin, M.O. (1997). "Item analysis and review" in M.O. Martin and D.L. Kelly (Eds.), *TIMSS technical report, volume II: Implementation and analysis*. Chestnut Hill, MA: Boston College.

Mullis, I.V.S., & Smith, T.A. (1996). "Quality Control Steps for Free- Response Scoring" in M.O. Martin and I.V.S. Mullis (Eds.), *Third International Mathematics and Science Study: Quality assurance in data collection*. Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Jones, C.O., and Garden, R.A. (1996). "Training sessions for free-response scoring and administration of the performance assessment" in M.O. Martin and I.V.S. Mullis (Eds.), *TIMSS: Quality assurance in data collection*. Chestnut Hill, MA: Boston College.

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

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Gregory, K.D., Garden, R.A., O'Connor, K.M., Chrostowski, S.J., & Smith, T.A. (2000). *TIMSS 1999 international mathematics report: Findings from IEA's repeat of the Third International Mathematics and Science Study at the eighth grade*. Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Gregory, K.D., Garden, R.A., O'Connor, K.M., Chrostowski, S.J., & Smith, T.A. (2000). *TIMSS 1999 international mathematics report: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the eighth grade*. Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Gregory, K.D., Garden, R.A., O'Connor, K.M., Chrostowski, S.J., & Smith, T.A. (2000). *TIMSS 1999 international mathematics report: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the eighth grade*. Chestnut Hill, MA: Boston College.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16(2), 159-176.

Muraki, E., & Bock, R.D. (1991). PARSCALE: *Parameter scaling of rating data*. Chicago, IL: Scientific Software, Inc.

Robitaille, D.F. & Garden, R.A. (1996). "Design of the Study" in D.F. Robitaille & R.A. Garden (Eds.), *TIMSS Monograph No. 2: Research questions & study design*. Vancouver, Canada: Pacific Educational Press.

Rubin, D.B. (1987). *Multiple imputation for non-response in surveys*. New York: John Wiley & Sons.

Schmidt, W. & Cogan, L. (1996). "Development of the TIMSS Context Questionnaires" in M.O. Martin & D.L. Kelly (Eds.), *Third International Mathematics and Science Study technical report, volume 1: Desing and development*. Chestnut Hill, MA: Boston College.

Schmidt, W.H., McKnight, C.C., Valverde, G.A., Houang, R.T., & Wiley, D.E. (1997a). *Many visions, many aims: A Cross- national investigation of curricular intentions in school mathematics*. Norwell, MA: Kluwer Academic Press.

Schmidt, W.H., Raizen, S.A., Britton, E.D., & Bianchi, L.J. (1997b *Many visions, many aims: A Cross- national investigation of curricular intentions in school science*. Norwell, MA: Kluwer Academic Press.

Smith, T.A., Martin, M.O., Mullis, I.V.S., & Kelly, D.L. (2000). *Profiles of student achievement in science at the TIMSS international benchmarks: U.S. performance and standards in an international context*, Chestnut Hill, MA: Boston College.

Thomas, N. (1993). Asymptotic corrections for multivariate posterior moments with factored likelihood functions. *Journal of Computational and Graphical Statistics*, *2*, 309-22.

TIMSS (1995). *Guide to checking, coding, and entering the TIMSS data*. Chestnut Hill, MA: Boston College.

TIMSS (1996a). *TIMSS mathematics items – Released set for population 2 (seventh and eighth grades*). Chestnut Hill, MA: Boston College.

TIMSS (1996b). *TIMSS science items – Released set for population 2 (seventh and eighth grades).* Chestnut Hill, MA: Boston College.

TIMSS (1997). *TIMSS 1999 School Sampling Manual–Version 2* (Doc. Ref.: TIMSS 1999 97-0012). Prepared by Pierre Foy, Statistics Canada. Chestnut Hill, MA: Boston College.

TIMSS (1997a). *TIMSS mathematics items – Released set for population 1 (third and fourth grades).* Chestnut Hill, MA: Boston College.

TIMSS (1997a). *Survey Operations Manual* (TIMSS 1999 Doc. Ref. No. 98-0026). Prepared by the International Study Center at Boston College. Chestnut Hill, MA: Boston College.

TIMSS (1997b). *TIMSS science items – Released set for population 1 (third and fourth grades).* Chestnut Hill, MA: Boston College.

TIMSS (1997). *TIMSS-R School Sampling Manual* (TIMSS 1999 Doc. Ref. 97-0012). Prepared by Pierre Foy, Statistics Canada. Chestnut Hill, MA: Boston College.

TIMSS (1997c). *School Coordinators Manual* (TIMSS 1999 Doc. Ref. 98-0024). Prepared by the International Study Center at Boston College. Chestnut Hill, MA: Boston College.

TIMSS (1997d). *Test Administrator Manual* (TIMSS 1999 Doc. Ref. 98-0025). Prepared by the IEA Data Processing Center. Chestnut Hill, MA: Boston College.

TIMSS (1998). *Manual for Entering the TIMSS-R Data* (Doc. Ref. No.: 98-0028). Prepared by the International Study Center at Boston College. Chestnut Hill, MA: Boston College.

TIMSS (1998a). *Scoring Guides for Mathematics and Science Free- Response Items* (TIMSS 1999 Doc. Ref. 98-0049). Prepared by the International Study Center at Boston College. Chestnut Hill, MA: Boston College.

TIMSS (1998a). *Survey Operations Manual–Main Survey* (Doc. Ref.: TIMSS 1999 98-0026). Prepared by the International Study Center. Chestnut Hill, MA: Boston College.

TIMSS (1998b). *Guidelines for the Translation Verification of the TIMSS-R Main Survey Instruments* (Doc. Ref. No. 98- 0042). Prepared by the International Study Center at Boston College. Chestnut Hill, MA: Boston College.

TIMSS (1998b). *Manual for Entering the TIMSS-R Data* (TIMSS 1999 Doc. Ref. 98-0028). Prepared by the International Study Center at Boston College. Chestnut Hill, MA: Boston College.

TIMSS (1998b). *School Coordinator Manual–Main Survey* (Doc. Ref.: TIMSS 1999 98-0024). Prepared by the International Study Center. Chestnut Hill, MA: Boston College.

TIMSS (1998c). *Manual for National Quality Control Observers* (TIMSS 1999 Doc. Ref. 98-0044). Prepared jointly by the International Study Center at Boston College and the IEA's Data Processing Center. Chestnut Hill, MA: Boston College.

Travers, K.J., & Westbury, I. (1989). *The IEA study of mathematics I: Analysis of mathematics curricula*. Oxford: Pergamon Press.

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

Wingersky, M., Kaplan, B.A., & Beaton, A.E. (1987). "Joint estimation procedures" in A. E. Beaton (Ed.), *Implementing the new design: The NAEP 1983-84 technical report* (pp.285-92) (No. 15-TR-20). Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

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

Yamamoto, K. and Kulick, E. (2000) "Scaling methodology and procedures for the TIMSS mathematics and science scales" in Martin, M.O., Gregory, K.D. and Stemler, S.E. (Eds.), *TIMSS 1999 technical report: IEA's repeat of the Third International Mathematics and Science Study at the eighth grade*. Chestnut Hill, MA: Boston College.

A Acknowledgements

TIMSS 1999 was truly a collaborative effort among hundreds of individuals around the world. Staff from the national research centers in each participating country, the International Association for the Evaluation for Educational Achievement (IEA), the International Study Center (ISC) at Boston College, advisors, and funding agencies worked closely to develop and implement TIMSS 1999. The project would not have been possible without the tireless efforts of all involved. Below, individuals and organizations are acknowledged for their contributions. Given that the implementation of TIMSS 1999 has spanned approximately four years and involved so many people and organizations, this list may not pay heed to all who contributed throughout the life of the project. Any omission is inadvertent. TIMSS 1999 also acknowledges the students, teachers, and school principals who contributed their time and effort to the study. This report would not be possible without them.

A.1 Funding Agencies

Funding for the international coordination of TIMSS 1999 was provided by the National Center for Education Statistics of the U.S. Department of Education, the U.S. National Science Foundation, the World Bank, and participating countries. Valena Plisko, Eugene Owen, and Patrick Gonzales of the National Center for Education Statistics; Larry Suter, Elizabeth VanderPutten, and Janice Earle of the National Science Foundation; and Marlaine Lockheed of the World Bank each played a crucial role in making TIMSS 1999 possible and for ensuring the quality of the study. Each participating country was responsible for funding national project costs and implementing TIMSS 1999 in accordance with the international procedures.

A.2 Management and Operations

TIMSS 1999 was conducted under the auspices of the IEA. TIMSS 1999 was co-directed by Michael O. Martin and Ina V.S. Mullis, and managed centrally by the staff of the International Study Center at Boston College, Lynch School of Education. Although the study was directed by the International Study Center, and its staff members implemented various parts of TIMSS 1999, important activities

also were carried out in centers around the world. In the IEA Secretariat, Hans Wagemaker, Executive Director, was responsible for overseeing fundraising and country participation. The IEA Secretariat also coordinated translation verification and recruiting of quality control monitors. The data were processed centrally by the IEA Data Processing Center in Hamburg. Statistics Canada was responsible for collecting and evaluating the sampling documentation from each country and for calculating the sampling weights. The Educational Testing Service in Princeton, New Jersey conducted the scaling of the achievement data.

IEA Secretariat

Hans Wagemaker, Executive Director
Barbara Malak, Manager Membership Relations
Leendert Dijkhuizen, Fiscal Officer

International Study Center at Boston College

Responsible for TIMSS and PIRLS

Michael O. Martin, Co-Director
Ina V.S. Mullis, Co-Director
Eugenio J. Gonzalez, Director of Operations and Data Analysis
Kelvin D. Gregory, TIMSS Project Coordinator
Teresa A. Smith, TIMSS Science Coordinator
Robert Garden, TIMSS Mathematics Coordinator
Kathleen O'Connor, TIMSS Benchmarking Coordinator
Dana L. Kelly, PIRLS Project Coordinator
Steven Chrostowski, Research Associate
Ce Shen, Research Associate
Julie Miles, Research Associate
Ann Kennedy, Research Associate
Steven Stemler, Research Associate
Joseph Galia, Statistician/Programmer
Andrea Pastelis, Database Manager
Kieran Brosnan, Technology Support Specialist
Christine Conley, Publications Design Manager
Jose R. Nieto, Publications Manager
Mario Pita, Data Graphics Specialist
Betty Hugh, Data Graphics Specialist
Susan Messner, Data Graphics Specialist
Isaac Li, Data Graphics Assistant
Kathleen Packard, Manager, Finance
Susan Comeau, Manager, Office Administration
Ann Tan, Manager, Conference Administration
Monica Guidi, Administrative Coordinator
Laura Misas, Administrative Coordinator

Statistics Canada

Pierre Foy, Senior Methodologist
Marc Joncas, Senior Methodologist
Andrea Farkas, Junior Methodologist
Salina Park, Cooperative Exchange Student

IEA Data Processing Center

Dirk Hastedt, Senior Researcher
Heiko Sibberns, Senior Researcher
Knut Schwippert, Senior Researcher
Caroline Dupeyrat, Researcher
Oliver Neuschmidt, Researcher
Stephan Petzchen, Research Assistant
Anneke Niemeyer, Research Assistant
Juliane Pickel, Research Assistant

Educational Testing Services

Kentaro Yamamoto, Principal Research Scientist
Ed Kulick, Manager, Research Data Analysis

Project Management Team

Michael Martin, International Study Center
Ina Mullis, International Study Center
Eugenio Gonzalez, International Study Center
Hans Wagemaker, IEA Secretariat
Dirk Hastedt, IEA Data Processing Center
Pierre Foy, Statistics Canada
Kentaro Yamamoto, Educational Testing Services
Eugene Johnson, American Institutes for Research

Sampling Referee

Keith Rust, Westat, Inc.

National Research Coordinators

The TIMSS 1999 National Research Coordinators and their staff had the enormous task of implementing the TIMSS 1999 design. This required obtaining funding for the project; participating in the development of the instruments and procedures; conducting field tests; participating in and conducting training sessions; translating the instruments and procedural manuals into the local language; selecting the sample of schools and students; working with the schools to arrange for the testing; arranging for data collection, coding, and data entry; preparing the data files for submission to the IEA Data Processing Center; contributing to the development of the international reports; and preparing national reports. The way in which the national centers operated and the resources that were available varied considerably across the TIMSS 1999 countries. In some countries, the tasks were conducted centrally, while in others, various components were subcontracted to other organizations. In some countries, resources were more than adequate, while in others, the national centers were operating with limited resources. Of course, across the life of the project, some NRCs have changed. This list attempts to include all past NRCs who served for a significant period of time as well as all the present NRCs. All of the TIMSS 1999 National Research Coordinators and their staff members are to be commended for their professionalism and their dedication in conducting all aspects of TIMSS.

Australia

Susan Zammit
Australian Council for Educ. Res.(ACER)
19 Prospect Hill Rd.
Private Bag 55
Camberwell, Victoria 3124

Belgium (Flemish)

Christiane Brusselmans-Dehairs
Jean-Pierre Verhaeghe
Vakgroep Onderwijskunde Universiteit Gent
Henri Dunantlaan 2
B 9000 Gent

Jan Van Damme
Afd. Didactiek
Vesaliusstraat 2
B-3000 Leuven

Ann Van Den Broeck
Dekenstraat 2
AFD.Didaktiek
3000 Leuven

Bulgaria

Kiril Bankov
Faculty of Mathematics and Informatics
University of Sofia
1164 Sophia

Canada

Alan Taylor
Applied Research and Evaluation Services (ARES)
University of British Columbia
6058 Pearl Avenue,
Burnaby, BC V5H 3P9

Richard Jones
Education Quality & Accountability Office(EQAO)
2 Carlton St., Suite 1200
Toronto, ON M5B2M9

Jean-Louis Lebel
Direction de la sanction des etudes
1035 rue De La Chevrotiere
26 etage
Quebec GIR 5A5

Michael Marshall
University of British Columbia
Faculty of Education, Rm 6
2125 Main Mall
Vancouver, BC V6T1Z4

Chile

Maria Ines Alvarez
Unidad de Curriculum y Evaluacion
Ministerio de Educacion
Alameda 1146
Sector B, Piso 8

Chinese Taipei

Jau-D Chen
Dean of General Affairs
National Taiwan Normal University
162, E. Hoping Rd. Sec. 1
Taipei, Taiwan 117

Cyprus

Constantinos Papanastasiou
Department of Education
University of Cyprus
P.O. Box 20537
Nicosia CY-1678

Czech Republic

Jana Paleckova
Institute for Information of Education (UIV)
Senovazne nam.26
111 21 Praha 1

England

Graham Ruddock
National Foundation for Educational Research (NFER)
The Mere, Upton Park
Slough, Berkshire
SL1 2DQ

Finland

Pekka Kupari
University of Jyvaskyla
Institute for the Educational Research
P. O. Box 35
SF – 40351 Jyvaskyla

Hong Kong, SAR

Frederick Leung
The University of Hong Kong –
Department of Curriculum
Faculty of Education, Rm. 219
Pokfulam Road

Hungary

Péter Vari
National Institute of Public Education
Centre for Evaluation Studies
Dorottya u.8, Pf 701/420
1051 Budapest

Indonesia

Jahja Umar
Examiniation Development Center
Jalan Gunung Sahari Raya – 4
Jakarta Pusat
Jakarta

Iran, Islamic Republic

Ali Reza Kiamanesh
Ministry of Education
196, Institute for Education Research
Keshavaraz Blvd.
Tehran, 14166

Israel

Ruth Zuzovsky
Tel Aviv University
School of Education
Center for Science and Technology Education
Ramat Aviv 69978

Italy

Anna Maria Caputo
Ministerio della Pubblica Istruzione
Centro Europeo Dell 'Educazione (CEDE)
5-Villa Falconieri
Frascati (Roma)
00044

Japan

Yuji Saruta
Hanako Senuma
National Institute for Educational Research (NIER)
6-5-22 Shimomeguro
Meguro-ku, Tokyo
153-8681

Jordan

Tayseer Al-Nhar
National Center for Human Resources Development
P. O. Box 560
Amman, Jordan 11941

Korea, Republic of

Sungsook Kim
Chung Park
Korea Institute of Curriculum & Evaluation (KICE)
KICE
25-1 Samchung-dong
GhongRo-Gu, Seoul
110-230

Latvia

Andrejs Geske
University of Latvia
IEA National Research Center
Jurmalas Gatve 74/76, Rm. 204A
Riga
LV-1083

Lithuania

Algirdas Zabulionis
National Examinations Center
Ministry of Education and Science
M. Katkaus 44
Vilnius
LT2051

Macedonia, Republic of

Anica Aleksova
Ministry of Education and Science
Bureau for Development of Education
Ruder Boskovic St. bb.
1 000 Skopje

Malaysia

Azmi Zakaria
Ministry of Education
Level 2,3 &5 Block J South
Pusat Bandar Damansara, Kuala Lumpur
50604

Moldova, Republic Of

Ilie Nasu
Ministry of Education and Science
University "A. Russo"
Str. Puschin 38
Balti 3100

Lidia Costiuc
1 Piata Mazzi Adunazi Nationale
Chisinau

Morocco

Mohamed Fatihi
Direction de l'Evaluation du Systeme Educatif
INNOVATIONS PEDAGOGIQUES
32 Boulevard Ibn Toumert
Place Bob Rouah, Rabat

Netherlands

Klaas Bos
University of Twente
Centre for Applied Research in Education (OCTO)
P.O. Box 217
7500 AE Enschede

New Zealand

Megan Chamberlain
Ministry of Education
CER Unit-Research Division
45-47 Pipitea Street
Thorndon, Wellington

Philippines

Ester Ogena
DOST-Science Education Institute
3F PTRI Blg
Bicutan, Taguig
Metro Manila 1604

Vivien Talisayon
Institute Of Science & Mathematics
Education Development
University of the Philippines UPISMED
Diliman, Quezon City

Romania

Gabriela Noveanu
Institute for Educational Sciences
Evaluation and Forecasting Division
Str. Stirbei Voda 37
Bucharest Ro-70732

Russian Federation

Galina Kovalyova
Center for Evaluating the Quality of
Education
Institute of General Secondary Education
ul. Pogodinskaya 8
Moscow
119905

Singapore

Cheow Cher Wong
Research and Evaluation Branch
Ministry of Education
1 North Buona Vista Dr /MOE Building
Singapore, Singapore
138675

Slovak Republic

Olga Zelmanova
Maria Berova
SPU-National Institute for Education
Pluhova 8, P. O. Box 26
Brastislava
830 00

Slovenia

Barbara Japelj
Educational Research Institute Ljubljana
Gerbiceva 62
Ljubljana
1000

South Africa

Sarah Howie
Human Sciences Research Council
134 Pretorius Street
Private Bag X41
Pretoria
0001

Thailand

Precharn Dechsri
Institute For the Promotion of Teaching
Science & Technology (IPST)
924 Sukhumvit Rd. Ekamai
Bangkok
10100

Tunisia

Ktari Mohsen
Ministere de l'Education
Boulevard Bab-Bnet
Tunis

Turkey

Yurdanur Atlioglu
Educational Research and Development
Directorate
Gazi Mustafa Kemal Bulvani
No 109/5-6-7
Maltepe, Ankara
06570

United States of America

Patrick Gonzales
National Center for Education Statistics
U.S. Department of Education
1990 K St., NW Rm 9071
Washington, DC 20006

TIMSS 1999 Advisory Committees

The International Study Center at Boston College was supported in its work by advisory committees. The Subject Matter Item Replacement Committee was instrumental in developing of TIMSS 1999 tests, and the Questionnaire Item Review Committee revised the TIMSS questionnaires. The Scale Anchoring Panel developed the descriptions of the international benchmarks in mathematics and science.

Subject Matter Item Replacement Committee

Mathematics

Antoine Bodin, France
Anna Maria Caputo, Italy
Nobert Delagrange, Belgium (Flemish)
Jan de Lange, Netherlands
Hee-Chan Lew, Republic of Korea
Mary Lindquist, United States
David Robitaille, Canada

Science

Hans Ernst Fischer, Germany
Galina Kovalyova, Russian Federation
Svein Lie, Norway
Masao Miyake, Japan
Graham Orpwood, Canada
Jana Strakova, Czech Republic
Carolyn Swain, England

Subject Area Coordinators

Bob Garden, New Zealand (Mathematics)
Teresa Smith, United States (Science)

Special Consultants

Chancey Jones, Mathematics
Christine O'Sullivan, Science

Questionnaire Item Review Committee

Im Hyung, Republic of Korea
Barbara Japelj, Slovenia
Trevor Williams, United States
Graham Ruddock, England
Klaas Bos, Netherlands

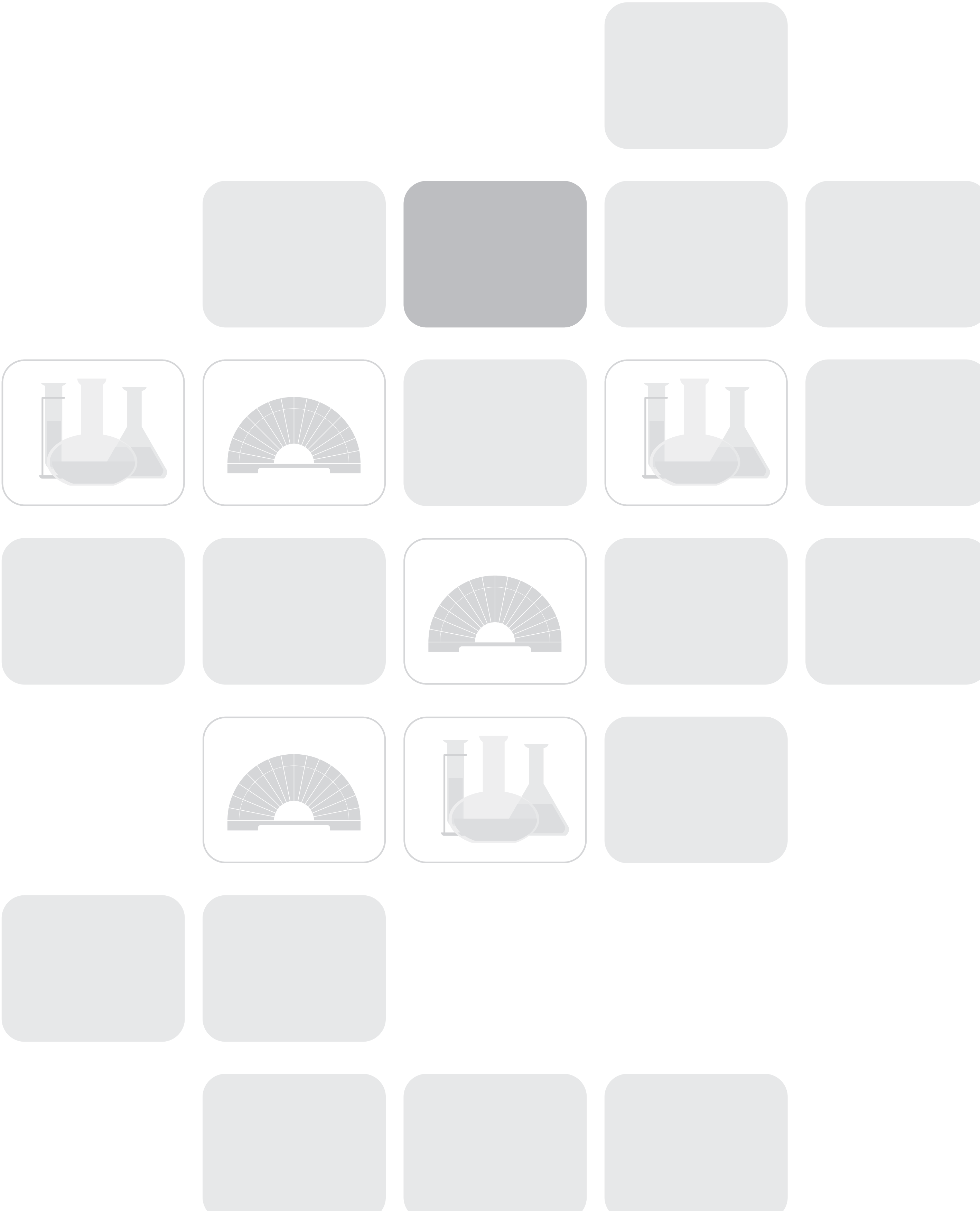
Scale Anchoring Committees

Mathematics

Anica Aleksova, Republic of Macedonia
Lillie Albert, United States
Kiril Bankov, Bulgaria
Jau-D Chen, Chinese Taipei
John Dossey, United States
Barbara Japelj, Slovenia
Mary Lindquist, United States
David Robitaille, Canada
Graham Ruddock, United Kingdom
Hanako Senuma, Japan
Pauline Vos, The Netherlands

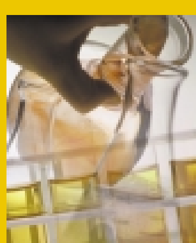
Science

Audrey Champagne, United States
Galina Kovalyova, Russian Federation
Jan Lokan, Australia
Jana Paleckova, Czech Republic
Senta Raizen, United States
Vivien Talisayon, Philippines
Hong Kim Tan, Singapore



The
International Association
for the Evaluation of
Educational Achievement

The
International
Study Center at
Boston College
Lynch School of Education



TIMSS 1999 USER GUIDE for the INTERNATIONAL DATABASE

IEA's Repeat
of the Third International
Mathematics and Science Study
at the Eighth Grade

ISBN 1-889938-21-1

TIMSS REPEAT •1999•

Boston College
Lynch School
of Education