Edited by:

Michael O. Martin

Ina V.S. Mullis

Steven J. Chrostowski

Contributors:
Alka Arora
Juliane Barth
Ralph Carstens
Steven J. Chrostowski
Dana Diaconu
Ebru Erberber
Pierre Foy
Joseph Galia
Robert A. Garden
Eugenio J. Gonzalez
Marc Joncas
Isaac Li
Barbara Malak
Michael O. Martin
Ina V.S. Mullis
Teresa A. Smith Neidorf
María José Ramírez

TIMSS

TIMSS 2003 Technical Report

Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades

International Association for the Evaluation of Educational Achievement

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

TIMSS 2003 Technical Report / Edited by: Michael O. Martin, Ina V.S. Mullis,
Steven J. Chrostowski

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

Library of Congress Catalog Card Number: 2004111600

ISBN: 1-889938-35-1

For more information about TIMSS contact:

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

Tel: +1-617-552-1600
Fax:+1-617-552-1203

E-mail: timss@bc.edu

URL: timss.bc.edu

Printed and bound in the United States.

Contents

1

Chapter 1

Overview of TIMSS 2003

Michael O. Martin and Ina V.S. Mullis

1.1 Introduction

Since pioneering cross-national studies of educational achievement with the First International Mathematics Study (FIMS) in 1964, the International Association for the Evaluation of Educational Achievement (IEA) has conducted almost 20 studies of student achievement in the curricular areas of mathematics, science, language, civics, and reading. The Third International Mathematics and Science Study (TIMSS) in 1994-1995 was the largest and most complex IEA study ever conducted, including both mathematics and science at third and fourth grades, seventh and eighth grades, and the final year of secondary school.

In 1999, TIMSS (now renamed the Trends in International Mathematics and Science Study) again assessed eighth-grade students in both mathematics and science to measure trends in student achievement since 1995. Also, 1999 represented four years since the first TIMSS, and the population of students originally assessed as fourth-graders had advanced to the eighth grade. Thus, TIMSS 1999 also provided information about whether the relative performance of these students had changed in the intervening years.

TIMSS 2003, the third data collection in the TIMSS cycle of studies, was administered at the eighth and fourth grades. For countries that participated in previous assessments, TIMSS 2003 provides three-cycle trends at the eighth grade (1995, 1999, 2003) and data over two points in time at the fourth grade (1995 and 2003). In countries new to the study, the 2003 results can help policy makers and practitioners assess their comparative standing and gauge the rigor and effectiveness of their mathematics and science programs.

This volume describes the technical aspects of TIMSS 2003 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.

1.2 Participants in TIMSS 2003

Exhibit 1.1 lists all the countries that have participated in TIMSS in 1995, 1999, or 2003 at fourth or eighth grade. In all, 67 countries have participated in TIMSS at one time or another. Of the 49 countries that participated in TIMSS 2003, 48 participated at the eighth grade and 26 at the fourth grade. Yemen participated at the fourth but not the eighth grade. The exhibit shows that at the eighth grade 23 countries also participated in TIMSS 1995 and TIMSS 1999. For these participants, trend data across three points in time are available. Eleven countries participated in TIMSS 2003 and TIMSS 1999 only, while three countries participated in TIMSS 2003 and TIMSS 1995. These countries have trend data for two points in time. Of the 12 new countries participating in the study, 11 participated at eighth grade and 2 at the fourth grade. Of the 26 countries participating in TIMSS 2003 at the fourth grade, 16 also participated in 1995, providing data at two points in time.

Following the success of the TIMSS 1999 benchmarking initiative in the United States,[1] in which 13 states and 14 school districts or district consortia administered the TIMSS assessment and compared their students' achievement to student achievement world wide, TIMSS 2003 included an international benchmarking program, whereby regions of countries could participate in the study to compare to international standards. TIMSS 2003 included four benchmarking participants at the eighth grade: the Basque Country of Spain, the U.S. state of Indiana, and the Canadian provinces of Ontario and Quebec. Indiana, Ontario, and Quebec participated also at the fourth grade. Having also participated in 1999, Indiana has data at two points in time at eighth grade. Ontario and Quebec participated also in 1995 and 1999, and so have trend data across three points in time at both grade levels.

1.3 Student Populations

TIMSS 2003 had as its intended target population all students at the end of their eighth and fourth years of formal schooling in the participating countries. However, for comparability with previous TIMSS assessments, the formal definition for the eighth-grade population specified all students enrolled in the upper of the two adjacent grades that contained the largest proportion of 13-year-old students at the time of testing. This grade level was intended to represent eight years of schooling, counting from the first year of primary or elementary schooling, and was indeed the eighth grade in most countries. Similarly, for the fourth-grade population, the formal definition specified all students enrolled in the upper of the two adjacent grades that contained the largest proportion of 9-year-olds. This grade level was intended to represent

1 See Mullis, Martin, Gonzalez, O'Connor, Chrostowski, Gregory, Garden, and Smith (2001) for the results of the benchmarking in mathematics and Martin, Mullis, Gonzalez, O'Connor, Chrostowski, Gregory, Smith, and Garden (2001) for the results in science.

four years of schooling, counting from the first year of primary or elementary schooling, and was the fourth grade in most countries.

1.4 Assessment Dates

TIMSS 2003 was administered near the end of the school year in each country. In countries in the Southern Hemisphere (where the school year typically ends in November or December) the assessment was conducted in October or November 2002. In the Northern Hemisphere, the school year typically ends in June; so in these countries the assessment was conducted in April, May, or June 2003.

1.5 Study Management and Organization

TIMSS 2003 was conducted under the auspices of the IEA. The study was directed by Michael O. Martin and Ina V.S. Mullis of the TIMSS & PIRLS International Study Center at Boston College, Lynch School of Education, where they also direct IEA's Progress in International Reading Literacy Study (PIRLS). The International Study Center was responsible for the design, development, and implementation of the study – including developing the assessment framework, assessment instruments, and survey procedures; ensuring quality in data collection; and analyzing and reporting the study results. Staff at the International Study Center worked closely with the organizations responsible for particular aspects of the study, the representatives of participating countries, and the TIMSS advisory committees.

In the IEA Secretariat, Hans Wagemaker, Executive Director, was responsible for overseeing fundraising and country participation. The IEA Secretariat also managed the ambitious translation verification effort conducted for the field test and main assessment and recruited international quality control monitors in each country. The IEA Data Processing Center was responsible for processing and verifying the data from the participating countries and for constructing the international database. Working closely with the Data Processing Center, Statistics Canada was responsible for collecting and evaluating the sampling documentation from each country and for calculating the sampling weights. Educational Testing Service in Princeton, New Jersey provided consultation on psychometric issues as well as technical support and software for scaling the achievement data. The Project Management Team, comprising the study directors and representatives from the International Study Center, IEA, Statistics Canada, and Educational Testing Service, met regularly throughout the study to discuss the study's progress, procedures, and schedule.

Exhibit 1.1 **Countries Participating in TIMSS 2003, 1999, and 1995**

Countries	Grade 8			Grade 4	
	2003	1999	1995	2003	1995
Argentina*	●	●			
Armenia	●			●	
Australia	●	●	●	●	●
Austria			●		●
Bahrain	●				
Belgium (Flemish)	●	●	●	●	
Belgium (French)			●		
Botswana	●				
Bulgaria	●	●	●		
Canada		●	●		●
Chile	●	●			
Chinese Taipei	●	●		●	
Colombia			●		
Cyprus	●	●	●	●	●
Czech Republic		●	●		●
Denmark			●		
Egypt	●				
England	●	●	●	●	●
Estonia	●				
Finland		●			
France			●		
Germany			●		
Ghana	●				
Greece			●		●
Hong Kong, SAR	●	●	●	●	●
Hungary	●	●	●	●	●
Iceland			●		●
Indonesia	●	●			
Iran, Islamic Rep. of	●	●	●	●	●
Ireland			●		●
Israel	●	●	●		●
Italy	●	●	●	●	●
Japan	●	●	●	●	●
Jordan	●	●			
Korea, Rep. of	●	●	●		●
Kuwait			●		●
Latvia	●	●	●	●	●
Lebanon	●				
Lithuania	●	●	●	●	
Macedonia, Rep. of	●	●			
Malaysia	●	●			
Moldova, Rep. of	●	●		●	

Exhibit 1.1 **Countries Participating in TIMSS 2003, 1999, and 1995** *(...Continued)*

Countries	Grade 8			Grade 4	
	2003	1999	1995	2003	1995
Morocco	●	●		●	
Netherlands	●	●	●	●	●
NewZealand	●	●	●	●	●
Norway	●		●	●	●
Palestinian Nat'l Auth.	●				
Philippines	●	●		●	
Portugal			●		●
Romania	●	●	●		
Russian Federation	●	●	●	●	
SaudiArabia	●				
Scotland	●		●	●	●
Serbia	●				
Singapore	●	●	●	●	●
Slovak Republic	●	●	●		
Slovenia	●	●	●	●	●
South Africa	●	●	●		
Spain			●		
Sweden	●		●		
Switzerland			●		
Syrian Arab Republic**	●				
Thailand		●	●		●
Tunisia	●	●		●	
Turkey		●			
United States	●	●	●	●	●
Yemen**				●	
Benchmarking Participants					
BasqueCountry, Spain	●				
IndianaState, US	●	●		●	
OntarioProvince, Can.***	●	●	●	●	●
QuebecProvince, Can.***	●	●	●	●	●

* Argentina administered the TIMSS 2003 data collection one year late, and did not score and process its data in time for inclusion in this report.

**Because the characteristics of their samples are not completely known, achievement data for Syrian Arab Republic and Yemen are presented in Appendix F of the International reports.

***Ontario and Quebec participated in TIMSS 1999 and 1995 as part of Canada.

Each participating country appointed a National Research Coordinator (NRC) and a national center responsible for all aspects of TIMSS 2003 within that country. The TIMSS & PIRLS International Study Center organized meetings of the NRCs several times a year to review study materials and procedures, and to provide training in student sampling, constructed-response item scoring, and data entry and database construction.

The TIMSS & PIRLS International Study Center was supported in its work by a number of advisory committees. The International Expert Panel in Mathematics and Science played a crucial role in developing the TIMSS 2003 frameworks and specifications for the assessment. The Mathematics and Science Item Development Task Forces coordinated the work of the National Research Coordinators in developing and reviewing the mathematics and science achievement items. The Science and Mathematics Item Review Committee reviewed and revised successive drafts of the achievement items and was an integral part of the scale anchoring process. The Questionnaire Item Review Committee revised the TIMSS context questionnaires for the 2003 assessment.

1.6 The TIMSS 2003 Assessment Frameworks

The development of the TIMSS 2003 assessment was a collaborative process spanning a two-and-a-half-year period and involving mathematics and science educators and development specialists from all over the world. Central to this effort was a major updating and revision of the existing TIMSS assessment frameworks to address changes during the last decade in curricula and the way science is taught. The resulting publication entitled *TIMSS Assessment Frameworks and Specifications 2003* serves as the basis of TIMSS 2003 and beyond (Mullis, Martin, Smith, Garden, Gregory, Gonzalez, Chrostowski, and O'Connor, 2003).

As shown in Exhibit 1.2, the mathematics and science assessment frameworks for TIMSS 2003 are framed by two organizing dimensions or aspects, a content domain and a cognitive domain. There are five content domains in mathematics (number, algebra, measurement, geometry, and data) and five in science (life science, chemistry, physics, earth science, and environmental science) that define the specific mathematics and science subject matter covered by the assessment. The cognitive domains, four in mathematics (knowing facts and procedures, using concepts, solving routine problems, and reasoning) and three in science (factual knowledge, conceptual understanding, and reasoning and analysis) define the sets of behaviors expected of students as they engage with the mathematics and science content.

Exhibit 1.2 **The Content and the Cognitive Domains of the Mathematics and Science Framework**

Mathematics		Science	
Content Domain		**Content Domain**	
Grade 8	Number	Grade 8	Life Science
	Algebra		Chemistry
	Measurement		Physics
	Geometry		Earth Science
	Data		Environmental Science
Grade 4	Number	Grade 4**	Life Science
	Patterns and Relationships*		Physical Science
	Measurement		Earth Science
	Geometry		
	Data		
Cognitive Domain		**Cognitive Domain**	
	Knowing Facts and Procedures		Factual Knowledge
	Using Concepts		Conceptual Understanding
	Solving Routine Problems		Reasoning and Analysis
	Reasoning		

* At fourth grade, the algebra content domain is called patterns and relationships.
**At the fourth grade, there are only three content areas in science, namely life science, physical science, and earth science.

1.7 Developing the TIMSS 2003 Assessment

Given TIMSS' ambitious goals for curriculum coverage and innovative problem solving tasks, as specified in the Frameworks and Specifications, the development of the assessment items required a tremendous cooperative effort, crucially dependent on the contribution of the National Research Coordinators (NRCs) during the entire process. To maximize the effectiveness of the contributions from national centers, the TIMSS & PIRLS International Study Center developed a detailed item-writing manual and conducted a workshop for countries that wished to provide items for the international item pool. At this workshop, two item development "Task Forces" reviewed general item-writing guidelines for multiple-choice and constructed-response items and provided specific training in writing mathematics and science items in accordance with the *TIMSS Assessment Frameworks and Specifications 2003*. The mathematics task force consisted of the mathematics coordinator and two experienced mathematics item writers, and similarly the science task force comprised the science coordinator and two experienced science item writers.

More than 2,000 items and scoring guides were drafted, and reviewed by the task forces. The items were further reviewed by the Science and Mathematics Item Review Committee, a group of internationally prominent mathematics and science educators nominated by participating countries to advise on subject-matter issues in the assessment. Committee members also helped to develop tasks and items to assess problem solving and scientific inquiry.

Participating countries field-tested the items with representative samples of students, and all of the potential new items were again reviewed by the Science and Mathematics Item Review Committee as well as by NRCs. The resulting TIMSS 2003 eighth-grade assessment contained 383 items, 194 in mathematics and 189 in science. The fourth grade assessment contained 313 items, 161 in mathematics and 152 in science.

Between one-third and two-fifths of the items at each grade level were in constructed-response format, requiring students to generate and write their own answers. Some constructed-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 used a multiple-choice format. In scoring the items, correct answers to most questions were worth one point. However, responses to some constructed-response questions (particularly those requiring extended responses) were evaluated for partial credit, with a fully correct answer being awarded two points. The total number of score points available for analysis thus somewhat exceeds the number of items.

Not all of the items in the TIMSS 2003 assessment were newly developed for 2003. To ensure reliable measurement of trends over time, the assessment included also items that had been used in the 1995 and 1999 assessments. For example, of the 426 score points available in the entire 2003 mathematics and science assessment, 47 came from items used also in 1995, 102 from items used also in 1999, and 267 from items used for the first time in 2003. At fourth grade, 70 score points came from 1995 items, and the remaining 267 from new 2003 items.

Every effort was made to ensure that the tests represented the curricula of the participating countries and that the items exhibited no bias toward or against particular countries. The final forms of the test were endorsed by the NRCs of the participating countries. In addition, countries had an opportunity to match the content of the test to their curriculum. They identified items measuring topics not covered in their intended curriculum. The information from this Test-Curriculum Matching Analysis, provided in Appendix C of the International Reports, indicates that omitting such items has little effect on the overall pattern of results.

1.8 TIMSS 2003 Assessment Design

With the large number of mathematics and science items, it was not possible for every student to respond to all items. To ensure broad subject-matter coverage without overburdening individual students, TIMSS 2003, as in the 1995 and 1999 assessments, used a matrix-sampling technique that assigns each assessment item to one of a set of item blocks, and then assembles student test booklets by combining the item blocks according to a balanced design. Each student takes one booklet containing both mathematics and science items. Thus, the same students participated in both the mathematics and science testing.

In the TIMSS 2003 assessment design, the 313 fourth-grade mathematics and science items and the 383 eighth-grade items were divided among 28 item blocks at each grade, 14 mathematics blocks labeled M01 through M14, and 14 science blocks labeled S01 through S14. Each block contained either mathematics items only or science items only. This general block design was the same for both grades, although the planned assessment time per block was 12 minutes for fourth grade and 15 minutes for eighth grade.

There were 12 student booklets at each grade level, with six blocks of items in each booklet. To enable linking between booklets, each block appears in two, three, or four different booklets. The assessment time for individual students was 72 minutes at fourth grade (six 12-minute blocks) and 90 minutes at eighth grade (six 15-minute blocks), which is comparable to that in the 1995 and 1999 assessments. The booklets were organized into two three-block sessions (Parts I and II), with a break between the parts.

The 2003 assessment was the first TIMSS assessment in which calculators were permitted, and so it was important that the design allow students to use calculators when working on the new 2003 items. However, because calculators were not permitted in TIMSS 1995 or 1999, the 2003 design also had to ensure that students did not use calculators when working on trend items from these assessments. The solution was to place the blocks containing trend items (blocks M01 – M06 and S01 – S06) in Part I of the test booklets, to be completed without calculators before the break. After the break, calculators were allowed for the new items (blocks M07 – M14 and S07 – S14). To provide a more balanced design, however, and have information about differences with calculator access, two mathematics trend blocks (M05 and M06) and two science trend blocks (S05 and S06) also were placed in Part II of one booklet each. Note that calculators were allowed only at the eighth grade, and not at the fourth grade.

1.9 Background Questionnaires

By gathering information about students' educational experiences together with their mathematics and science achievement on the TIMSS assessment, it is possible to identify factors or combinations of factors related to high achievement. As in previous assessments, TIMSS in 2003 administered a broad array of questionnaires to collect data on the educational context for student achievement. For TIMSS 2003, a concerted effort was made to streamline and upgrade the questionnaires. The TIMSS 2003 contextual framework (Mullis, et al., 2003) articulated the goals of the questionnaire data collection and laid the foundation for the questionnaire development work.

Across the two grades and two subjects, TIMSS 2003 involved 11 questionnaires. *National Research Coordinators* completed four questionnaires. With the assistance of their curriculum experts, they provided detailed information on the organization, emphasis, and content coverage of the mathematics and science curriculum at fourth and eighth grades. The *fourth- and eighth-grade students* who were tested answered questions pertaining to their attitudes towards mathematics and science, their academic self-concept, classroom activities, home background, and out-of-school activities. The *mathematics and science teachers* of sampled students responded to questions about teaching emphasis on the topics in the curriculum frameworks, instructional practices, professional training and education, and their views on mathematics and science. Separate questionnaires for mathematics and science teachers were administered at the eighth grade, while to reflect the fact that most younger students are taught all subjects by the same teacher, a single questionnaire was used at the fourth grade. The principals or heads of *schools* at the fourth and eighth grades responded to questions about school staffing and resources, school safety, mathematics and science course offerings, and teacher support.

1.10 Translation and Verification

The TIMSS data collection instruments were prepared in English and translated into 34 languages. Of the 49 countries and four benchmarking participants, 17 collected data in two languages and one country, Egypt, in three languages – Arabic, English, and French. In addition to translation, it sometimes was necessary to modify the international versions for cultural reasons, even in the countries that tested wholly or partly in English. This process represented an enormous effort for the national centers, with many checks along the way. The translation effort included (1) developing explicit guidelines for translation and cultural adaptation; (2) translation of the instruments by the national centers in accordance with the guidelines, using two or more independent translations; (3) consultation with subject-matter experts on cultural adaptations to ensure that the meaning and difficulty of items did not change; (4) verification of translation quality by professional translators from an independent translation company; (5) corrections by the national centers in accordance with the suggestions made; (6) verification by the International Study Center that corrections were made; and (7) a series of statistical checks after the testing to detect items that did not perform comparably across countries.

1.11 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 covered 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 NRCs' report documenting procedures used in the study. In addition, the TIMSS & PIRLS International Study Center considered it essential to monitor compliance with standardized procedures. NRCs were asked to nominate one or more persons unconnected with their national center 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 roles and responsibilities.

In all, 50 quality control monitors drawn from the 49 countries and four Benchmarking participants participated in the training. Where necessary, quality control monitors who attended the training session were permitted to recruit other monitors to assist them in covering the territory and meeting the testing timetable. All together, the international quality control monitors and those trained by them observed 1,147 testing sessions (755 for grade 8 and 392 for grade 4), and conducted interviews with the National Research Coordinator in each of the participating countries.

The results of the interviews indicate that, in general, NRCs had prepared well for data collection and, despite the heavy demands of the schedule and shortages of resources, were able to conduct the data collection efficiently and professionally. Similarly, the TIMSS tests appeared to have been administered in compliance with international procedures, including the activities before the testing session, those during testing, and the school-level activities related to receiving, distributing, and returning material from the national centers.

1.12 Scoring the Constructed-Response Items

Because a large proportion of the assessment time was devoted to constructed-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. Although not used in this report, analyses of responses based on the second digit should provide insight into ways to help students better understand science concepts and problem-solving approaches.

To ensure reliable scoring procedures based on the TIMSS rubrics, the International Study Center prepared detailed guides containing the rubrics and explanations of how to implement them, together with example student responses for the various rubric categories. These guides, along with training packets containing extensive examples of student responses for practice in applying the rubrics, were used as a basis for intensive training in scoring the constructed-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.

To gather and document empirical information about agreement among scorers in each country, TIMSS arranged to have systematic samples of at least 100 student responses to each item scored independently by two readers. The results showed a high degree of agreement for both the correctness score (the first digit) and for the two-digit diagnostic score. At the eighth grade, the percentage of exact agreement between scorers averaged 99 and 97 percent for the correctness score in mathematics and science, respectively, and 97 and 92 percent for the diagnostic score. At fourth grade, the figures were 99 and 96 percent for the mathematics and science correctness score and 97 and 92 percent for the diagnostic score. The TIMSS data from the reliability studies indicate that scoring procedures were robust for the mathematics and science items, especially for the correctness score used for the analyses in the International reports.

TIMSS 2003 also took steps to show that those constructed-response items from 1999 that were used in 2003 were scored in the same way in both assessments. In anticipation of this, countries that participated in TIMSS 1999 sent samples of scored student booklets from the 1999 eighth-grade data collection to the IEA Data Processing Center, where they were digitally scanned and stored in presentation software for later use. As a check on scoring consistency from 1999 to 2003, staff members working in each country on scoring the 2003 eighth-grade data were asked also to score these 1999 responses using the DPC software. The items from 1995 that were used in TIMSS 2003 all were in multiple-choice format, and therefore scoring reliability was not an issue. There was a high degree of scoring consistency, with 92 percent exact agreement, on average, internationally, in mathematics and 98 percent in science between the scores awarded in 1999 and those given by the 2003 scorers. There was somewhat less agreement at the diagnostic score level, with 93 percent exact agreement, on average, in mathematics and 81 percent in science.

To monitor the consistency with which the scoring rubrics were applied across countries, TIMSS collected from the Southern-Hemisphere countries that administered TIMSS in English a sample of 150 student responses to 41 constructed-response mathematics and science questions. This set of student responses was then sent to each Northern-Hemisphere country having scorers proficient in English and scored independently by one or if possible two of these scorers. All 150 responses to each of the 41 items were scored by 37 scorers from the countries that participated. Agreement across countries was defined in terms of the percentage of these scores that were in exact agreement. The results showed that scorer reliability across countries was high, particularly in mathematics, with the percent exact agreement averaging 96 percent across the mathematics items and 87 percent across the science items for the correctness score and 92 percent and 76 percent across mathematics and science items, respectively, for the diagnostic score.

1.13 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 creation of 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 TIMSS 2003 data were checked and double-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. In general, the items exhibited very good psychometric properties in all countries. In the few instances where there were poor item statistics (negative point-biserials for the key, large item-by-country interactions, and statistics indicating lack of fit with the model), these were a result of translation, adaptation, or printing errors.

1.14 Scaling the TIMSS Achievement Data

Deriving reliable student achievement scores from a large-scale assessment measuring trends over time like TIMSS poses a difficult challenge. Firstly, because of the ambitious coverage goals of TIMSS 2003, there was not enough testing time for a student to complete the entire assessment, and so a matrix-sampling design was adopted whereby each student's test booklet contained just a part of the assessment. Although this solved the problem of administering the assessment, it complicated the calculation of student achievement scores, since not all students took the same set of items, and the items that students did take were not all equally difficult. Secondly, in measuring trends over time (1995, 1999, 2003, and so on), it was not possible for TIMSS to keep reusing the same mathematics and science achievement items. In order to keep the assessment at the cutting edge of mathematics and science education, it was necessary to replace older items with new material at each cycle. In addition, TIMSS has a policy of publishing a large proportion of the items used in each assessment so that educators, policy makers, and the public may have a good understanding of the mathematics and science addressed by the assessment. Accordingly, the composition of the assessment evolves at each assessment cycle, as items are published and used for illustrative purposes and new items are developed to replace the published items. This further complicated the calculation of student achievement scores.

To meet the challenge of estimating student achievement, TIMSS relies primarily on item response theory (IRT) scaling methods. With IRT scaling, students' scores do not depend on taking the same set of items, and so this methodology is particularly useful when different blocks of items and different samples of students have to be linked. This being the case, IRT methodology was preferred by TIMSS for developing comparable estimates of performance for all students, since students answered different test items depending upon which of the 12 test booklets they received. The IRT analysis provides a common scale on which performance can be compared across countries. In addition to providing a basis for estimating mean achievement, scale scores permit estimates of how students within countries vary and provide information on percentiles of performance.

In TIMSS 2003, 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 generalized partial credit models for items with 0, 1, or 2 available score points. The IRT scaling method produces a score by averaging the responses of each student to the items that he or she took in a way that takes into account the difficulty and discriminating power of each item. As with any method of scaling student achievement, measurement is most reliable when a student responds to a large number of items, and is less reliable when the number of items is small. In the matrix-sampling approach adopted by TIMSS, with each student responding to a limited number of items, and given TIMSS' ambitious reporting goals – scales for two subjects (mathematics and science) and for five content domains in each subject – each student may respond to just a few items related to a particular scale.

To improve reliability, the TIMSS scaling methodology draws on information about students' background characteristics as well as their responses to the achievement items. This approach, known as "conditioning," enables reliable scores to be produced even though individual students responded to relatively small subsets of the total mathematics or science item pool. Rather than estimating student scores directly, TIMSS combines information about item characteristics, student responses to the items that they took, and student background information to estimate student achievement distributions. Having determined the overall achievement distribution, TIMSS estimates each student's achievement conditional on the student's responses to the items that they took and the student's background characteristics. To account for error in this imputation process, TIMSS draws five such estimates, or "plausible values," for each student on each of the scales, and incorporates the variability between the five estimates in the standard error of any statistics reported.

The TIMSS mathematics and science achievement scales were designed to provide reliable measures of student achievement spanning 1995, 1999, and 2003. The metric of the scale was established originally with the 1995 assessment. Treating equally all the countries that participated in 1995 at the eighth grade, the TIMSS scale average over those countries was set at 500 and the standard deviation at 100. The same applied for the fourth-grade assessment. Since the countries varied in size, each country was weighted to contribute equally to the mean and standard deviation of the scale. The average and standard deviation of the scale scores are arbitrary and do not affect scale interpretation. To preserve the metric of the original 1995 scale, the 1999 eighth-grade assessment was scaled using students from the countries that participated in both 1995 and 1999. Then students from the countries that tested in 1999 but not 1995 were assigned scores on the basis of the scale.

At the eighth grade, TIMSS developed the 2003 scale in the same way as in 1999, preserving the metric first with students from countries that participated in both 1999 and 2003, and then assigning scores on the basis of the scale to students tested in 2003 but not the earlier assessment. At fourth grade, because there was no assessment in 1999, the 2003 and 1995 data were linked directly together using students from countries that participated in both assessments, and the students tested in 2003 but not 1995 were assigned scores on the basis of the scale.

In addition to the scales for mathematics and science overall, TIMSS created IRT scales for each of the mathematics and science content domains for the 2003 data. These included number, algebra, measurement, geometry, and data in mathematics; and life science, chemistry, physics, earth science, and environmental science in science.[2] However, insufficient common items were used in 1995 and 1999 to establish reliable IRT content area scales for trend purposes.

1.15 Data Analysis and Reporting

The TIMSS 2003 International Mathematics Report (Mullis, Martin, Gonzalez, and Chrostowski, 2004) and the TIMSS 2003 International Science Report (Martin, Mullis, Gonzalez, and Chrostowski, 2004) summarize fourth- and eighth- grade students' mathematics and science achievement, respectively, in each participating country. The reports present trend results from 1995 and 1999 at the eighth grade, as well as from 1995 for the fourth grade. Average achievement is reported separately for girls and for boys.

To provide additional information about mathematics and science achievement among high- and low-achieving students, TIMSS reported the percentage of students in each country performing at each of four international benchmarks of student achievement. Selected to represent the range of performance shown by students internationally, the advanced benchmark was 625, the high benchmark was 550, the intermediate benchmark was 475,

2 At the fourth grade, scales were constructed only for life science, physical science, and earth science.

and the low benchmark was 400. Although the fourth- and eighth-grade scales are different, the same benchmark points were used at both grades. To enhance this reporting approach, TIMSS conducted a scale anchoring analysis to describe achievement of students at those four points on the scales. Scale anchoring is a way of describing students' performance at different points on a scale in terms of what they know and can do. It involves a statistical component, in which items that discriminate between successive points on the scale are identified, and a judgmental component, in which subject-matter experts examine the items and generalize to students' knowledge and understandings. Complementing this approach further, the TIMSS 2003 International Reports present examples of mathematics and science items that anchor at each of the benchmarks, and display student performance in each country on the example items.

TIMSS 2003 collected a wide array of information about the homes, schools, classrooms, and teachers of the participating students, as well as about the mathematics and science curriculum in each country. The TIMSS 2003 International Reports summarize much of this information, combining data into composite indices showing an association with achievement where appropriate. In particular, student mathematics and science achievement is described in relation to characteristics of the home, curriculum coverage, classroom instruction, and school environment.

Because the statistics presented in the international reports are estimates of national performance based on samples of students, rather than the values that could be calculated if every student in every country had answered every question, it is important to have measures of the degree of uncertainty of the estimates. The jackknife procedure was used to estimate the standard error associated with each statistic presented in this report. The jackknife standard errors also include an error component due to variation among the five plausible values generated for each student. The use of confidence intervals, based on the standard errors, provides a way to make inferences about the population means and proportions in a manner that reflects the uncertainty associated with the sample estimates. An estimated sample statistic plus or minus two standard errors represents a 95 percent confidence interval for the corresponding population result.

References

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., O'Connor, K.M., Chrostowski, S.J., Gregory, K.D., Garden, R.A., and Smith, T.A. (2001), *Mathematics Benchmarking Report TIMSS 1999 – Eighth Grade: Achievement for U.S. States and Districts in an International Context*. Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzalez, E.J., Chrostowski, S.J, and O'Connor, K.M. (2003), *TIMSS Assessment Frameworks and Specifications 2003 (2nd Edition)*, Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., and Chrostowski, S.J. (2004), *TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: Boston College.

Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., O'Connor, K.M., Chrostowski, S.J., Gregory, K.D., Smith, T.A., and Garden, R.A. (2001), *Science Benchmarking Report TIMSS 1999 – Eighth Grade: Achievement for U.S. States and Districts in an International Context*. Chestnut Hill, MA: Boston College.

Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., and Chrostowski, S.J. (2004), *TIMSS 2003 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: Boston College.

2

Chapter 2

Developing the TIMSS 2003 Mathematics and Science Assessment and Scoring Guides

Teresa Smith Neidorf and Robert Garden

2.1 Overview

The development of the TIMSS 2003 mathematics and science assessment was a collaborative process spanning a two-and-a-half-year period, from September 2000 to March 2003, and involving mathematics and science educators and development specialists from all over the world. The work began with a major updating and revision of the existing TIMSS assessment frameworks to address changes during the last decade in curricula and the way mathematics and science are taught (Mullis, Martin, Smith, Garden, Gregory, Gonzalez, Chrostowski, & O'Connor, 2003). The assessment development work was based firmly on the new assessment frameworks and specifications.

Meeting the specifications of the TIMSS 2003 assessment frameworks required a large number of new mathematics and science items to be developed at both fourth and eighth grades. With support and training from the TIMSS International Study Center, National Research Coordinators (NRCs), from participating countries contributed a large pool of items for review and field testing. The International Study Center established two task forces, one in mathematics and one in science,[1] to manage the item development process. To help review, select, and revise items for the assessment and to ensure their mathematical and scientific accuracy, the International Study Center convened the Science and Mathematics Item Review Committee (SMIRC), an international committee of prominent mathematics and science experts

1 The mathematics task force consisted of Robert Garden, TIMSS Mathematics Coordinator, Chancey Jones of Educational Testing Service in the United States, and Graham Ruddock of the National Foundation for Educational Research in England. The science task force consisted of Teresa Smith Neidorf, TIMSS Science Coordinator, Christine O'Sullivan, formerly the Science Coordinator for the U.S. National Assessment for Educational Progress (NAEP), and Svein Lie, University of Oslo, formerly the chair of the TIMSS 1995 Subject Matter Advisory Committee.

nominated by participating countries and representing a range of nations and cultures.[2]

Since the test items were developed in English and translated into 34 languages by the participating countries, both the SMIRC and NRCs were important in identifying any items that might prove difficult to translate consistently.

To ensure that TIMSS 2003 reflects an international perspective, both the framework and test development procedures included substantial contributions from the international community. Exhibit 2.1 provides an overview of the process. This chapter describes the steps taken in developing the TIMSS 2003 mathematics and science assessment, with sections covering frameworks development, the mathematics and science assessment specifications, development of mathematics and science items and scoring guides, and the assessment booklet design.

2.2 Developing the TIMSS 2003 Assessment Frameworks

For the TIMSS 2003 assessment, the curriculum frameworks used as the basis for the 1995 and 1999 TIMSS assessments (Robitaille, McKnight, Schmidt, Britton, Raizen, & Nicol, 1993) were extensively revised and updated. This effort was conducted by the TIMSS International Study Center at Boston College in collaboration with the National Research Coordinators of the TIMSS countries and with guidance from an international Expert Panel. The Expert Panel was made up of 29 internationally recognized experts and included mathematicians and scientists, curriculum experts, and educational practitioners, researchers, and assessment specialists.[3]

The framework development process took approximately one year, beginning in September 2000. Work to update the frameworks began with a review of the TIMSS 1999 curriculum data to identify mathematics and science topics emphasized in the curricula of the TIMSS countries. In addition, a survey of NRCs of more than 20 countries planning to participate in TIMSS 2003, administered in September 2000, provided recommendations for the percentage of the TIMSS 2003 assessment to be devoted to each mathematics and science content area at fourth and eighth grades and to identify any curriculum areas that should receive greater or less emphasis than in the TIMSS 1999 assessment. The TIMSS International Study Center used the results of the review and survey to prepare initial framework discussion documents for the First Expert Panel Meeting.

2 See Appendix A for a list of the members of the Science and Mathematics Item Review Committee.

3 See Appendix A for a list of the members of the Expert Panel.

Exhibit 2.1 **Overview of the TIMSS 2003 Framework and Test Development Process**

Date(s)		Group and Activity
September	2000	National Research Coordinators Complete preliminary survey on recommended coverage of mathematics and science content in the TIMSS 2003 assessment.
September – October	2000	TIMSS International Study Center Review results from TIMSS 1999 curriculum questionnaires and TIMSS 2003 NRC survey; prepare initial framework discussion documents for First Expert Panel Meeting.
November	2000	First Expert Panel Meeting (Boston) Make recommendations for coverage of major content and cognitive domains, updated assessment topics within content areas, and initial draft of TIMSS mathematics and science assessment frameworks.
December	2000	TIMSS International Study Center Develop detailed specifications for assessment topics at fourth and eighth grades and prepare first draft of TIMSS assessment frameworks.
February	2001	First National Research Coordinators Meeting (Hamburg) Review first draft of TIMSS assessment frameworks.
March – April	2001	National Research Coordinators Complete survey of Mathematics and Science Curriculum Topics.
April – May	2001	TIMSS International Study Center Compile TIMSS 2003 Mathematics and Science Curriculum Topics survey results and prepare second draft of frameworks.
May	2001	Second Expert Panel Meeting (Amsterdam) Review and approve second draft of TIMSS 2003 Assessment Frameworks incorporating revisions from the first NRC meeting and results of mathematics and science frameworks topic survey completed by National Research Coordinators; generate preliminary ideas for problem-solving and inquiry tasks.
June	2001	Second National Research Coordinators Meeting (Montreal) Review and approve final draft of TIMSS 2003 Assessment Framework. TIMSS Item-Writing Workshop
June – July	2001	National Research Coordinators Develop and submit items to the International Study Center.
August – September	2001	Mathematics and Science Task Forces Assemble, review and revise international item pool; develop additional items to cover framework.
September	2001	TIMSS International Study Center Publish first edition of the TIMSS Assessment Frameworks and Specifications 2003.
September	2001	First Science and Mathematics Item Review Committee Meeting (Boston) Review/refine international item pool; generate prototype ideas for problem-solving and inquiry tasks.

Exhibit 2.1 **Overview of the TIMSS 2003 Framework and Test Development Process**
(...Continued)

Date(s)		Group and Activity
October	2001	Second Science and Mathematics Item Review Committee Meeting (Portsmouth) Review, revise and select preferred and alternate items for field test; develop problem-solving and inquiry tasks.
October – December	2001	Mathematics and Science Task Forces Assemble draft field test item blocks and problem-solving and inquiry tasks.
December	2001	Third National Research Coordinators Meeting (Madrid) Review and approve field test item blocks and problem-solving and inquiry tasks.
December – January	2002	TIMSS International Study Center Conduct teacher review of problem-solving and inquiry tasks; incorporate final revisions to items and tasks based on NRC and teacher reviews.
January – February	2002	TIMSS International Study Center Conduct small-scale item trial of constructed-response items and problem-solving and inquiry tasks; distribute field test instruments; update scoring guides; prepare field test scoring training materials.
February – April	2002	National Research Coordinators: Translate field test instruments. IEA: Verify field test translations.
March	2002	Fourth National Research Coordinators Meeting (Ghent) Field test scoring training
April	2002	U.S. National Center for Education Statistics Cognitive Laboratory Investigation of Problem-Solving and Inquiry Tasks
April – June	2002	Field test administration
June – July	2002	TIMSS International Study Center Review field test item statistics; revise problem-solving and inquiry tasks; assemble draft main survey item blocks.
July	2002	Third Science and Mathematics Item Review Committee Meeting (Oslo) Review field test results and draft item blocks and scoring guides for main survey.
August	2002	Fifth National Research Coordinators Meeting (Tunis) Review and approve item blocks and scoring guides for main survey.
September	2002	TIMSS International Study Center Conduct small-scale trial of final problem-solving and inquiry tasks; distribute main survey instruments; update main survey scoring guides; prepare main survey scoring training materials.
September – October	2002	National Research Coordinators (southern hemisphere): Translate main survey test instruments. IEA: Verify main survey translations for southern hemisphere countries.
October – December	2002	Main survey administration in southern hemisphere countries
November	2002	Southern hemisphere scoring training for the main survey (Wellington)

Exhibit 2.1 **Overview of the TIMSS 2003 Framework and Test Development Process**
(...Continued)

Date(s)		Group and Activity
December	2002	TIMSS International Study Center Update TIMSS Assessment Frameworks and Specifications document to include example items and from the field test and a revised test booklet design; distribute final version of main survey scoring guides.
December – March	2003	National Research Coordinators (northern hemisphere): Translate main survey test instruments. IEA: Verify main survey translations.
February	2003	Publish second edition of the TIMSS Assessment Frameworks and Specifications 2003.
March	2003	Sixth National Research Coordinators Meeting (Bucharest) Northern hemisphere scoring training for the main survey
March – June	2003	Main survey administration in northern hemisphere countries

At its first meeting, in November 2000, the Expert Panel made recommendations concerning how the assessment time in TIMSS 2003 should be distributed across the mathematics and science content areas at each grade level; made suggestions for the major assessment topics that should be included; and discussed calculator usage and the inclusion of scientific inquiry in the 2003 assessment. They also discussed how the "performance expectations" aspect of the original frameworks might be reformulated as a set of broadly-defined cognitive domains for mathematics and science. Maintaining alignment of the TIMSS 2003 content domains with the reporting categories in TIMSS 1995 and 1999 and the measurement of trend were important considerations. Following the meeting, the International Study Center prepared a first draft of the mathematics and science assessment frameworks incorporating the recommendations of the Expert Panel for review by the NRCs.

The first draft of the frameworks contained initial recommendations for the distribution of the assessment across content and cognitive domains and specific assessment objectives for a broad range of mathematics and science topics at the fourth and eighth grade. The draft frameworks document was reviewed and discussed at the first meeting of TIMSS 2003 National Research Coordinators in February 2001, with representatives from more than 40 countries. Some adjustments to the distributions of assessment time across content and cognitive domains were made in response to NRC suggestions. NRCs also gave input on the appropriateness of the assessment topics for the populations of students being assessed.

Following the meeting, four extensive assessment topic questionnaires (mathematics and science at fourth and eighth grades) were distributed to each NRC to be completed with the assistance of experts in mathematics and science curriculum in each country. The questionnaires asked countries to indicate for every mathematics and science assessment topic in the draft frameworks: i) if the topic is addressed by their curriculum at the appropriate grade level, and ii) whether the topic should be included in the TIMSS 2003 international assessment (even if the topic has not been included in their curriculum by that grade level). Results were obtained from 36 countries and were used to refine the set of topics in the frameworks, focusing on those that were included in the curricula or recommended for inclusion in TIMSS by a significant number of participating countries. In the retained set, nearly all topics were included in the curricula of the majority of countries, and for many topics in more than 90 percent of countries.

A second draft of the frameworks document, incorporating the results of the NRC survey, was further refined and improved at the Second Expert Panel Meeting in May 2001. In July 2001, the International Study Center distributed a third draft of the frameworks to the Expert Panel members and the NRCs for review. Comments and suggestions on this draft were incorporated into the final version, *TIMSS Assessment Frameworks and Specifications 2003* (Mullis et al., 2001), published September 2001. A second edition of the frameworks, incorporating example mathematics and science items from the field test and a revised test booklet design, was published in February 2003 (Mullis et al., 2003). During the process of updating the TIMSS assessment frameworks for 2003, the expert panelists and national representatives reaffirmed the importance of emphasizing problem solving, reasoning and inquiry in the outcomes to be assessed, and this is reflected in the final version of the frameworks.

2.3 Mathematics Assessment Framework and Specifications

The mathematics assessment framework for TIMSS 2003 is framed by two organizing dimensions, a content dimension and a cognitive dimension, analogous to those used in the earlier TIMSS assessments. There are five content domains: number, algebra, measurement, geometry, and data. There are four cognitive domains: knowing facts and procedures, using concepts, solving routine problems, and reasoning. The two dimensions and their domains are the foundation of the mathematics assessment. The content domains define the specific mathematics subject matter covered by the assessment, and the cognitive domains define the sets of behaviors expected of students as they engage with the mathematics content. Exhibit 2.2 shows the target percentages of the total mathematics assessment time to be devoted to each of the content and cognitive domains at fourth and eighth grades.

Exhibit 2.2 **Target Percentages of TIMSS 2003 Mathematics Assessment Devoted to Content and Cognitive Domains by Grade Level**

	Grade 4	Grade 8
Mathematics Content Domains		
Number	40%	30%
Algebra*	15%	25%
Measurement	20%	15%
Geometry	15%	15%
Data	10%	15%
Mathematics Cognitive Domains		
Knowing Facts and Procedures	20%	15%
Using Concepts	20%	20%
Solving Routine Problems	40%	40%
Reasoning	20%	25%

* At fourth grade, the algebra content domain is called patterns and relationships.

2.3.1 Content Domains

For each of the five content domains, the mathematics framework identifies several topic areas to be included in the assessment, as shown in Exhibit 2.3. For example, *number* is further categorized by *whole numbers, fractions and decimals, integers,* and *ratio, proportion, and percent.* Each topic area is presented as a list of objectives covered in a majority of participating countries, at either fourth or eighth grade. The organization of topics across the content domains reflects some minor revision in the reporting categories used in the 1995 and 1999 assessments. However, each of the trend items from 1995 and 1999 may be mapped directly into the content domains defined for 2003.

2.3.2 Cognitive Domains

To respond correctly to TIMSS test items, students need to be familiar with the mathematics content of the items. Just as important, however, items were designed to elicit the use of particular cognitive skills. The assessment framework presents detailed descriptions of the skills and abilities that make up the cognitive domains and that will be assessed in conjunction with the content. These skills and abilities should play a central role in developing items and achieving a balance in learning outcomes assessed by then items in fourth and eighth grades. The student behaviors used to define the mathematics framework have been classified into four cognitive domains, as follows:

Exhibit 2.3 Main Topics Included in the Mathematics Content Domains

Content Domains	Main Topics
Number	Whole numbers Fractions and decimals Integers (grade 8 only) Ratio, proportion, and percent
Algebra	Patterns Algebraic expressions (grade 8 only) Equations and formulas Relationships
Measurement	Attributes and units Tools, techniques, and formulas
Geometry	Lines and angles Two- and three-dimensional shapes Congruence and similarity Locations and spatial relationships Symmetry and transformations
Data	Data collection and organization Data representation Data interpretation Uncertainty and probability (grade 8 only)

Knowing Facts and Procedures: *Facts* encompass the factual knowledge that provide the basic language of mathematics and the essential mathematical facts and properties that form the foundation for mathematical thought. *Procedures* form a bridge between more basic knowledge and the use of mathematics for solving routine problems, especially those encountered by people in their daily lives. Students need to be efficient and accurate in using a variety of computational procedures and tools.

Using Concepts: Familiarity with mathematical concepts is essential for the effective use of mathematics for problem solving, for reasoning, and thus for developing mathematical understanding. Knowledge of concepts enables students to make connections between elements of knowledge, make extensions beyond their existing knowledge, and create mathematical representations.

Solving Routine Problems: Problem solving is a central aim of teaching school mathematics and features prominently in school mathematics textbooks. Routine problems may be standard in classroom exercises designed to provide practice in particular methods or techniques. Some of these problems may be set in a quasi-real context, and may involve extended knowledge of mathematical properties (e.g., solving equations). Though they range in difficulty, routine problems are expected to be sufficiently familiar to students that they essentially involve selecting and applying learned procedures.

Reasoning: Mathematical reasoning involves the capacity for logical, systematic thinking. It includes intuitive and inductive reasoning based on patterns and regularities that can be used to arrive at solutions to non-routine problems, i.e., problems very likely to be unfamiliar to students. Such problems may be purely mathematical or may have real-life settings, and involve application of knowledge and skills to new situations, with interactions among reasoning skills usually a feature.

Examples of the behaviors associated with each of the cognitive domains may be found in Mullis et al. (2003).

2.3.3 Communicating Mathematically

Communicating mathematical ideas and processes is important for many aspects of living and fundamental to the teaching and learning of mathematics. In the TIMSS framework, communication is not a separate cognitive domain but rather an overarching dimension across all mathematics content areas and processes. Communication is fundamental to each of the four TIMSS cognitive domains (*knowing facts and procedures, using concepts, solving routine problems, and reasoning*), and students' communication in and about mathematics should be regarded as assessable in each of these areas. Students in TIMSS may demonstrate communication skills through description and explanation, such as describing or discussing a mathematical object, concept, or model. Communication also occurs in using mathematical terminology and notation, demonstrating the procedure used in solving an equation, or using particular representational modes to present mathematical ideas.

2.3.4 Calculator Policy

The TIMSS policy on calculator use at the eighth grade is to give students the best opportunity to operate in settings that mirror their classroom experience. Beginning with 2003, calculators were permitted but not required for newly-developed eighth-grade assessment materials. Participating countries could decide whether or not their students were allowed to use calculators for the new items. Since calculators were not permitted at the eighth grade in the 1995 or 1999 assessments, the 2003 eighth-grade test booklets were designed so that items from these assessments were placed in the first half and items new in 2003 placed in the second half. Where countries chose to permit eighth-grade students to use calculators, they could use them for the second half of the booklet only. For the fourth-grade assessment, TIMSS 2003 continued the 1995 policy of not permitting calculator use.

2.4 Science Assessment Framework and Specifications

The science assessment framework for TIMSS 2003, like the mathematics framework, is framed by two organizing dimensions, a content dimension and a cognitive dimension. There are five content domains: life science, chemistry, physics, earth science, and environmental science, and three cognitive domains: factual knowledge, conceptual understanding, and reasoning and analysis. Exhibit 2.4 shows the target percentages of the total science assessment time to be devoted to each of the science content and cognitive domains for fourth and eighth grades. In contrast to TIMSS 1999, where a separate reporting category of "Scientific Inquiry and the Nature of Science" was included, the TIMSS 2003 framework treats scientific inquiry as a separate assessment strand that overlaps all of the fields of science and has both content- and skills-based components. Although scientific inquiry is not treated as a separate reporting category in TIMSS 2003, the framework specifies that outcomes related to scientific inquiry will represent up to 15 percent of the total science assessment time at each grade level to permit some level of reporting student performance in this area. Further descriptions of the assessment specifications for the content domains, cognitive domains, and scientific inquiry assessment strand are provided in the following sections.

Exhibit 2.4 **Target Percentages of TIMSS 2003 Science Assessment Devoted to Content and Cognitive Domains by Grade Level**

	Grade 4	Grade 8
	Science Content Domains	
Life Science	45%	30%
Physical Science	35%	*
Chemistry	*	15%
Physics	*	25%
Earth Science	20%	15%
Environmental Science	*	15%
	Science Cognitive Domains	
Factual Knowledge	40%	30%
Conceptual Understanding	35%	35%
Reasoning and Analysis	25%	35%

* At fourth grade, Physical Science included Physics and Chemistry topics. Also, a few Environmental Science topics that addressed the use of conservation of natural resources and changes in environments were included in Earth Science and Life Science.

2.4.1 Content Domains

For each of the science content domains, the framework identifies several main topic areas that are to be included in the assessment as shown in Exhibit 2.5. Most of the main topics are appropriate for both grades, but some topics are included at the eighth grade only, as indicated. For each main topic area, the frameworks document includes a list of specific subtopics or assessment objectives appropriate for each grade level. This structure of the frameworks highlights the development of knowledge and abilities across the grades.

Exhibit 2.5 **Main Topics Included in the Science Content Domains**

Content Domain	Main Topics
Life Science	Types, characteristics, and classification of living things
	Structure, function, and life processes in organisms
	Cells and their functions (grade 8 only)
	Development and life cycles of organisms
	Reproduction and heredity
	Diversity, adaptation, and natural selection
	Ecosystems
	Human health
Chemistry	Classification and composition of matter
	Particulate structure of matter (grade 8 only)
	Properties and uses of water
	Acids and bases (grade 8 only)
	Chemical change
Physics	Physical states and changes in matter
	Energy types, sources and conversions
	Heat and temperature
	Light
	Sound and vibration (grade 8 only)
	Electricity and magnetism
	Forces and motion
Earth Science	Earth's structure and physical features
	Earth's processes, cycles and history
	Earth in the solar system and the universe
Environmental Science	Changes in population (grade 8 only)
	Use and conservation of natural resources
	Changes in environments

2.4.2 Cognitive Domains

The set of skills and abilities to be demonstrated by students in responding to items across the science topics is organized into the three broad cognitive domains specified in the framework – factual knowledge, conceptual understanding, and reasoning and analysis. The exact nature of behaviors elicited by the TIMSS items in each of these categories varies between fourth and eighth grade in accordance with the increased cognitive ability, maturity, instruction, experience, and conceptual understanding of students at the higher grade level. A brief description of each cognitive domain and the set of skills and abilities required by TIMSS items corresponding to each are listed below.

Factual Knowledge: This refers to students' knowledge base of relevant science facts, information, tools, and procedures. Items may require students to recall/recognize accurate statements about science facts and concepts; demonstrate knowledge/use of correct scientific terms; describe scientific processes, properties, characteristics, structure, function, and relationships; and demonstrate knowledge about the use of scientific tools and procedures.

Conceptual Understanding: Students should be able to demonstrate a grasp of the relationships that explain the physical world and relate the observable to more abstract or general concepts. Items may require students to provide examples to illustrate general concepts; compare/contrast and classify objects, materials and organisms; use diagrams/models; relate underlying concepts to observed or inferred properties/behaviors; extract/apply textual, tabular or graphical information; find solutions to problems involving the direct application of concepts; and provide explanations.

Reasoning and Analysis: This includes problem-solving and scientific reasoning processes involved in the more complex tasks related to science. Items may require students to analyze/interpret problems; integrate/synthesize a number of factors or related concepts across mathematics and science; hypothesize/predict; design investigations and procedures; analyze/interpret data; draw conclusions; generalize; evaluate; and justify explanations and problem solutions.

2.4.3 Scientific Inquiry

The scientific inquiry strand is assessed through longer problem-solving and inquiry tasks as well as some individual items that require students to apply scientific inquiry skills in a practical context. While not full scientific investigations, the tasks are designed to require a basic understanding of the nature of science and investigation and elicit some of the skills essential to the scientific inquiry process. Tasks may include some portion of the following major phases in the scientific inquiry process:

- Formulating questions and hypotheses
- Designing investigations
- Collecting, representing, analyzing, and interpreting data
- Drawing conclusions and developing explanations based on evidence

The same general assessment outcomes related to scientific inquiry are appropriate for both fourth and eighth grades, but the specific understandings and abilities to be demonstrated increase in complexity across grades. The items and tasks developed to measure scientific inquiry skills are set in content-based contexts. These items are, therefore, classified with respect to content and cognitive categories as well as scientific inquiry and will contribute to the appropriate content reporting scale.

2.5 Developing Mathematics and Science Items and Scoring Guides

Test development for TIMSS 2003 involved developing a set of items aligned with the *TIMSS Assessment Frameworks and Specifications* in each mathematics and science content and cognitive domain. In addition to the target percentages of assessment time to be devoted to the mathematics and science content and cognitive domains, the frameworks give guidelines for the distribution of testing time across item formats, specifying that at least one-third of the assessment should come from constructed-response items. Since approximately half of the eighth-grade items from TIMSS 1999 and one-third of the fourth-grade items from TIMSS 1995 had been kept secure and were to be included, these trend items were taken into account in allocating the test development effort to the different assessment areas for TIMSS 2003. Item development blueprints, specifying the approximate number of mathematics and science items to be developed in each content area in the frameworks, formed the basis for test development for TIMSS 2003. These blueprints were created by:

- estimating the number of items needed in the final test based on the total score points and percentage of score points in each content domain specified in the frameworks,
- distributing this number of items across the mathematics and science main topic areas in accordance with their breadth of content,
- accounting for the number of trend items already included in each topic area,
- ensuring coverage of the cognitive domains and appropriate numbers of multiple-choice and constructed-response items, and
- scaling up the number of items to be developed to allow for attrition during the item selection and field-testing process.

This section describes the test development procedure, including the consideration of trend items, development of the international item pool including problem-solving and inquiry tasks, item review and revision, field testing, item selection for the main survey, and the development of scoring guides for the constructed-response items.

2.5.1 Trend Items

In developing the TIMSS 2003 test blueprints, the trend items from 1995 and 1999 were mapped into the content and cognitive categories in the new 2003 frameworks. As shown in Exhibits 2.6 and 2.7, the mathematics and science trend items cover a range of content domains at both grades. Eighth-grade trend items include both multiple-choice and constructed-response items, while fourth-grade trend items are nearly all multiple-choice. Therefore, a larger proportion of constructed-response items needed to be developed for grade 4.

Exhibit 2.6 **Mathematics Trend Items at Grade 4 and Grade 8 by Content Domain and Item Format**

	Grade 4 Trend Items			Grade 8 Trend Items		
Content Domain	**Multiple Choice**	**Constructed Response**	**Total**	**Multiple Choice**	**Constructed Response**	**Total**
Number	19	0	19	19	6	25
Algebra*	2	0	2	11	5	16
Measurement	8	0	8	8	8	16
Geometry	4	0	4	11	1	12
Data	4	0	4	10	0	10
Total	**37**	**0**	**37**	**59**	**20**	**79**

* Called Patterns and Relationships at Grade 4.

Exhibit 2.7 **Science Trend Items at Grade 4 and Grade 8 by Content Domain and Item Format**

	Grade 4 Trend Items			Grade 8 Trend Items		
Content Domain	**Multiple Choice**	**Constructed Response**	**Total**	**Multiple Choice**	**Constructed Response**	**Total**
Life Science	11	1	12	12	5	17
Physical Science	9	0	9	--	--	--
Chemistry	--	--	--	13	1	14
Physics	--	--	--	14	8	22
Earth Science	11	1	12	10	2	12
Environmental Science	--	--	--	4	5	9
Total	**31**	**2**	**33**	**53**	**21**	**74**

2.5.2 Developing the International Item Pool for TIMSS 2003

Test development for TIMSS 2003 was an international collaborative process, involving participants from more than 30 countries. To maximize the effectiveness of the contributions from national centers, the International Study Center developed a detailed item-writing manual and conducted a workshop for countries that wished to provide items for the international item pool. At this workshop, the mathematics and science task forces reviewed general item-writing guidelines for multiple-choice and constructed-response items and provided specific training in writing mathematics and science items in accordance with the *TIMSS Assessment Frameworks and Specifications 2003*. After the training sessions, participants were organized into item-writing subgroups by mathematics and science content domains for the development and review of items. Nearly 200 draft items were developed at the item-writing workshop.

Following the workshop, national centers developed additional items in mathematics and/or science for the fourth or eighth grade in accordance with their interest and capacity. To maximize contributions from international item writers and ensure adequate item development in the appropriate mathematics and science content areas, some specifications were given by the International Study Center to focus item development in areas not already covered by the trend items. Draft items were submitted by the national centers to the International Study Center, which coordinated the contributions from participating countries and managed the overall test development and review process to ensure that the TIMSS tests were aligned with the assessment frameworks.

Each item from the national centers was submitted with an item-writing form that identified the portion of the framework that the item was designed to assess – content domain, main topic and specific assessment objective, and the primary cognitive domain. Science items also were designated as to whether or not they were intended to measure knowledge and skills associated with the scientific inquiry strand. This development process resulted in an initial item pool of more than 1300 items across both grades, with contributions from 35 countries covering a broad range of mathematics and science topics.

2.5.3 Item Review and Revision

The mathematics and science task forces assembled, reviewed, and revised the draft items submitted by participating countries and confirmed the classification of items with respect to the frameworks. They also developed additional items for areas of the frameworks not well covered by the country submissions. The resultant item pool of more than 2000 items covered a wide array of topics in the mathematics and science content domains at each grade level and reflected the range of cognitive domains and item types specified in the frameworks. The task forces then made a preliminary selection from among these draft items for review by the Science and Mathematics Item Review Committee (SMIRC).

The SMIRC conducted its initial item review work in two meetings, the first in September and the second in October 2001. Working from test development blueprints identifying the number of items needed in each content domain, the SMIRC made much progress in choosing among alternative items, refining the most promising items, and supplementing this set of items in content areas lacking coverage.

Between the second SMIRC meeting and the third NRC meeting in December 2001, the mathematics and science task forces continued the work of developing, reviewing and revising the items for the field test. The draft field test items were organized into a set of “preferred” and “alternate” item blocks based on input from the SMIRC (see section 2.5.5). At the third NRC meeting, the “preferred” item blocks were reviewed in plenary with all NRCs. The “alternate” item blocks were made available for review in separate review sessions, and NRCs provided feedback on these items in comment sheets. Both “preferred” and “alternate” items were subsequently revised in line with suggestions received from NRCs. In general, the items for the field test were well received, and NRCs were satisfied that the items consituted a very satisfactory field test item pool.

2.5.4 Developing the Problem-Solving and Inquiry Tasks

To address the importance placed in the frameworks on the assessment of problem-solving, reasoning and scientific inquiry, a set of tasks were developed to assess how well students can draw on and integrate a variety of processes and understandings in mathematics and science to conduct investigations and solve problems. At the first NRC meeting, it was decided that from an operational perspective, it was important that the tasks developed for TIMSS 2003 be less demanding to administer than the performance assessment conducted in TIMSS 1995. Specifically, the tasks needed to be self-contained, involve minimal equipment, and be integrated into the main test administration without any special accommodations or additional testing sessions. Thus, a major challenge for TIMSS 2003 was to develop a set of relevant problem-solving and inquiry tasks that would satisfy the requirements set forth by the Expert Panel and the national representatives.

The development of tasks was an evolutionary process, starting with a "brainstorming" session of international mathematics and science experts at the second Expert Panel meeting in May 2001. The expert panel developed a number of innovative prototype ideas for investigative or "real-world" tasks, many of which integrated ideas across mathematics and science. At the TIMSS item-writing workshop for participating countries, the approach to developing problem-solving and inquiry tasks was discussed, and some ideas were submitted by national centers as part of the international item development process. Much of the development for the problem-solving and inquiry tasks occurred at meetings of the Science and Mathematics Item Review Committee. At the first SMIRC meeting in September 2001, the initial set of ideas for tasks put forth by the Expert Panel and submitted from national centers were discussed for their appropriateness and feasibility, and considerable progress was made in drafting tasks using these ideas as a starting point. The ISC staff and a subset of SMIRC members further refined the initial drafts in the following few weeks, and these first drafts were presented to the full SMIRC at their second meeting in October 2001. The drafts were reviewed, and a subset was selected and substantially elaborated, at the second SMIRC meeting. Additional tasks were developed to ensure that the set of tasks covered a range of content areas in mathematics and science.

Following the second meeting of the SMIRC, ISC staff and the mathematics and science task force members continued to work on the tasks and prepare them for the presentation and review of field test materials at the second NRC meeting. A number of modifications recommended by the NRCs were incorporated following the meeting. It was suggested at the NRC meeting that the accessibility, reading level, and appropriateness of content and terminology for fourth- and eighth-grade students be further evaluated, particularly for the science tasks. To address this concern, the ISC recruited two experienced fourth-grade and eighth-grade science teachers in the Boston area to review the revised science tasks. The feedback from the teachers was very positive overall, indicating that most of the content was now grade appropriate and the tasks were interesting and engaging. After some revisions in layout, content, and language based on the results of the teacher review, a small-scale item pilot of the problem-solving and inquiry tasks and other constructed-response items was conducted in February 2002 in seven countries that tested in English. This pilot yielded a total of approximately 4500 student responses at each grade level, with 30 to 40 responses to each task. The results of this international pilot provided valuable information about how the tasks functioned internationally and were used primarily to refine the scoring guides and obtain student responses for use in preparing scoring training materials for the field test.

A total of 19 tasks (9 at fourth grade and 10 at eighth grade) were selected for the field test. Each task included a series of related test items, mostly constructed-response, that were linked by a common theme and involved an investigation or extended problem-solving situation. Some of the mathematics tasks involved manipulatives such as cardboard rulers or geometric tiles; no equipment or manipulatives were required for the science tasks. Although some of the initial ideas for science tasks involved the use of equipment, during further development stages it was decided that the type of equipment required was not feasible in the test administration setting. Each of the tasks in the field test was designed to take up to 12 minutes at the fourth grade and up to 15 minutes at the eighth grade, the length of one assessment block (see section 2.6.1 on booklet/block design). Exhibits 2.8 and 2.9 describe the problem-solving and inquiry tasks selected for the field test and the main content covered in each for the fourth grade and eighth grade, respectively.

Results from the international field test (section 2.5.5) were used to select the problem-solving and inquiry tasks that performed best internationally for the main survey. In addition, a cognitive laboratory investigation of the field-test version of the problem-solving and inquiry tasks was conducted by the United States National Center for Education Statistics. This involved working with a small group of students to probe their understanding of the demands of the tasks and to uncover any conceptual difficulties encountered in them. The results of the international field test as well as the experiences from the cognitive laboratory investigation were used to inform the selection process and to make revisions to improve the clarity of directions, layout, reading level, use of manipulatives, and scoring guides for the main survey. In general, the problem-solving and inquiry tasks selected for the main survey were shortened from the field test version. In some cases, an entire task or large portions of a task were selected. In other cases, individual items within tasks were selected and adapted to function as stand-alone items. As shown in Exhibit 2.10, a total of 13 problem-solving and inquiry tasks were selected for the main survey (6 at fourth grade and 7 at eighth grade).

Exhibit 2.8 **Problem-Solving and Inquiry Tasks Selected for the Field Test – Grade 4**

Name of Task	Description	Main Content
	Mathematics Tasks	
Geometry Tiles	Students are given three types of square tiles (black, white, and triangle tiles half black and half white) that can be placed together to form patterns. Students create two-dimensional shapes; compute fraction of pattern that is black; create patterns satisfying given conditions.	Geometry and Number
Number Tiles	Students are given number tiles marked from 0 to 9 that can be used to create addition, subtraction and multiplication problems. By choosing the place value of the numbers (units or tens), students combine their tiles to create problems that give a total closest to a given number, and to create the largest possible answer.	Number
Trading Cards	Three types of trading cards can be exchanged according to equivalency rules. Students compute how many cards they would get if they trade n cards of a certain type by another type of cards; explain how to maximize the number of cards they could get by trading; and infer conversion rules.	Number
Reversible Numbers	Presents examples of reversible numbers (e.g., 66, 121, 3003) and a general rule to make reversible numbers starting from two-digit numbers. Students provide examples of reversible numbers meeting certain conditions; create reversible numbers following one- or two-step rules; justify why reversible numbers cannot have three different digits; evaluate rules to create reversible numbers.	Number
Map It!	Students are shown maps drawn to scale indicating several locations. Using a cardboard ruler, students measure distance in centimetres between towns; estimate distance in kilometres; infer which towns are closer; compute time required to travel from one town to another; mark new plausible locations in the map so that they satisfy given conditions.	Measurement and Number
	Science Tasks	
Oceans and Tidepools	Presents textual and graphical information about the oceans and tidepools and a series of exploratory questions involving food chains, features of organisms, and ocean resources; students make predictions, provide explanations; select set-ups to investigate the effect of salt level on seaweed.	Life Science
Garden	Presents a practical situation involving a plan for a garden and a series of questions about plant growth and dispersal, light conditions, and importance/control of insects; students make predictions; provide explanations; interpret diagram; extract tabular information; relate position of sun and light conditions to complete table of plants in each area.	Life Science and Earth Science
Patterns on Earth	Presents historical information about measuring time using observed patterns (phases of the moon, daily cycle of the sun, appearance of shadows, periodic motion of pendulums); students evaluate graphical representations; complete diagrams; extend and relate patterns to time measurements; relate periodic motion of a pendulum to gravity.	Earth Science and Physical Science
Light and Color	Presents a practical situation involving an investigation of the effect of the light source on the color of materials; students describe and interpret results of the investigation; draw conclusions; make predictions and generalize results to new situations; compare with situations where color changes are due to changes in materials.	Physical Science

Exhibit 2.9 **Problem-Solving and Inquiry Tasks Selected for the Field Test – Grade 8**

Name of Task	Description	Main Content
	Mathematics Tasks	
Geometry Tiling	Provides four identical geometry tiles and several grids showing how tiles can be placed to form patterns. Students place tiles on a grid to make a pattern symmetrical about a given line; extend geometric patterns using symbols to represent the position of the tiles; and create whole new symmetrical patterns using symbols.	Geometry, Number, and Algebra
Class Trip	Students are given a map, bus timetables, trip rates per student, and a series of conditions that must be met in planning a class trip. Students estimate distances; compute costs for different trip options; evaluate if conditions can be met; decide upon which trip to make; and justify their choice.	Measurement, Number, and Data
Red and Black Tiles	Presents red and black tiles that can be combined to form square shapes with a given pattern but having different sizes. Students extend numeric and geometric patterns; identify number of tiles of each type required to form a shape of a given size; and infer the general algebraic expression to find out the number of tiles needed for any shape.	Algebra
Phone Plans	Presents two telephone payment plans involving fixed and variable costs. Students read and interpret data from a table to decide which plan would be cheapest under a range of conditions and justify their selection of a plan.	Data
Bird House	Students are given plans for making a wooden birdhouse. Working from the scale drawings in the plans, and using a ruler, students determine the actual size of the wood pieces required to build the birdhouse. They also infer the size of a missing piece, and draw it.	Measurement, Number, and Geometry
Number Triangles	Presents number triangles with some numbers missing, and an adding rule to combine the existing numbers to determine the missing numbers. Students determine how to create different combinations of odd and even numbers, and how to get positive and negative integers; they also identify ranges of values that satisfy given conditions.	Number
	Science Tasks	
Oceans	Presents textual and graphical information about the oceans and a series of exploratory questions involving food webs, adaptations of organisms, resources, and human exploration using sonar technology; students make predictions; provide explanations; interpret graphical information; describe procedures.	Life Science and Earth Science
Galapagos Islands	Presents textual and graphical information about the Galapagos Islands and a series of exploratory questions involving formation, arrival of organisms, impact of humans, adaptations and competition among species; students make predictions; interpret graphical data; draw conclusions; provide explanations.	Life Science and Earth Science
Metal Crown	Presents an investigation of a crown of unknown composition; students predict observable properties; describe a procedure to determine volume and density; evaluate results from repeated measures; draw conclusion by comparing measurements to density/cost data for various metals.	Physical Science
Light Filters	Presents practical situations involving color change due to the light source or to changes in materials; students interpret and explain results of an investigation of the effect of light sources/filters on color; apply knowledge of chemical change to a new situation (dye fade); design an investigation of the effect of light source on dye fade.	Physical Science

Exhibit 2.10 **Problem-Solving and Inquiry Tasks Selected for the Main Survey at Grade 4 and Grade 8**

Grade 4		Grade 8	
Name of Task	**Content Domains***	**Name of Task**	**Content Domains***
	Mathematics Tasks		
Geometry Tiles	Geometry (2) Number (4)	Geometry Tiling	Geometry (4) Algebra (1)
Number Tiles	Number (7)	Class Trip	Measurement (2) Number (2) Data (6)
Trading Cards	Number (6)	Red and Black Tiles	Algebra (8)
Marytown (portions of original Map It task)	Measurement (3)	Phone Plans	Data (6)
	Science Tasks		
Garden	Life Science (7) Earth Science (1)	Life in the Oceans (portions of original Oceans task)	Life Science (7)
Light and Color	Physical Science (7)	Galapagos Islands	Life Science (7)
		Metal Crown	Physics (4) Chemistry (3)

* The number of score points in each content domain is indicated in parentheses. The tasks range from three to ten score points.

2.5.5 Field Test

To evaluate the international performance of the new items developed for TIMSS 2003, a full-scale field test was conducted at both the fourth and eighth grades during the period April to June 2002. In total, 41 countries participated in the eighth-grade field test and 20 countries in the fourth grade. The field test in each country was administered to a random sample of a minimum of 25 schools, with two classrooms per school. To ensure that an adequate number of items were available for selection, substantially more items were field tested (1-1/2 to 2 times) than were needed in the assessment, particularly constructed-response items and items in content areas not already covered by trend items from 1995 and 1999.

Including the problem-solving and inquiry tasks, a total of 435 items were included in the fourth-grade field test, 229 in mathematics and 206 in science. At the eighth grade, a total of 386 items were included in the field test, 190 in mathematics and 196 in science. Since some constructed-response items contribute two score points, this corresponds to a total number of score points of 242 in mathematics and 248 in science at the fourth grade, and 211 in mathematics and 239 in science at the eighth grade.

2.5.6 Item Selection for the Main Survey

International item analysis of the results from the field test was used to inform the review and selection of items and tasks for the main survey. Data almanacs were produced containing basic item statistics for each country and internationally to evaluate the item difficulty, how well items discriminated between high-and low-performing students, the effectiveness of distracters in multiple-choice items, scoring reliability for constructed-response items, the frequency of occurrence of diagnostic codes used in the scoring guides, and whether there were any biases towards or against individual countries or in favor of boys or girls.

The TIMSS International Study Center conducted an initial review of the field-test results in early July 2002, using data from 36 countries at the eighth grade and 19 countries at the fourth grade that were available for analysis at that time. This review included NRC input from field test survey activities reports, feedback on items and scoring guides, and translation verification reports to identify any items with translation or cultural issues affecting international item performance. On the basis of this review, the mathematics and science coordinators identified the set of test items they felt would be most appropriate for use in the assessment, taking into account individual item statistics as well as alignment with the frameworks. Draft blocks of items for the assessment were then assembled for review by the Science and Mathematics Item Review Committee.

At its third meeting on July 15 - 18, 2002, the SMIRC reviewed the proposed item blocks, examining the field test item statistics to identify any anomalies. Items that did not work well were replaced with alternate items from the same content area. The problem-solving and inquiry tasks received particular attention and improvements were made where necessary. Revisions to items included improving graphics and item layout, clarifying stems, and revising distracters selected by very low percentages of students. In a few instances, item format was changed from multiple-choice to constructed-response or vice-versa. The final set of items selected was chosen to provide an appropriate balance in content coverage, level of difficulty, and item types.

Based on the recommendations of the SMIRC, the International Study Center prepared draft instruments for the assessment to be reviewed by the National Research Coordinators at their fifth meeting in August 2002. The draft instruments were well received and widely discussed by NRCs, who recommended a number of additional improvements that were incorporated into the final instruments distributed in September 2002. A total of 243 new items at the fourth grade and 230 items at the eighth grade were selected for the main survey. Including both trend and new items, the final tests include 313 items at the fourth grade and 383 items at the eighth grade. Exhibits 2.11 and 2.12 show the distribution of new and trend items in the main survey by subject and item format for fourth and eighth grades, respectively, and reflect the individual items and all item subparts included in multi-part items and problem-solving and inquiry tasks. Between 40 and 50 percent of the total score points are contributed by constructed-response items at both grades, which exceeds the minimum proportion of one-third specified in the frameworks.

Exhibit 2.11 **Distribution of New and Trend Items in the TIMSS 2003 Main Survey by Subject and Item Format – Grade 4**

	Number of Items				
Item Format	**New Items**	**Trend Items**	**Total (New + Trend)**	**Total Score Points**	**Percentage of Score Points**
		Mathematics Items			
Multiple Choice	55	37	92	92	54%
Constructed Response	69	0	69	77	46%
Total Mathematics Items	**124**	**37**	**161**	**169**	
		Science Items			
Multiple Choice	60	31	91	91	54%
Constructed Response	59	2	61	77	46%
Total Science Items	**119**	**33**	**152**	**168**	
		All Items			
Multiple Choice	115	68	183	183	54%
Constructed Response	128	2	130	154	46%
Total Items	**243**	**70**	**313**	**337**	

Exhibit 2.12 **Distribution of New and Trend Items in the TIMSS 2003 Main Survey by Subject and Item Format – Grade 8**

Item Format	Number of Items: New Items	Number of Items: Trend Items	Number of Items: Total (New + Trend)	Total Score Points	Percentage of Score Points
Mathematics Items					
Multiple Choice	69	59	128	128	60%
Constructed Response	46	20	66	87	40%
Total Mathematics Items	**115**	**79**	**194**	**215**	
Science Items					
Multiple Choice	56	53	109	109	52%
Constructed Response	59	21	80	102	48%
Total Science Items	**115**	**74**	**189**	**211**	
All Items					
Multiple Choice	125	112	237	237	56%
Constructed Response	105	41	146	189	44%
Total Items	**230**	**153**	**383**	**426**	

2.5.7 Scoring of Constructed-Response Items

In the TIMSS 2003 assessment, constructed-response items made up more than 40 percent of the total assessment time, and a large number of constructed-response items were developed and field tested. Scoring guide development for the constructed-response items was a considerable effort and an integral part of the test development process for TIMSS 2003. This section describes the TIMSS general scoring method, the scoring guide development process, and the scoring training materials and procedures.

2.5.7.1 The TIMSS General Scoring Method

TIMSS 2003 used the same approach to scoring as the previous TIMSS assessments. As in TIMSS 1995 and 1999, both short-answer items and extended-response items were included in the assessment. Short-answer items typically are worth one score point and require a numerical response in mathematics or a brief descriptive response in science. Extended-response items are worth a maximum of two score points and require students to show their work or provide explanations using words and/or diagrams to demonstrate their conceptual and procedural knowledge. The generalized scoring guides for mathematics and science items developed for TIMSS 1999 (Exhibit 2.13) also were applied in TIMSS 2003.

Exhibit 2.13 **TIMSS Generalized Scoring Guide for Mathematics and Science Items**

Mathematics	Science
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	**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 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	**1 Point** A one-point response is only partially correct. The response demonstrates only a partial understanding of the science 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 when required • 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 completely incorrect, irrelevant, or incoherent.	**0 Points** A zero-point response is seriously inaccurate or inadequate, irrelevant, or incoherent.
Short-Answer Items	
1 Point A one-point response is correct. The response indicates that the student has completed the task correctly.	**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.	**0 Points** A zero-point response is completely incorrect, irrelevant, or incoherent.

Each constructed-response item has its own scoring guide that utilizes a two-digit scoring scheme to provide diagnostic information. The first digit designates the correctness level of the response: 2 for a two-point response, 1 for a 1-point response, and 7 for an incorrect 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. A second digit of 0-5 may be used for pre-defined international codes at each correctness level, while a second digit of 9 corresponds to "other" types of responses that fall within the appropriate correctness level but do not fit any of the pre-defined international codes. A special code (99) is given for completely blank responses. In general, only a few diagnostic codes are used to track high-frequency correct or partial approaches or common misconceptions and errors, and a particular effort was made in TIMSS 2003 to minimize the number of diagnostic codes used. In addition to the international codes, second digit codes of 7 and 8 may be used by national centers to monitor specific responses not already captured by the internationally-defined codes. The general TIMSS two-digit scoring scheme is summarized in Exhibit 2.14.

Exhibit 2.14 **TIMSS Two-Digit Scoring Scheme for Constructed-Response Items**

Two- Point Items			One-Point Items		
Correctness Level	**International Code(s)**		**Correctness Level**	**International Code(s)**	
Correct Responses	20 – 25:	category/method #1 - #5	Correct Responses	10 – 15:	category/method #1- #5
	29:	other correct method		19:	other correct method
Partial Responses	10 – 15:	category/method #1- #5	Incorrect Responses	70 – 75:	misconception/error #1- #5
	19:	other partial method		79:	other error
Incorrect Responses	70 – 75:	misconception/error #1 - #5	Blank	99	
	79:	other error			
Blank	99				

2.5.7.2 Developing the TIMSS 2003 Scoring Guides

Items and scoring guides were developed in parallel, with draft scoring guides provided by item writers along with their item submissions. Scoring guides were further developed during item review and revision by the mathematics and science task forces and at the first two meetings of the Science and Mathematics Item Review Committee. Draft field-test versions of the scoring guides were reviewed by National Research Coordinators at their third NRC meeting. In February 2002, prior to the field test, a small-scale pilot of fourth- and eighth-grade constructed-response items was conducted in seven countries that tested in English. This pilot included all of the problem-solving and inquiry tasks as well as other items with more challenging scoring guides. Results from the pilot were used to finalize scoring guides for the field test by identifying common responses and clarifying the threshold for correct versus partial or incorrect responses. Selected student responses from the pilot were included as examples in the scoring guides and materials for scoring training for the field test.

In general, the scoring reliability from the field test was quite high, with an average percent agreement correctness of more than 90 percent for nearly all items. However, scoring relibility data did suggest some scoring guide revisions. NRC feedback on their scoring experiences during the field test also was used to make improvements in the scoring guides. In addition, sets of student booklets from the field test were collected from all of the English-test countries as sources of example student responses to clarify codes and prepare scoring training materials for the assessment.

During the review of the main survey test instruments at the fifth NRC meeting in August 2002, the changes recommended at the SMIRC meeting were discussed and NRCs made some additional suggestions for revisions to the scoring guides. Because so many changes were made to the problem-solving and inquiry tasks after the field test, these were included in a second small-scale item trial conducted in September 2002 in five countries that test in English. Student responses from this trial provided examples for the final scoring guides and for scoring training materials. The scoring guides and training materials were used during the first international scoring training session in November 2002 for southern hemisphere countries. A few additional revisions and clarifications were suggested by the national representatives at this training session. These were incorporated into the guides prior to their general distribution in December 2002.

Scoring guides for the trend constructed-response items (35 items from the eighth-grade 1999 assessment and 2 items from the fourth-grade 1995 assessment) were essentially unchanged from the versions used in the previous assessments, except for some modifications made to be consistent with the TIMSS 2003 format.[4]

2.5.7.3 Scoring Training Materials and Procedures

As in previous assessments, the International Study Center used a "train-the-trainers" approach to provide training on the international procedures for scoring the TIMSS 2003 constructed-response items. National Research Coordinators and/or other personnel responsible for training scorers in each country participated in training sessions for the field test and the main survey. In each of these sessions, the general TIMSS scoring approach was reviewed, and participants were then trained on a subset of constructed-response items. The subset of items was selected to reflect a range of scoring guide types and situations encountered across the TIMSS mathematics and science items and included some of the most complicated scoring guides.

Training was organized into four sessions by subject and grade (mathematics fourth and eighth grades and science fourth and eighth grades) conducted by the mathematics and science coordinators and task force members. Participants received the international version of the scoring guides and a binder for each subject/grade combination containing a set of prescored

4 Scoring guides for a few eighth-grade science items from 1999 were simplified to reduce the number of diagnostic codes. In all cases, the overall scoring strategy was retained to ensure score-level reliability from 1999 to 2003.

example student responses illustrating the diagnostic codes and the rationale used to score the responses and a set of 10-20 unscored practice responses for each item. The student responses were selected from the international small-scale item pilot and field-test booklets.

The purpose of the international scoring training was to present a model for use in each country and an opportunity to practice and resolve scoring issues with the most difficult items. The training teams discussed the need for NRCs to prepare comparable materials for training in their own country for all constructed-reponse items and a larger number of practice responses for the more challenging scoring guides during the national training sessions. The following general procedures were followed in the scoring training for each item:

- Participants read the item and its scoring guide.
- Trainers discussed the rationale and methodology of the scoring guide.
- Trainers presented and discussed the set of prescored example student responses.
- Participants scored the set of practice student responses.
- Trainers led a group discussion of the scores given to the practice responses to reach a common understanding of the interpretation and application of the scoring guide.

Scoring training for the field test was conducted at the fourth NRC meeting in March 2002. Two full days of scoring training were devoted to the science items, with one day for each grade. For mathematics, training for both grades was done over a total of one and one-half days.

Scoring training for the assessment was conducted in the same fashion as for the field test, with separate sessions devoted for each subject/grade combination. For the assessment scoring training, 40 total items were included for eighth grade – 20 mathematics items and 20 science items. This set of items represents nearly 30 percent of the constructed-response items in the eighth-grade assessment. For fourth grade, 14 items were included for mathematics, and 16 items were included for science. This represents more than 25 percent of the constructed-response items in the fourth-grade assessment. For each grade, at least one item from each of the problem-solving and inquiry tasks was selected for training.

Two main scoring training sessions were conducted for the 2003 assessment, one for countries on a southern hemisphere schedule and one for countries on a northern hemisphere schedule.[5] The first was held in November 2002 in Wellington, New Zealand, for southern-hemisphere countries. The second, held in March 2003 in conjunction with the sixth NRC meeting

5 An extra scoring training session was organized in May 2003 for northern hemisphere countries that were unable to attend the main training session.

in Bucharest, Romania, was for the remaining countries. At each session, a full day of training was devoted to each subject for eighth grade and a little less for fourth grade (about a half day for mathematics and three-quarters for science). After the completion of scoring training, code sheets for the example and practice papers were distributed to NRCs for use in organizing scoring training materials in their own countries.

2.6 Assessment Booklet Design

In order to cover the frameworks, the pool of items and tasks included in the TIMSS asssessment is extensive and would require much more testing time than could be alloted for individual students (about seven hours at grade 8 and five and one-half hours at grade 4). Therefore, as in the 1995 and 1999 assessments, TIMSS 2003 uses a matrix-sampling technique that involves dividing the entire assessment pool into a set of unique item blocks, distributing these blocks across a set of booklets, and rotating the booklets among the students. Each student takes one booklet containing both mathematics and science items.[6]

2.6.1 Block and Booklet Design

The TIMSS design for 2003 divides the 313 items at fourth grade and 383 items at eighth grade into 28 item blocks at each grade, 14 mathematics blocks labeled M01 through M14, and 14 science blocks labeled S01 through S14. Each block contains either mathematics items only or science items only. This general block design, shown in Exhibit 2.15, is the same for both grades, although for the assessment time is 12 minutes for fourth-grade blocks and 15 minutes for eighth-grade blocks. At the eighth grade, six blocks in each subject (blocks 01 – 06) contain secure items from 1995 and 1999 to measure trends and eight blocks (07 – 14) contain new items developed for TIMSS 2003. Since fourth grade was not included in the 1999 assessment, trend items from 1995 only were available, and these were placed in the first three blocks. The remaining 11 blocks contain items new in 2003.

In the TIMSS 2003 design, the 28 blocks of items are distributed across 12 student booklets, as shown in Exhibit 2.16. Each booklet consists of six blocks of items. To enable linking between booklets, each block appears in two, three, or four different booklets. The assessment time for individual students is 72 minutes at fourth grade and 90 minutes at eighth grade, which is comparable to that in the 1995 and 1999 assessments.

6 See Mullis et al. (2003) for more information on the assessment booklet design.

Exhibit 2.15 **General Design of the TIMSS 2003 Matrix-Sampling Blocks**

Source of Items	Mathematics Blocks	Science Blocks
Trend Items (TIMSS 1995 or 1999)	M01	S01
Trend Items (TIMSS 1995 or 1999)	M02	S02
Trend Items (TIMSS 1995 or 1999)	M03	S03
Trend Items (TIMSS 1999)	M04	S04
Trend Items (TIMSS 1999)	M05	S05
Trend Items (TIMSS 1999)	M06	S06
New Replacement Items (TIMSS 2003)	M07	S07
New Replacement Items (TIMSS 2003)	M08	S08
New Replacement Items (TIMSS 2003)	M09	S09
New Replacement Items (TIMSS 2003)	M10	S10
New Replacement Items (TIMSS 2003)	M11	S11
New Replacement Items (TIMSS 2003)	M12	S12
New Replacement Items (TIMSS 2003)	M13	S13
New Replacement Items (TIMSS 2003)	M14	S14

The booklets are organized into two three-block sessions (Parts I and II), with a break in between each part. Since the use of calculators was introduced for the first time in TIMSS 2003 at the eighth grade, this had an impact on the booklet design. To ensure that calculators could be used for the new items but not for the trend items from 1995 and 1999, the trend items (blocks 01 – 06) were placed in Part I of the test booklets to be completed without calculators before the break. After the break, calculators were allowed for the new items (blocks 07 – 12) at eighth grade but not fourth grade. To provide a more balanced design, however, two mathematics trend blocks (M05 and M06) and two science trend blocks (S05 and S06) also were placed in Part II of one booklet each.

Exhibit 2.16 **Booklet Design for TIMSS 2003 – Grade 4 and Grade 8**

	Assessment Blocks					
Student Booklet	**Part I**			**Part II**		
Booklet 1	M01	M02	S06	S07	M05	M07
Booklet 2	M02	M03	S05	S08	M06	M08
Booklet 3	M03	M04	S04	S09	M13	M11
Booklet 4	M04	M05	S03	S10	M14	M12
Booklet 5	M05	M06	S02	S11	M09	M13
Booklet 6	M06	M01	S01	S12	M10	M14
Booklet 7	S01	S02	M06	M07	S05	S07
Booklet 8	S02	S03	M05	M08	S06	S08
Booklet 9	S03	S04	M04	M09	S13	S11
Booklet 10	S04	S05	M03	M10	S14	S12
Booklet 11	S05	S06	M02	M11	S09	S13
Booklet 12	S06	S01	M01	M12	S10	S14

2.6.2 Assembling Item Blocks

The assessment blocks were assembled to create a balance across blocks and booklets with respect to content domain, cognitive domain, and item format. Although a balance was achieved at the overall assessment level, the distribution of item types varies across blocks. The trend blocks from 1995 (blocks 01 – 03) contain mostly multiple-choice items, while the blocks containing the problem-solving and inquiry tasks have a higher proportion of constructed-response items. Each block contains an average of 12 score points at fourth grade and 15 score points at eighth grade, and the percentage of score points from constructed-response items in each block ranges from 0 to about 80 percent. On average, there are 6-7 multiple-choice items, 4-5 short-answer items, and 0-1 extended-response items per block at the fourth grade. At the eighth grade, there are 8-9 multiple-choice items, 3-4 short-answer items, and 1-2 extended-response items per block, on average. Depending on the exact number of multiple-choice, short-answer, and extended-response items in each block, the total number of items in a block ranges from 10 to 13 at fourth grade and from 11 to 16 at eighth grade.

2.6.3 Incorporating Trend Items

In TIMSS 1995 and 1999, items were organized into 26 item clusters (labeled A through Z). Clusters A-R contained sets of both mathematics and science items, clusters S-V only mathematics items, and clusters W-Z only science items. After the 1995 assessment, clusters A-H (containing nearly all multiple-choice items) were held secure for future assessments; clusters I-Z were released and replaced with new items in the 1999 assessment. Since the fourth grade was not included in the 1999 assessment, only clusters A-H from 1995 are available as trend items for the 2003 assessment, and these clusters contain nearly all multiple-choice items.

At the eighth grade, clusters I-Z contained items developed for the 1999 assessment. At the end of TIMSS 1999, the "even" clusters (B, D, F, etc.) were released and the "odd" clusters (A, C, E, etc.) were held secure as trend items for the 2003 assessment. Therefore, the following clusters of trend items at the eighth grade are available for the 2003 assessment:

- 1995 items: A, C, E, G (mathematics and science)
- 1999 items: I, K, M, O, Q (mathematics and scence); S, U (mathematics); W, Y (science)

Because of the new booklet and block design specified in the TIMSS 2003 frameworks, the trend item clusters from 1995 and 1999 were reorganized for the TIMSS 2003 assessment. In accordance with the TIMSS 2003 test design, mathematics and science items from 1995 were assigned to blocks M01-M03 or S01-S03. Most items from 1999 were assigned to blocks M04-M06 or S04-S06, although some were assigned to blocks M01-M03 or S01-S03 where there were insufficient 1995 items to fill these blocks. In addition, some new items were added to fill the trend blocks; in particular, blocks M04-M06 and S04-S06 contain all new items at the fourth grade. The assignment of 1995 and 1999 trend item clusters to the TIMSS 2003 item blocks and the resulting distribution of score points across assessment years is summarized for the fourth and eighth grades in Exhibits 2.17 and 2.18, respectively.

2.6.4 Alignment with the Mathematics and Science Frameworks

The test development process for TIMSS 2003 successfully produced fourth- and eighth-grade assessments aligned with the *TIMSS Assessment Frameworks and Specifications 2003*. Details about the coverage of the frameworks are given separately for the fourth- and eighth-grades in the following sections.

Exhibit 2.17 **TIMSS 2003 Mathematics and Science Blocks – Grade 4: Number of Items from 1995 Trend Clusters and Score Points by Assessment Year**

Block	Number of Items from Trend Clusters*	Score Points by Assessment Year		
		1995	2003	Total
Mathematics Blocks				
M01	C(4), E(4), G(4)	12	0	12
M02	A(3), D(5), F(5)	13	0	13
M03	A(2), B(5), H(5)	12	0	12
M04 – M14	–	0	132	132
Mathematics Total		**37**	**132**	**169**
Science Blocks				
S01	A(4), B(4), F(3)	11	0	11
S02	D(2), G(5), H(4)	11	0	11
S03	C(5), D(2), E(4)	11	0	11
S04 – S14	–	0	135	135
Science Total		**33**	**135**	**168**
Overall Total Score Points		**70**	**267**	**337**

* The number of items from each trend cluster is indicated in parentheses. Items in clusters A-H were developed for the 1995 assessment; grade 4 was not included in the 1999 assessment. Blocks M04-M14 and S04-S14 contain only new items developed for TIMSS 2003.

2.6.4.1 Fourth-Grade Assessment

Exhibit 2.19 shows the distribution of score points across content and cognitive domains in the fourth-grade mathematics assessment. The percentage of score points across both content and cognitive categories is very close to the target percentages specified in the frameworks (Exhibit 2.2). Exhibit 2.20 shows the score-point distribution for the fourth-grade science assessment, as well as the score points in the scientific inquiry assessment strand (see Exhibit 2.4 for the science framework target percentages). For both mathematics and science, items reflecting the full range of cognitive domains are included in each content domain. About 10 percent of the score points in science, covering a wide range of science content, also contribute to the scientific inquiry strand.

Exhibit 2.18 **TIMSS 2003 Mathematics and Science Blocks – Grade 8: Number of Items from 1995/1999 Trend Clusters and Score Points by Assessment Year**

Block	Number of Items from Trend Clusters*	Score Points by Assessment Year			
		1995	1999	2003	Total
		Mathematics Blocks			
M01	A(6), G(6)	12	0	3	15
M02	C(5), Q(10)	5	10	0	15
M03	E(6), O(9)	6	9	0	15
M04	I(9), S(7)	0	17	0	17
M05	K(9), U(4)	0	16	0	16
M06	M (8)	0	8	7	15
M07 – M14	–	0	0	122	122
Mathematics Total		**23**	**60**	**132**	**215**
		Science Blocks			
S01	E(6), K(10)	6	10	0	16
S02	A(6), C(6)	12	0	3	15
S03	G(6), O(8)	6	8	0	14
S04	M(6), W(4)	0	11	4	15
S05	I(11), Y(3)	0	15	0	15
S06	Q (8)	0	8	7	15
S07 – S14	–	0	0	121	121
Science Total		**24**	**52**	**135**	**211**
Overall Total Score Points		**47**	**112**	**267**	**426**

* The number of items from each trend cluster is indicated in parentheses. Items in clusters A-H were developed for the 1995 assessment; items in clusters I-Z were developed for the 1999 assessment. Blocks M07-M14 and S07-S14 contain only new items developed for TIMSS 2003.

Exhibit 2.19 **Distribution of Score Points in the TIMSS 2003 Mathematics Assessment by Content and Cognitive Domains – Grade 4**

Content Domain	Cognitive Domain				Total Score Points	Percentage of Score Points
	Knowing Facts and Procedures	Using Concepts	Solving Routine Problems	Reasoning		
Number	15	17	27	9	68	40%
Patterns and Relationships	3	5	9	8	25	15%
Measurement	9	3	12	9	33	20%
Geometry	12	8	4	1	25	15%
Data	0	6	9	3	18	11%
Total Score Points	**39**	**39**	**61**	**30**	**169**	
Percentage of Score Points	**23%**	**23%**	**36%**	**18%**		

Exhibit 2.20 **Distribution of Score Points in the TIMSS 2003 Science Assessment by Content and Cognitive Domains, and Scientific Inquiry Strand – Grade 4**

Content Domain	Cognitive Domain: Factual Knowledge	Cognitive Domain: Conceptual Understanding	Cognitive Domain: Reasoning and Analysis	Total Score Points	Percentage of Score Points	Scientific Inquiry Score Points
Life Science	28	28	16	72	43%	4
Physical Science	16	26	17	59	35%	12
Earth Science	15	16	6	37	22%	1
Total Score Points	**59**	**70**	**39**	**168**		**17**
Percentage of Score Points	**35%**	**42%**	**23%**			

In accordance with the frameworks, a range of item types is reflected in the TIMSS 2003 assessment, including multiple-choice, short-answer, and extended-response items. Exhibit 2.21 shows the breakdown of the fourth-grade mathematics and science items by item type and cognitive domain, indicating that each content domain covers a range of item types.

Exhibit 2.21 **Number of Mathematics and Science Items in TIMSS 2003 by Item Type and Content Domain – Grade 4**

Content Domain	Item Type: Multiple Choice	Item Type: Short Answer	Item Type: Extended Response	Total Number of Items
Mathematics Items				
Number	30	31	2	63
Patterns and Relationships	16	7	1	24
Measurement	23	10	0	33
Geometry	12	11	1	24
Data	11	5	1	17
Total Mathematics Items	**92**	**64**	**5**	**161**
Science Items				
Life Science	41	23	1	65
Physical Science	29	20	4	53
Earth Science	21	13	0	34
Total Science Items	**91**	**56**	**5**	**152**
Total Overall Items	**183**	**120**	**10**	**313**

TIMSS reports trends in student achievement in mathematics in the major content domains of each subject. To facilitate linking to previous assessments, TIMSS 2003 includes items from 1995 in the fourth grade and from 1995 and 1999 in the eighth grade in each content domain. Exhibit 2.22 shows, for the fourth-grade assessment, the number of score points in mathematics and science contributed by items used previously in 1995 and by those used for the first time in 2003. In mathematics, the number of score points in the five content domains ranges from a maximum of 19 (Number) to a minimum of 2 (Patterns and Relationships). In science, there are between 9 and 12 score points from the 1995 assessment in the content domains. Because there are relatively few items and score points from the 1995 assessment in most content domains, TIMSS 2003 developed achievement scales linking 1995 and 2003 for mathematics and science overall, but not for individual content domains. However, the TIMSS 2003 design makes provision for sufficient trend items to develop achievement scales linking the content domains from 2003 onwards, i.e., to 2007, 2011, and so on.

Exhibit 2.22 **Number of Score Points in TIMSS 2003 from Each Assessment Year by Mathematics and Science Content Domain – Grade 4**

	Assessment Year			
Content Domain	**From 1995**	**From 1999**	**New in 2003**	**Total 2003**
		Mathematics		
Number	19	N/A	49	68
Patterns and Relationships	2	N/A	23	25
Measurement	8	N/A	25	33
Geometry	4	N/A	21	25
Data	4	N/A	14	18
Total in Mathematics	**37**	**N/A**	**132**	**169**
		Science		
Life Science	12	N/A	60	72
Physical Science	9	N/A	50	59
Earth Science	12	N/A	25	37
Total in Science	**33**	**N/A**	**135**	**168**
Total Overall	**70**	**N/A**	**267**	**337**

N/A: Not Applicable – TIMSS was not administered at fourth grade in 1999.

The block and booklet design for TIMSS 2003 ensures that the student booklets contain an appropriate balance of mathematics and science content. Exhibit 2.23 shows the number of mathematics and science score points available in each fourth-grade booklet. The number of score points per booklets ranges from 71 to 80, with an average of 75. In accordance with the frameworks, in booklets 1-6 about two-thirds of the score points come from mathematics items and one-third from science. Conversely, in booklets 7-12 about two-thirds of the score points come from science items and one-third from mathematics. All student booklets contain items from each of the mathematics and science content domains.

Exhibit 2.23 **Maximum Number of Score Points in TIMSS 2003 in Each Booklet by Mathematics and Science Content Domain – Grade 4**

	Booklet											
Content Domain	1	2	3	4	5	6	7	8	9	10	11	12
Mathematics												
Number	19	18	23	17	21	20	5	7	10	11	11	8
Patterns and Relationships	6	6	6	8	6	5	7	3	4	2	3	6
Measurement	12	13	7	9	11	12	6	8	2	3	5	6
Geometry	4	7	7	9	4	7	2	3	5	6	4	3
Data	8	5	6	5	4	3	4	3	2	2	4	2
Total in Mathematics	**49**	**49**	**49**	**48**	**46**	**47**	**24**	**24**	**23**	**24**	**27**	**25**
Science												
Life Science	13	10	13	13	9	11	18	22	19	21	29	24
Physical Science	6	9	10	8	7	10	12	14	20	25	14	17
Earth Science	8	10	4	4	9	7	18	16	11	7	10	9
Total in Science	**27**	**29**	**27**	**25**	**25**	**28**	**48**	**52**	**50**	**53**	**53**	**50**
Total Overall	**76**	**78**	**76**	**73**	**71**	**75**	**72**	**76**	**73**	**77**	**80**	**75**

2.6.4.2 Eighth-Grade Assessment

Exhibit 2.24 shows the distribution of score points across content and cognitive domains in the TIMSS 2003 eighth-grade mathematics assessment. The percentage of score points is close to the target percentages (Exhibit 2.2) for nearly all content and cognitive categories, although the assessment has a somewhat higher percentage of items in *knowing facts and procedures* and a lower percentage in *solving routine problems*. Exhibit 2.25 shows the distribution of score points across content and cognitive domains in the eighth-grade science assessment, as well as the number of score points in each content domain that also pertain to the scientific inquiry assessment strand. The percentages of score points in the content and cognitive domains of the science assessment also are close to their targets (see Exhibit 2.4). As with the fourth-grade assessment, items reflecting a range of cognitive domains are included in each of the mathematics and science content domains at the eighth grade. About 14 percent of the score points in science, covering a wide range of science content, also contribute to the scientific inquiry strand.

Exhibit 2.24 **Distribution of Score Points in the TIMSS 2003 Mathematics Assessment by Content and Cognitive Domains – Grade 8**

	Cognitive Domain					
Content Domain	**Knowing Facts and Procedures**	**Using Concepts**	**Solving Routine Problems**	**Reasoning**	**Total Score Points**	**Percentage of Score Points**
Number	15	11	27	7	60	28%
Algebra	13	12	10	18	53	25%
Measurement	9	2	15	8	34	16%
Geometry	7	8	10	9	34	16%
Data	1	6	14	13	34	16%
Total Score Points	**45**	**39**	**76**	**55**	**215**	
Percentage of Score Points	**21%**	**18%**	**35%**	**26%**		

Exhibit 2.25 **Distribution of Score Points in the TIMSS 2003 Science Assessment by Content and Cognitive Domains and Scientific Inquiry Strand – Grade 8**

	Cognitive Domain					
Content Domain	**Factual Knowledge**	**Conceptual Understanding**	**Reasoning and Analysis**	**Total Score Points**	**Percentage of Score Points**	**Scientific Inquiry Score Points**
Life Science	24	24	17	65	31%	8
Chemistry	7	16	11	34	16%	6
Physics	7	23	19	49	23%	9
Earth Science	12	13	8	33	16%	1
Environmental Science	9	4	17	30	14%	6
Total Score Points	**59**	**80**	**72**	**211**		**30**
Percentage of Score Points	**28%**	**38%**	**34%**			

Exhibit 2.26 shows the number of multiple-choice, short-answer, and extended-response items in each content domain for the eighth-grade assessment. As in the fourth grade, each of the content domains at eighth grade includes a range of item types.

Exhibit 2.26 **Number of Mathematics and Science Items in TIMSS 2003 by Item Type and Content Domain – Grade 8**

Content Domain	Item Type: Multiple Choice	Item Type: Short Answer	Item Type: Extended Response	Total Number of Items
Mathematics Items				
Number	43	11	3	57
Algebra	29	13	5	47
Measurement	19	9	3	31
Geometry	22	6	3	31
Data	15	8	5	28
Total Mathematics	**128**	**47**	**19**	**194**
Science Items				
Life Science	29	17	8	54
Chemistry	20	10	1	31
Physics	28	15	3	46
Earth Science	22	9	0	31
Environmental Science	10	8	9	27
Total Science	**109**	**59**	**21**	**189**
Total Items	**237**	**106**	**40**	**383**

To study trends in eighth-grade student mathematics and science achievement, TIMSS 2003 included items from the TIMSS 1995, 1999, and 2003 assessments. Exhibit 2.27 shows the number of score points in mathematics and science contributed by items used previously in 1995 and in 1999 as well as by those used for the first time in 2003. Among items from 1995, the number of score points in each content domain ranges from 3 to 6, and among 1999 items, from 6 to 20. TIMSS 2003 developed achievement scales linking 1995, 1999, and 2003 for mathematics and science overall, but because there are relatively few items and score points from the 1995 and 1999 assessments in content domains, TIMSS did not develop scales for measuring trends in individual content domains. However, the TIMSS 2003 design makes provision for sufficient trend items to develop achievement scales linking the content domains from 2003 onwards, i.e., to 2007, 2011, and so on. TIMSS used average percents correct to show changes in performance in the content domains from 1999 to 2003.

Exhibit 2.27 **Number of Score Points in TIMSS 2003 from Each Assessment Year by Mathematics and Science Content Domain – Grade 8**

Content Domain	Assessment Year			Total 2003
	From 1995	From 1999	New in 2003	
Mathematics				
Number	6	20	34	60
Algebra	6	11	36	53
Measurement	4	14	16	34
Geometry	4	8	22	34
Data	3	7	24	34
Total in Mathematics	**23**	**60**	**132**	**215**
Science				
Life Science	6	12	47	65
Chemistry	4	11	19	34
Physics	5	17	27	49
Earth Science	6	6	21	33
Environmental Science	3	6	21	30
Total in Science	**24**	**52**	**135**	**211**
Total Overall	**47**	**112**	**267**	**426**

Exhibit 2.28 shows the maximum number of score points in mathematics, science, and overall and the distribution of score points across the mathematics and science content domains for each booklet in the eighth-grade assessment. The total score points in each booklet ranges from 90 to 97, with an average of 94. As in the fourth grade, about two-thirds of score points are from mathematics items in booklets 1-6, and about two-thirds of score points are from science items in booklets 7-12. Each booklet covers the full range of mathematics and science content domains.

Exhibit 2.28 **Maximum Number of Score Points in TIMSS 2003 in Each Booklet by Mathematics and Science Content Domain – Grade 8**

Content Domain	Booklet 1	2	3	4	5	6	7	8	9	10	11	12
Mathematics												
Number	19	22	19	14	25	17	11	10	8	7	8	7
Algebra	12	12	15	22	7	15	4	6	9	6	7	9
Measurement	13	7	11	13	11	5	4	8	8	4	5	4
Geometry	8	11	11	7	14	10	5	4	5	4	5	5
Data	11	8	6	8	4	12	8	3	2	8	5	6
Total Mathematics	**63**	**60**	**62**	**64**	**61**	**59**	**32**	**31**	**32**	**29**	**30**	**31**
Science												
Life Science	10	12	7	7	9	10	20	19	18	23	21	23
Chemistry	6	4	5	6	4	5	10	10	11	8	10	10
Physics	6	8	9	6	7	9	17	11	13	19	14	13
Earth Science	4	5	3	5	8	5	9	12	11	7	7	10
Environmental Science	6	4	9	4	4	3	7	10	11	11	8	8
Total Science	**32**	**33**	**33**	**28**	**32**	**32**	**63**	**62**	**64**	**68**	**60**	**64**
Total Overall	**95**	**93**	**95**	**92**	**93**	**91**	**95**	**93**	**96**	**97**	**90**	**95**

2.6.5 Item Release Policy

TIMSS 2003 is the third assessment in a series of regular four-year studies, providing trend data from 1995 and 1999. As in previous assessments, the design for TIMSS 2003 and beyond (2007, 2011, etc.) provides for retaining some of the items for the measurement of trend and releasing some items into the public domain. In TIMSS 2003, half of the 14 assessment blocks in each subject will be released after the assessment results for 2003 are published. The released blocks will include all three mathematics and three science blocks containing trend items from 1995 (blocks M01 – M03, S01 – S03), one mathematics and one science block of trend items from 1999 (blocks M04 and S04)[7] and three blocks of new mathematics and science items and tasks developed for 2003 (blocks M09, M10, and M13; S09, S10, and S13). As item blocks are released, new items will be developed to take their place, and the release policy for future assessments will ensure that item blocks are cycled out after three assessments.

In the assignment of items to blocks in TIMSS 2003, particular attention was paid to balancing the blocks with respect to content domain to

7 At fourth grade, these blocks contain new 2003 items.

ensure that adequate numbers of items are held secure in each area for the purposes of measuring trend in future studies. In addition, the placement of the problem-solving and inquiry tasks results in about half of the tasks being retained and half being released after 2003. The released item set provides valuable information for interpreting the international and national reports and for use in secondary analyses. Therefore, it is also important that the released set be representative of the overall test to provide as much information as possible about the nature and scope of the test. Exhibits 2.29 and 2.30 show the number of secure and released items from the TIMSS 2003 assessment for fourth and eighth grades broken down by content domain. Approximately half of the items overall and in each content domain are released and half are kept secure. At the fourth grade, however, more than half of the items in the Number content domain and less than half of the items in Patterns and Relationships are released.

Exhibit 2.29 **Number of Items in Each Mathematics and Science Content Domain by Release Status in TIMSS 2003 – Grade 4**

Content Domain	Secure	Released	Total
Mathematics			
Number	24	39	63
Patterns and Relationships	17	7	24
Measurement	19	14	33
Geometry	12	12	24
Data	9	8	17
Total Mathematics	**81**	**80**	**161**
Science			
Life Science	32	33	65
Physical Science	29	24	53
Earth Science	15	19	34
Total Science	**76**	**76**	**152**
Total Overall	**157**	**156**	**313**

Exhibit 2.30 **Number of Items in Each Mathematics and Science Content Domain by Release Status in TIMSS 2003 – Grade 8**

Content Domain	Secure	Released	Total
Mathematics			
Number	26	31	57
Algebra	23	24	47
Measurement	14	17	31
Geometry	15	16	31
Data	17	11	28
Total Mathematics	**95**	**99**	**194**
Science			
Life Science	27	27	54
Chemistry	14	17	31
Physics	23	23	46
Earth Science	15	16	31
Environmental Science	15	12	27
Total Science	**94**	**95**	**189**
Total Overall	**189**	**194**	**383**

References

Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzalez, E.J., Chrostowski, S.J., & O'Connor, K.M. (2001), *TIMSS Assessment Frameworks and Specifications 2003*, Chestnut Hill, MA: TIMSS International Study Center, Boston College.

Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzalez, E.J., Chrostowski, S.J., & O'Connor, K.M. (2003), *TIMSS Assessment Frameworks and Specifications 2003 (2nd ed.)*, Chestnut Hill, MA: TIMSS International Study Center, Boston College.

Robitaille, D.F., McKnight, C.C., Schmidt, W.H., Britton, E., Raizen, S., & Nicol, C. (1993), *Curriculum Frameworks for Mathematics and Science: TIMSS Monograph No. 1*, Vancouver, BC: Pacific Educational Press.

3

Chapter 3

Developing the TIMSS 2003 Background Questionnaires

Steven J. Chrostowski

3.1 Overview

For a fuller appreciation of what the TIMSS achievement results mean and how they may be used to improve student learning in mathematics and science, it is important to understand the contexts in which students learn. Therefore, TIMSS collects extensive information about the contexts for learning mathematics and science by administering a range of background questionnaires. Four types of background questionnaires were used in TIMSS 2003 to gather information at various levels of the educational system: (i) curriculum questionnaires addressed issues of system-wide curriculum design and support and curricular emphasis in mathematics and science; (ii) a school questionnaire asked school principals/headmasters of the students tested to provide information about curricular and instructional arrangements, school resources, and school climate; (iii) teacher questionnaires asked mathematics and science teachers of the students tested about their preparation to teach, their teaching activities and approaches, their attitudes toward teaching the subject matter, and the curriculum that is implemented in the classroom; and (iv) a questionnaire for the students tested sought information about their home backgrounds, their attitudes toward learning mathematics and science, and their experiences in learning these subjects.

The questionnaires were based on the contextual framework included in the *TIMSS Assessment Frameworks and Specifications 2003* (Mullis, Martin, Smith, Garden, Gregory, Gonzalez, Chrostowski, & O'Connor, 2003). The contextual framework specifies the major characteristics of the educational and social contexts to be studied and identifies the areas to be addressed in the background questionnaires. Questionnaires were developed at both the fourth and eighth grades.

Because TIMSS is a trend study designed to measure change in student achievement in mathematics and science over time, it was important to retain many of the questions included in the background questionnaires in prior cycles of TIMSS for use in TIMSS 2003. Here the focus was on retaining those questions that were found to be most valuable in analysis and reporting in prior cycles of TIMSS. However, at the same time, it was also important to refine some questions and add new ones to address emerging research areas of interest. In particular, TIMSS 2003 added new questions on teacher preparation and professional development, and on the use of information technology for teaching and learning. In order to allow for such expansion in the questionnaires while also keeping response burden manageable, it was necessary to delete questions from earlier cycles of the study, and the focus here was on questions that were not included in reporting TIMSS results. In general, great effort was made to streamline the questionnaires in order to keep response burden to a minimum.

The conceptual framework underlying TIMSS uses the curriculum, broadly defined, as the major organizing concept to explain international variation in student achievement. The TIMSS curriculum model has three aspects: the intended curriculum, the implemented curriculum, and the attained curriculum. These represent, respectively, the mathematics and science that society intends for students to learn and how the education system should be organized to facilitate this learning; what is actually taught in classrooms, who teaches it, and how it is taught; and finally, what students have learned, and what they think about these subjects. Based on this model, TIMSS collects, through the background questionnaires, information about the factors likely to influence students' learning of mathematics and science at the national (or regional), school, classroom, and student level.

This chapter describes the contextual framework underlying the questionnaires, the process used to develop the questionnaires, and their content.

3.2 Contextual Framework for the Background Questionnaires

Just as the mathematics and science frameworks describe the content and cognitive domains to be assessed in those subjects, the contextual framework identifies the major characteristics of the educational and social contexts to be examined with a view toward improving student learning in mathematics and science.

3.2.1 Development of the Contextual Framework

In conjunction with the updating of the original TIMSS assessment frameworks in mathematics and science (see Chapter 2), a new contextual framework was developed by the TIMSS & PIRLS International Study Center (ISC) in collaboration with the TIMSS 2003 Expert Panel.[1] The contextual framework, like the mathematics and science assessment frameworks, went through an extensive and widely consultative development process spanning approximately one year. This work was supported by a grant from the U.S. National Science Foundation, in response to the proposal "A New TIMSS for a New Century." The three overarching goals of this proposal were to update the TIMSS frameworks to ensure that the latest developments in mathematics and science would be addressed by the TIMSS 2003 assessment, develop detailed specifications of the mathematics and science that should be covered in the TIMSS 2003 assessments, and articulate key policy issues that should be addressed in the TIMSS 2003 background questionnaires, i.e., teacher preparation and professional development, and the use of information technology in the classroom.

The development work on the frameworks began in September 2000 when the ISC distributed a survey to the National Research Coordinators (NRCs) seeking their suggestions for areas where the mathematics and science frameworks needed strengthening and revision and potential areas for inclusion in the contextual framework. In regard to the contextual framework and background questionnaires, some of the issues NRCs identified for exploration were:

- the relationship between student achievement and well-defined national curriculum and examinations;
- teacher preparation and professional development;
- student mobility and transience;
- school climate;
- simplifying the language used in the fourth-grade questionnaires;
- pruning the questionnaires by deleting items that have proven to be unreliable or not useful in analysis and reporting; and
- improving the layout of the questionnaires and organizing questionnaire items into logical blocks.

Development work on the contextual framework continued with the first meeting of the Expert Panel in November 2000 in Boston. The primary tasks of the Expert Panel regarding the contextual framework were to identify the main policy issues and new research questions to address in the background questionnaires, and to discuss data sources and methods of data

1 See Appendix A for a list of members of the Expert Panel.

collection. The first Expert Panel meeting included a discussion of the policy issues addressed in TIMSS 1999, an overview of the TIMSS 1999 background questionnaires, an articulation of the key policy issues to be addressed in TIMSS 2003, and a discussion of potential data sources and methods to collect contextual information for TIMSS 2003. Panel members agreed that there was a need to focus on a limited number of policy issues. The panel recognized the need to ensure that the questionnaires used in TIMSS 2003 maintain continuity with previous TIMSS surveys in order to measure trend, yet at the same time recognized the tension between the dual needs of addressing new policy areas while also streamlining the questionnaires in order to minimize response burden.

Following the first meeting of the Expert Panel, staff at the International Study Center prepared a model of the contextual framework for discussion at the First TIMSS 2003 National Research Coordinators' Meeting, held in February 2001 in Hamburg, Germany. NRCs emphasized that in developing the TIMSS 2003 questionnaires, the questions used in past TIMSS reports should be retained, and questions not used should be deleted. Also, the total time devoted to each questionnaire should not exceed that in TIMSS 1999. NRCs were asked to submit suggestions for the contextual framework, including areas of study and specific questions to include in the background questionnaires.

From March through April 2001, following the first NRC meeting, ISC staff further developed the assessment frameworks based on the input from NRCs. The revised frameworks were reviewed by the Expert Panel at its second meeting, held in May 2001 in Amsterdam, the Netherlands. The Expert Panel suggested the following topics for further exploration:

- **Teacher training:** The link between teacher training and later teaching effectiveness could be investigated. This could include the type of teacher training institution attended by teachers, the curriculum offered, the length of training and the amount of teaching practice, the use of technology in teacher training, and teacher competency standards.
- **Professional development:** Topics suitable for exploration include who provides the professional development, the nature of the professional development, the incentives for engaging in professional development, and the attractiveness of teaching as a profession.
- **Technology:** A central question to investigate would be level of access to the Internet by students and teachers, and how the Internet is used to facilitate teaching and learning. Additional topics that could be addressed include the ability of students to judge the quality of information they obtain via the Internet, and potential problems associated with Internet use.

Based on the input from the Expert Panel, ISC staff further revised the assessment frameworks for final review and approval by NRCs at the Second TIMSS 2003 National Research Coordinators' Meeting, held in June 2001 in Montreal, Canada. National Research Coordinators provided additional input on the frameworks, and upon incorporating some new suggestions from NRCs, the International Study Center published the first edition of the *TIMSS Assessment Frameworks and Specifications 2003* in September 2001.[2] In addition to the mathematics, science, and contextual frameworks, this document also includes a chapter on the planned assessment design.

3.2.2 Content of the Contextual Framework

The TIMSS contextual framework describes the contextual areas to be studied, and provides direction for development of the curriculum, school, teacher and student background questionnaires. The contextual framework encompasses five broad areas that interact with each other to impact student achievement:

- the curriculum;
- the schools;
- teachers and their preparation;
- classroom activities and characteristics;
- the students.

In particular, the framework focuses on the curricular goals of the education system and how the system is organized to attain and sustain those goals; the educational resources provided and how the school is organized to foster teaching and learning; the teaching force and how it is educated and supported; the topics that are taught and the learning activities that go on in the classroom; and the students' home background and learning support and the attitudes they bring to school.

The following sections briefly summarize the main areas included in the contextual framework.

3.2.2.1 The Curriculum

The TIMSS contextual framework sees curriculum development as a process involving consideration of the society which the education system serves, the needs and aspirations of the students, the nature and function of learning, and the formulation of statements on what learning is important. Building on past IEA experience, the TIMSS contextual framework addresses five broad aspects of the intended curriculum in mathematics and science: formulating

2 The second edition of the frameworks was published in February 2003, and features example mathematics and science achievement items used in the field test but not the main data collection, as well as a revised assessment design chapter.

the curriculum; defining the scope and content of the curriculum; organizing the curriculum, monitoring and evaluating the implemented curriculum; and providing curricular materials and support.

3.2.2.2 The Schools

In the TIMSS contextual model, the school is the institution through which the goals of the curriculum are implemented. TIMSS focuses on a set of indicators of school quality that research has shown to characterize schools that function as well-managed integrated systems supportive of teaching and learning. These include: organization of the school; school goals; roles of the school principal; resources to support mathematics and science learning; parental involvement; and a disciplined school environment.

3.2.2.3 Teachers and Their Preparation

Teachers are the primary agents of curriculum implementation in the TIMSS contextual model. Regardless of how closely prescribed the curriculum, or how explicit the textbook, the actions of the teacher in the classroom can greatly affect student learning. What teachers know and are able to do is of critical importance. In this area, TIMSS focuses on a set of indicators related to having highly qualified teachers in the classroom. These include: academic preparation and certification; teacher recruitment; teacher assignment; teacher induction; teaching experience; teaching styles; and professional development.

3.2.2.4 Classroom Activities and Characteristics

Although the school provides the general context for learning, it is in the classroom setting and through the guidance of the teacher that most teaching and learning take place. Aspects of the implemented curriculum that are most readily studied in the classroom include the curriculum topics that are actually taught, the pedagogical approaches used, the materials and equipment available, and the conditions under which learning takes place, including the size and composition of the class and the amount of classroom time devoted to mathematics and science education. Here the TIMSS contextual framework addresses several areas: curriculum topics taught; instructional time; homework; assessment; classroom climate; use of information technology; calculator use; emphasis on scientific investigation; and class size.

3.2.2.5 The Students

Students come to school from different backgrounds and with different experiences that affect their attitudes toward learning mathematics and science and their academic performance in these subjects. In this area TIMSS focuses on: students' home background and resources for learning; their prior experiences; and their attitudes toward learning.

3.3 The TIMSS 2003 Background Questionnaires

The TIMSS 2003 contextual framework served as the foundation in developing the TIMSS 2003 background questionnaires. As mentioned above, four types of background questionnaires were used to collect information regarding the contexts in which students learn mathematics and science.

- The **curriculum questionnaire** addressed issues of the intended national curriculum in mathematics and science. Four versions of this questionnaire were administered: fourth-grade mathematics, fourth-grade science, eighth-grade mathematics, and eight-grade science.
- The **school questionnaire** asked school principals or headmasters to provide information about the school contexts for the teaching and learning of mathematics and science. There were separate versions for fourth grade and eighth grade.
- The **teacher questionnaire** collected information about the teachers' preparation and professional development, their pedagogical activities, and the implemented curriculum. At fourth grade there was one questionnaire that addressed both mathematics and science, and at eighth grade there were separate versions for mathematics teachers and science teachers.
- The **student questionnaire** sought information about the students' home backgrounds and their experiences in learning mathematics and science. There were separate versions for fourth grade and eighth grade, and at eighth grade there were different versions for countries where eighth-grade science is taught as a single integrated subject and countries where it is taught as separate subjects (i.e., biology, chemistry, physics, earth science).

3.3.1 Development of the Background Questionnaires

Like the contextual framework, the TIMSS 2003 background questionnaires were developed through an iterative and widely collaborative process that spanned slightly more than one year. This process involved the TIMSS & PIRLS International Study Center, National Research Coordinators, the Questionnaire Item Review Committee (QIRC), and the IEA Data Processing Center. The process included a series of reviews of draft instruments, a field test of the questionnaires, a review of the field-test data, and a revision of the field-test instruments for use in the main data collection.

The development work began at the second NRC meeting in June 2001, when NRCs reviewed the TIMSS 1999 questionnaires in conjunction with the TIMSS 2003 contextual framework to advise what should be included in the 2003 assessment. Where questionnaire items had been used in the TIMSS 1999 international reports, NRCs decided that in general these items should be retained, preferably in the same form in order to measure trend. Items not reported in TIMSS 1999 were to be modified or deleted. NRCs also suggested to add or expand questions regarding the type of homework that students do, whether students get support for homework outside of school, the types of threats to safety that students experience, how teachers are licensed and evaluated, and the types of professional development that teachers undergo.

Working from the contextual framework and the TIMSS 1999 questionnaire review conducted by NRCs, staff at the International Study Center produced drafts of all the background questionnaires during the period of June through September 2001. The drafts were sent to members of the Questionnaire Item Review Committee for their review.[3] The first meeting of the Questionnaire Item Review Committee was held in October 2001 in Washington, D.C., at which the draft questionnaires were reviewed in detail. QIRC members suggested many improvements, as well as ways to reduce response burden by eliminating some questions thought to be less useful for reporting purposes. Following this meeting, the suggested revisions were implemented, and the revised drafts were submitted to further internal review at the ISC. The draft questionnaires were then provided to NRCs for their review at the Third TIMSS 2003 National Research Coordinators' Meeting, held in December 2001 in Madrid, Spain. NRCs suggested a number of improvements to the questionnaires that were to be field tested, and these revisions were implemented by the ISC during January 2002, in preparation for the field test. The field-test instruments were then provided to NRCs for translation, production, and administration.[4]

The TIMSS 2003 field test was conducted during April through June 2002. One of the primary purposes of the field test was to check across participating countries whether the questionnaires were appropriate for the measurement purposes for which they were designed. Although the questionnaires were adapted from previous versions, because there were a number of additions and refinements in the 2003 version it was necessary to field test them.[5] In all, 20 out of 26 countries participated in the field test at the fourth grade, and 41 of 48 countries participated at the eighth grade.

3 See Appendix A for a list of members of the Questionnaire Item Review Committee.

4 Please see Chapter 4 for more information about the translation and verification process.

5 The curriculum questionnaires were not administered in the field test.

After administering the field test, countries prepared their data files and sent them to the IEA Data Processing Center for checking and cleaning. After the field-test data were verified and transformed into the international format, they were sent to the International Study Center for analysis, and for review by the QIRC and NRCs. To facilitate review of the questionnaire data, the ISC prepared three data almanacs each for fourth and eighth grades, one for the school questionnaire, one for the teacher questionnaire, and one for the student questionnaire. For every country that participated, each almanac displayed student-weighted distributions of responses to each item in the questionnaires. For categorical variables, the weighted percentage of respondents choosing each option was shown together with the corresponding average student achievement in mathematics and science. For questions with numeric responses, the mean, mode, and selected percentiles were given. The almanacs were the basic data summaries that were used by ISC staff, the QIRC, and NRCs in assessing the quality of the field-test instruments and in making suggestions for the instruments to be used in the main data collection.

The initial review of the field-test results was conducted by the International Study Center in early July 2002. The questionnaire items were reviewed in terms of how well they worked both across countries and within individual countries. Based on this review, ISC staff made some improvements to the school, teacher, and student questionnaires, upon consultation with the QIRC. Also at this time, drafts of the curriculum questionnaires (which were not field tested) were completed.

At its second meeting, in July 2002 in Amsterdam, QIRC members reviewed the field-test results for the school, teacher, and student questionnaires, examining the statistics for each item and determining if there were any anomalies. Items that did not work well were deleted. The committee also discussed potential improvements suggested by the ISC, suggested modifications to some items, and arrived at a set of recommended changes to be brought before NRCs at their next meeting. The QIRC also proposed some refinements to the draft curriculum questionnaires.

During the latter half of July 2002, staff at the International Study Center prepared draft instruments for the main survey and documented the recommended changes from the field-test version for review by NRCs at the Fifth TIMSS 2003 National Research Coordinators' Meeting, held in late July and early August 2002 in Tunis, Tunisia. The draft instruments were well received and widely discussed by NRCs, who recommended a number of additional improvements. A substantial organizational change was made to the fourth grade teacher questionnaire, to facilitate data collection in countries where mathematics and science at fourth grade were taught by different teachers. Immediately after the NRC meeting, ISC staff finalized the instruments, and these were provided to NRCs during the latter part of August, for translation, production, and administration in the main TIMSS 2003 data collection, which was held during September through November 2002 in southern hemisphere countries and during February through July 2003 in northern hemisphere countries.

3.3.2 Content of the Background Questionnaires

The curriculum, school, teacher, and student questionnaires used in TIMSS 2003 were developed from the TIMSS 1999 questionnaires. While most of the questions were thematically similar in both assessments, some questions from 1999 were eliminated, some were modified with the intention of refining them, and some new questions were introduced in 2003, either as replacements for eliminated items or to provide additional information in areas deemed important to the study. In general, every effort was made to streamline the questionnaires in order to limit response burden. Based upon the guidelines specified in the contextual framework, new emphasis was placed upon the areas of teacher preparation and professional development, and the access to and use of technology for teaching and learning.

The organization of the questionnaires was improved so that the questions were more clearly organized into logical blocks, each with a heading. The design and layout also was improved to make the questionnaires easier to complete, especially where filter questions were used. Parallel questions were used in different questionnaires to measure the same constructs from different sources, and wherever possible the wording of such questions was identical. Questions that addressed the focus areas of teacher preparation and professional development, and use of technology for teaching and learning, were included in the four different questionnaire types.

The content of the TIMSS 2003 background questionnaires used to collect information about the contexts for learning mathematics and science is described below.

3.3.2.1 Curriculum Questionnaire

The fourth- and eighth-grade curriculum questionnaires for mathematics and science were addressed to National Research Coordinators, who were asked to supply information about their nation's mathematics and science curricula in the target grades, drawing on the expertise of curriculum specialists in their countries. The curriculum questionnaires were designed to collect basic information about the organization of and support for the intended mathematics and science curriculum in each country, and whether the mathematics and science topics included in the TIMSS 2003 assessment were included in the country's intended curriculum through the target grade. The four versions of the curriculum questionnaire were the same in structure and very similar in content, with the mathematics and science versions tailored to the subject matter and grade level wherever necessary. One notable difference was that the eighth-grade science curriculum questionnaire included a question asking whether eighth-grade science was taught as a single integrated subject or as separate science subjects.

Some of the central questions addressed in the curriculum questionnaire included:

- Is there a national curriculum in mathematics/science at the target grade?
- Does the country administer public examinations in mathematics/science that have consequences for individual students?
- What methods are used to support and monitor implementation of the national mathematics/science curriculum?
- How does the national curriculum address the issue of students with different levels of ability?
- What aspects of the teaching and learning of mathematics/science are emphasized in the national curriculum?
- What are the requirements for becoming a mathematics/science teacher, and is there a process to license or certify teachers?
- Are the topics included in the TIMSS 2003 assessment included in the national curriculum, and if so, for what proportion of students, and at what grades are the topics intended to be taught?

The complete contents of the TIMSS 2003 mathematics and science curriculum questionnaires at fourth and eighth grades are described in Exhibit 3.1.

3.3.2.2 School Questionnaire

The fourth- and eighth-grade school questionnaires were to be completed by the school principal or headmaster of each school sampled for the study. They were designed to collect information concerning some of the major factors thought to influence student achievement in mathematics and science. The fourth- and eighth-grade versions of the school questionnaire are nearly identical, although two of the questions are tailored to the appropriate grade. The school questionnaire was designed to be completed in about 30 minutes.

Exhibit 3. 1 **Content of the TIMSS 2003 Mathematics and Science Curriculum Questionnaires at the Eighth and Fourth Grades**

Item Number				Item Content	Description
Mathematics Grade 8	Mathematics Grade 4	Science Grade 8	Science Grade 4		
1	1	1	1	National curriculum	Whether the country has a national mathematics/science curriculum at the target grade, the year introduced, and whether under revision
-	-	2	-	Separate sciences	Whether science is taught as separate subjects by eighth grade, and the specific subjects and grades taught
2	2	3	2	Public examinations	Whether the country administers public examinations in mathematics/science that have consequences for individual students, the authority that administers such examinations, and the grades at which they are given
3	3	4	3	Methods used to help implement the national curriculum	Whether the country uses various methods to help monitor implementation of the national mathematics/science curriculum at the target grade
4	4	5	4	Specification of instructional time	Whether the national curriculum specifies the percentage of instructional time intended to be devoted to mathematics/science at various grades, and the percentage of time designated
5	5	6	5	Differentiation of the curriculum	How the national mathematics/science curriculum at the target grade addresses the issue of students with different levels of ability
6	6	7	6	Emphasis on approaches and processes	How much emphasis the national mathematics/science curriculum at the target grade places on various approaches and processes
7	7	-	-	Policy on calculator use	Whether the national mathematics curriculum contains statements/policies on the use of calculators at the target grade, and a brief description of such policies
-	-	8	7	Policy on emphasis given scientific inquiry	Whether the national science curriculum contains statements/policies about the emphasis that should be placed on scientific inquiry at the target grade, and a brief description of such policies
8	8	9	8	Policy on computer use	Whether the national mathematics/science curriculum contains statements/policies on the use of computers at the target grade, and a brief description of such policies

Exhibit 3. 1 **Content of the TIMSS 2003 Mathematics and Science Curriculum Questionnaires at the Eighth and Fourth Grades** *(...Continued)*

Item Number				Item Content	Description
Mathematics Grade 8	Mathematics Grade 4	Science Grade 8	Science Grade 4		
9	9	10	9	Preparation of teachers in how to teach the intended curriculum	Whether mathematics/science teachers at the target grade receive specific preparation in how to teach the intended curriculm as part of their pre-service or in-service education, and a brief description of such preparation
10	10	11	10	Teaching requirements	Whether mathematics/science teachers at the target grade must fulfill various requirements in order to teach
11	11	12	11	Licensure process	Whether there is a process to license or certify mathematics/science teachers at the target grade, and what entity licenses the teachers
12	12	13	12	The teaching of the TIMSS topics	Whether the TIMSS mathematics/science topics are included in the national curriculum through the target grade, the proportion of students intended to be taught the topics, and the grade(s) at which the topics are intended to be taught

Some of the main questions addressed in the school questionnaire were:

- What is the school climate like?
- What are the school's expectations of parents?
- How does the school organize mathematics/science instruction for students with different levels of ability?
- How difficult was it to fill mathematics/science teaching vacancies, and were any incentives used to recruit or retain teachers?
- What types of professional development activities did mathematics/science teachers engage in?
- How safe is the school environment?
- Is the school's capacity to provide instruction affected by a shortage of various resources?
- What is the availability of computers for educational purposes in the school, and how many have access to the Internet?

The complete contents of the TIMSS 2003 school questionnaires at fourth and eighth grades are described in Exhibit 3.2.

Exhibit 3. 2 **Content of the TIMSS 2003 School Questionnaires at the Eighth and Fourth Grades**

Item Number		Item Content	Description
Grade 8	Grade 4		
1	1	Grade levels	Grade range of the school
2	2	Enrollment	Total school enrollment in all grades and in the target grade
3	3	Community size	Size of the community in which the school is located
4	4	Absenteeism	Percentage of students absent from school on a typical school day
5	5	Stability/ mobility of student body	Percentage of students enrolled at the beginning of the school year who were still enrolled at the time of testing, and percentage of students who enrolled after the beginning of the school year
6	6	Students' background	Percentage of students who come from economically disadvantaged or affluent homes, and percentage of students whose native language is the language of the test
7	7	School climate	Principal's perception of teachers' job satisfaction and expectations for student achievement; of parental support and involvement; and of students' regard for school property and desire to do well in school
8	8	Principal's experience	Number of years as a principal of this school
9	9	Principal's time allocation	Percentage of time principal spends on various activities across the school year
10	10	Parental involvement	Whether the school expects parents to participate in various activities
11	11	Instructional time	Number of days per year and days per week the school is open for instruction, and number of hours of instructional time in a typical day
12	12	Differentation of mathematics curriculum	How the school organizes mathematics instruction for students with different levels of ability
13	13	Tracking in mathematics	Whether the students are grouped by ability in their mathematics classes
14	14	Enrichment/ remedial mathematics	Whether the school offers enrichment and remedial courses in mathematics

Exhibit 3. 2 **Content of the TIMSS 2003 School Questionnaires at the Eighth and Fourth Grades** *(...Continued)*

Item Number		Item Content	Description
Grade 8	Grade 4		
15	15	Differentiation of science curriculum	How the school organizes science instruction for students with different levels of ability
16	16	Tracking in science	Whether the students are grouped by ability in their science classes
17	17	Enrichment/ remedial science	Whether the school offers enrichment and remedial courses in science
18	18	Teacher vacancies	Difficulty in filling teacher vacancies in mathematics, science, and computer science/information technology (4th grade version does not ask about specific subjects)
19	19	Incentives for teachers	Whether the school uses incentives to recruit or retain teachers in mathematics, science, and/or other subjects (4th grade version does not ask about specific subjects)
20	20	Professional development	Frequency with which teachers participated in various types of professional development activities during the school year
21	21	Teacher evaluation	Whether the school uses various procedures in evaluating mathematics and science teachers
22	22	Student behavior	Frequency and severity of various problematic student behaviors occurring in the school
23	23	Instructional resources	Degree to which the school's capacity to provide instruction is affected by shortages or inadequacy of various resources
24	24	Computers	Number of computers available for educational purposes, and proportion of computers with access to the Internet
25	25	Technology support	Whether there is anyone available to help teachers use information and communication technology for teaching and learning, and description of that person

3.3.2.3 Teacher Questionnaire

The teacher questionnaires were designed to gather information about the classroom contexts for the teaching and learning of mathematics and science, and about the implemented curriculum in these subjects. For each participating school at the fourth grade, there was one teacher questionnaire addressed to the classroom teacher of the sampled class. At eighth grade, for each sampled school a single mathematics class was sampled for the TIMSS 2003 assessment.[6] The mathematics teacher of that class was asked to complete a mathematics teacher questionnaire, and the science teacher(s) of the students

6 In some circumstances it was necessary to sample two classes to yield the desired sample size. Please see Chapter 5 for more information on sample design.

in that class was asked to complete a science teacher questionnaire, which paralleled that for the mathematics teacher. Although the general background questions were essentially the same for all versions, questions pertaining to instructional practices, content coverage, and teachers' views about teaching the subject matter were tailored toward mathematics or science. Many questions, such as those related to classroom characteristics and activities, and homework and assessment, were answered with respect to the specific classes of the sampled TIMSS students. Because the fourth- and eighth-grade versions of the teacher questionnaire were designed to be similar in length, and because the fourth-grade version included questions about both mathematics and science, some questions had to be eliminated or shortened in the fourth-grade version.

Some of the primary questions addressed in the teacher questionnaire were:

- What is teachers' educational background, and do they have a teaching license or certificate?
- How many years of pre-service teacher training did teachers have, and how many years have they been teaching?
- How ready do teachers feel they are to teach various topics at the target grade?
- What types of professional development have teachers participated in?
- What is the teaching load of teachers, and how do they spend their time both during and outside the formal school day (eighth grade only)?
- What are teachers' attitudes toward teaching the subject matter, and their perceptions regarding school climate and school safety?
- What instructional activities are provided to the students in the TIMSS class, and how do the students spend their time during their mathematics and science lessons?
- Do various student- and resource-related factors limit how teachers instruct the students in the TIMSS class (eighth grade only)?
- What percentages of time are devoted to the various mathematics and science content areas in teaching the TIMSS class?
- When have the students in the TIMSS class been taught the topics included in the TIMSS 2003 assessment?
- Do students have calculators available to them, and how do they use them (mathematics only)
- Do students have computers available to them, and how do they use them?
- How much homework is assigned to students?

- How often are students given a test or examination, and what types of questions are included (eighth grade only)?

The TIMSS 2003 teacher questionnaires were designed to take about 45 minutes to complete. The complete contents of the TIMSS 2003 teacher questionnaires are described in Exhibit 3.3 for the eighth grade and in Exhibit 3.4 for the fourth grade.

3.3.2.4 Student Questionnaire

Each student in the sampled fourth- and eighth-grade TIMSS classes completed a student questionnaire, which sought information about the student's home background and resources for learning, their attitudes about mathematics and science, and their experiences in learning these subjects. The fourth- and eighth-grade versions of the student questionnaire were thematically and organizationally similar to each other. Some questions were identical in the two versions, while for other questions the language was simplified in the fourth-grade version or the specific content of the question was altered to be appropriate to the fourth grade. The fourth-grade questionnaire was shorter in length than the eighth-grade version.

As in TIMSS 1999, two versions of the eighth-grade questionnaire were used, a *general science version* intended for countries where eighth-grade science is taught as a single integrated subject, and a *separate science subjects version* intended for countries where eighth-grade science is taught as separate subject (e.g., biology, earth science, chemistry, physics); countries administered the version that was consistent with the way in which science instruction was organized at the eighth grade. 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 subjects version, the same 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. Although the two versions differed with respect to the science questions, the general background and mathematics-related questions were identical across the two forms.

The student questionnaire was designed to gather information on some of the major factors thought to influence student achievement in mathematics and science. Some of the central questions addressed in the student questionnaire included:

- What are students' general demographic backgrounds – age, gender, native language, country of origin, household size?
- What are the resources for learning in the students' homes?

- What is the educational attainment of the students' parents, and what are the students' own educational aspirations?
- What is students' affinity for learning mathematics and science, and how do they perceive success in and the utility of learning mathematics and science?
- What types of learning activities do students engage in in their mathematics and science lessons?
- Do students use a computer, where, and for what learning activities?
- What are students' perceptions about school climate and school safety?
- How do students spend their time outside of school?
- How much homework do students do?

The TIMSS 2003 student questionnaires were designed to take about 30 minutes to complete. The complete contents of the TIMSS 2003 student questionnaires are described in Exhibit 3.5 for the eighth grade and in Exhibit 3.6 for the fourth grade.

Exhibit 3. 3 **Content of the TIMSS 2003 Mathematics and Science Teacher Questionnaires at the Eighth Grade**

Item Number			
Mathematics Teacher Questionnaire	**Science Teacher Questionnaire**	**Item Content**	**Description**
1	1	Age	Teacher's age
2	2	Gender	Teacher's gender
3	3	Teaching experience	Number of years as a teacher
4	4	Formal education	Highest level of formal education completed by the teacher
5	5	Teacher training	Number of years of pre-service teacher training completed by the teacher
6	6	Major area of study	Teacher's major area of study during post-secondary education
7	7	Teaching requirements	Requirements the teacher had to satisfy in order to become a teacher
8	8	Teaching license	Whether the teacher has a teaching license or certificate, and the type of license
9	9	Preparation to teach	How ready the teacher feels to teach the topics included in the TIMSS mathematics/science test
10	10	Teaching load	Number of periods for which the teacher is formally scheduled per week for various activities, and number of minutes in a period
11	11	Extra working time	Number of hours teacher spends on teaching-related activities outside the formal school day
12	12	Teacher interactions	Frequency of various types of interactions the teacher has with colleagues
13	13	Professional development	Whether the teacher participated in various types of professional development activities
14	14	Attitudes toward subject	Teacher's beliefs about the nature of mathematics/science and how the subject should be taught.
15	15	School setting	Teacher's perceptions about the adequacy of the school facility and about school safety
16	16	School climate	Teacher's perception of teachers' job satisfaction and expectations for student achievement; of parental support and involvement; and of students' regard for school property and desire to do well in school
17	17	Class size	Number of students in the sampled class
18	18	Time spend teaching subject	Minutes per week the teacher teaches mathematics/science to the sampled class
19	19	Textbook	Whether a textbook(s) is used as a primary or supplementary resource
20	20	Student learning activities	Percentage of time students spend doing various learning activities in a typical week
21	21	Content-related activities	Frequency with which the teacher asks students to do various content-related activities in mathematics/science

Exhibit 3. 3 **Content of the TIMSS 2003 Mathematics and Science Teacher Questionnaires at the Eighth Grade** *(...Continued)*

Item Number			
Mathematics Teacher Questionnaire	**Science Teacher Questionnaire**	**Item Content**	**Description**
22	22	Factors limiting teaching	Extent to which the teacher perceives various student and resource factors to limit teaching
23	23	Emphasis on content areas	Percentage of time spent on mathematics/science content areas over the course of the year
24	24	Topic coverage	When the students were taught the TIMSS mathematics/science topics, by content area
25	-	Calculator use policy	Whether the students are permitted to use calculators during mathematics lessons
26	-	Calculator availability	Proportion of students that have access to calculators during mathematics lessons
27	-	Graphing calculator availability	Proportion of students that have access to graphing calculators during mathematics lessons
28	-	Calculator use	Frequency with which the students use calculators for various learning activities
29	-	Calculators in test/exams	How often the students are allowed to use calculators during tests or examinations
30	25	Computer availability	Whether the students have access to computers during mathematics/science lessons and whether computers have access to Internet
31	26	Computer use	Frequency with which the students use computers for various learning activities
32	27	Homework	Whether the teacher assigns mathematics/science homework
33	28	Frequency of homework	How often the teacher assigns mathematics/science homework
34	29	Amount of homework	Number of minutes it would take an average student to complete a mathematics/science homework assignment
35	30	Type of homework	Frequency with which the teacher assigns various types of homework
36	31	Use of homework	How often the teacher uses mathematics/science homework for various purposes
37	32	Assessment	Frequency with which the teacher gives a mathematics/science test or examination
38	33	Question format	Item formats the teacher typically uses in mathematics/science tests or examinations
39	34	Type of questions	Types of questions the teacher uses in mathematics/science tests or examinations

Exhibit 3. 4 **Content of the TIMSS 2003 Teacher Questionnaire at the Fourth Grade**

Item Number	Item Content	Description
1	Age	Teacher's age
2	Gender	Teacher's gender
3	Teaching experience	Number of years as a teacher
4	Formal education	Highest level of formal education completed by the teacher
5	Teacher training	Number of years of pre-service teacher training completed by the teacher
6	Major area of study	Teacher's major area of study during post-secondary education
7	Teaching requirements	Requirements the teacher had to satisfy in order to become a teacher
8	Teaching license	Whether the teacher has a teaching license or certificate, and the type of license
9	School climate	Teacher's perception of teachers' job satisfaction and expectations for student achievement; of parental support and involvement; and of students' regard for school property and desire to do well in school
10	School setting	Teacher's perceptions about the adequacy of the school facility and about school safety
11	Teacher interactions	Frequency of various types of interactions the teacher has with colleagues
12	Preparation to teach mathematics	How ready the teacher feels to teach the topics included in the TIMSS mathematics test
13	Professional development in mathematics	Whether the teacher participated in various types of professional development activities for mathematics teaching
14	Mathematics class size	Number of students in the sampled class for mathematics, and number of those in the fourth grade
15	Time spend teaching mathematics	Minutes per week the teacher teaches mathematics to the sampled class
16	Mathematics textbook	Whether a textbook(s) is used as a primary or supplementary resource in teaching mathematics
17	Student learning activities in mathematics	Percentage of time students spend doing various learning activities in a typical week of mathematics lessons
18	Calculator use policy	Whether the students are permitted to use calculators during mathematics lessons
19	Calculator availability	Proportion of students that have access to calculators during mathematics lessons
20	Calculator use	Frequency with which the students use calculators for various learning activities
21	Calculators in test/exams	How often the students are allowed to use calculators during tests or examinations
22	Computer availability for mathematics	Whether the students have access to computers during mathematics lessons and whether computers have access to the Internet
23	Computer use in mathematics	Frequency with which the students use computers for various learning activities in mathematics

Exhibit 3. 4 **Content of the TIMSS 2003 Teacher Questionnaire at the Fourth Grade**
(...Continued)

Item Number	Item Content	Description
24	Mathematics content-related activities	Frequency with which the teacher asks students to do various content-related activities in mathematics
25	Emphasis on mathematics content areas	Percentage of time spent on mathematics content areas over the course of the year
26	Mathematics topic coverage	When the students were taught the TIMSS mathematics topics, by content area
27	Mathematics homework	Whether the teacher assigns mathematics homework
28	Frequency of mathematics homework	How often the teacher assigns mathematics homework
29	Amount of mathematics homework	Number of minutes it would take an average student to complete a mathematics homework assignment
30	Preparation to teach science	How ready the teacher feels to teach the topics included in the TIMSS science test
31	Professional development in science	Whether the teacher participated in various types of professional development activities for science teaching
32	Science class size	Number of students in the sampled class for science, and number of those in the fourth grade
33	Time spend teaching science	Minutes per week the teacher teaches science to the sampled class
34	Science textbook	Whether a textbook(s) is used as a primary or supplementary resource in teaching science
38	Student learning activities in science	Percentage of time students spend doing various learning activities in a typical week of science lessons
35	Computer availability for science	Whether the students have access to computers during science lessons and whether computers have access to the Internet
36	Computer use in science	Frequency with which the students use computers for various learning activities in science
37	Science content-related activities	Frequency with which the teacher asks students to do various content-related activities in science
39	Preparation to teach science	How ready the teacher feels to teach the topics included in the TIMSS science test
40	Science homework	Whether the teacher assigns science homework
41	Frequency of science homework	How often the teacher assigns science homework
42	Amount of science homework	Number of minutes it would take an average student to complete a science homework assignment

Exhibit 3. 5 **Content of the TIMSS 2003 Student Questionnaire at the Eighth Grade**

Item Number			
General science version	**Separate science subjects version**	**Item Content**	**Description**
1	1	Age	Month and year of student's birth
2	2	Gender	Student's gender
3	3	Language	Student's frequency of use of the language of the test at home
4	4	Books in the home	Number of books in the student's home
5	5	Home possessions	Educational resources and general possessions in the student's home
6	6	Parents' education	Highest level of education completed by mother and father
7	7	Educational expectations	Level of education the student expects to complete
8	8	Liking math-ematics	How much the student likes and feels competent at mathematics
9	9	Valuing math-ematics	Importance and value the student attributes to mathematics
10	10	Learning activi-ties in math-ematics	Frequency with which student does various learning activities in mathematics lessons
11	-	Liking science	How much the student likes and feels competent at science
12	-	Valuing science	Importance and value the student attributes to science
13	-	Learning activi-ties in science	Frequency with which student does various learning activities in science lessons
-	11	Study biology	Whether the student is studying biology this year
-	12	Liking biology	How much the student likes and feels competent at biology
-	13	Valuing biology	Importance and value the student attributes to biology
-	14	Learning activi-ties in biology	Frequency with which student does various learning activities in biology lessons
-	15	Study earth science	Whether the student is studying earth science this year
-	16	Liking earth science	How much the student likes and feels competent at earth science
-	17	Valuing earth science	Importance and value the student attributes to earth science
-	18	Learning activi-ties in earth science	Frequency with which student does various learning activities in earth science lessons

Exhibit 3. 5 **Content of the TIMSS 2003 Student Questionnaire at the Eighth Grade**
(...Continued)

Item Number			
General science version	Separate science subjects version	Item Content	Description
-	19	Study chemistry	Whether the student is studying chemistry this year
-	20	Liking chemistry	How much the student likes and feels competent at chemistry
-	21	Valuing chemistry	Importance and value the student attributes to chemistry
-	22	Learning activities in chemistry	Frequency with which student does various learning activities in chemistry lessons
-	23	Study physics	Whether the student is studying physics this year
-	24	Liking physics	How much the student likes and feels competent at physics
-	25	Valuing physics	Importance and value the student attributes to physics
-	26	Learning activities in physics	Frequency with which student does various learning activities in physics lessons
14	27	Computers	Whether student uses a computer, where uses it, and frequency with which student uses a computer for various educational activities
15	28	School climate	Student's affinity for school, and perception of other students' motivation in school and teachers' expectations and care of students
16	29	Safety in school	Whether the student experienced being the object of problematic behaviors by other students
17	30	Out-of-school activities	Frequency with which student does various non-academic activities and homework outside of school
18	31	Extra lessons/ tutoring	Frequency of extra lessons or tutoring in mathematics and science
19	32	Mathematics homework	Frequency and amount of mathematics homework
20	32	Science homework	Frequency and amount of science homework
21	33	Persons living in home	Number of people living at home
22	34	Parents born in country	Whether mother and father were born in country
23	35	Student born in country	Whether student was born in country, and if not age at which student emigrated

Exhibit 3. 6 **Content of the TIMSS 2003 Student Questionnaire at the Fourth Grade**

Item Number	Item Content	Description
1	Age	Month and year of student's birth
2	Gender	Student's gender
3	Language	Student's frequency of use of the language of the test at home
4	Books in the home	Number of books in the student's home
5	Home possessions	Educational resources and general possessions in the student's home
6	Liking mathematics	How much the student likes and feels competent at mathematics
7	Learning activities in mathematics	Frequency with which student does various learning activities in mathematics lessons
8	Liking science	How much the student likes and feels competent at science
9	Learning activities in science	Frequency with which student does various learning activities in science lessons
10	Computers	Whether student uses a computer, where uses it, and frequency with which student uses a computer for various educational activities
11	School climate	Student's affinity for school, and perception of other students' motivation in school and teachers' expectations and care of students
12	Safety in school	Whether the student experienced being the object of problematic behaviors by other students
13	Out-of-school activities	Frequency with which student does various non-academic activities and homework outside of school
14	Extra lessons	Frequency of extra lessons or tutoring in mathematics and science
15	Mathematics homework	Frequency and amount of mathematics homework
16	Science homework	Frequency and amount of science homework
17	Persons living in home	Number of people living at home
18	Parents born in country	Whether mother and father were born in country
19	Student born in country	Whether student was born in country, and if not age at which student emigrated

References

Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzalez, E.J., Chrostowski, S.J., & O'Connor, K.M. (2003), *TIMSS Assessment Frameworks and Specifications 2003 (2nd ed.)*, Chestnut Hill, MA: Boston College.

4

Chapter 4

Translation and Cultural Adaptation of the TIMSS 2003 Instruments

Steven J. Chrostowski and Barbara Malak

4.1 Overview

The TIMSS 2003 data collection instruments (achievement tests and background questionnaires) were developed and prepared in English by the TIMSS & PIRLS International Study Center (ISC) at Boston College, with contribution from the National Research Coordinators (NRCs) of participating countries. The assessment instruments were subsequently translated by the participating countries into their local languages of instruction, 34 in total. Of the 49 countries and four Benchmarking participants in the TIMSS 2003 data collection, 17 collected data in two languages and one in three languages. The most common languages of testing were English (18 countries) and Arabic (10 countries).

The translation process was designed to ensure standardization of instruments across countries. Each country was expected to follow procedures established by the ISC for translating the test instruments into the national language and cultural context. These guidelines were provided to all NRCs in the *TIMSS 2003 Survey Operations Manual* (TIMSS, 2002a), and were further elaborated and discussed at relevant NRC meetings.

Before the translated instruments were administered to students, they went through a rigorous process of translation verification and review to ensure that they were translated accurately and were internationally comparable. This process was managed by the IEA Secretariat in Amsterdam. As a critical part of the translation verification process, the translated instruments for each country were checked by independent verifiers against the TIMSS 2003 international version to assess the comparability of translation. Verifiers reviewed the translated instruments and documented any deviations from the international version. National Research Coordinators received a Translation Verification Report that identified corrections or improvements considered necessary by the verifiers. When all necessary corrections had been implemented by NRCs, the International Study Center reviewed the revised instruments, suggested additional improvements, and gave final approval to the countries to print and administer the materials.

Translation verification was conducted both for the TIMSS 2003 field test and the main data collection.[1] For the achievement tests, the bulk of the translation effort took place prior to the field test, as there were few changes to the test items selected from the field test for use in the main data collection. The background questionnaires, however, were substantially revised after the field test and therefore required a second major translation effort. For the 44 participants in the field test, verification was conducted at both stages of the study. This allowed these countries to practice the translation procedures prior to the main data collection. It also gave them an additional opportunity to check the translations of items used in both the field test and main data collection.

All countries that participated in TIMSS 2003 submitted their most important instruments for translation verification. However, some countries did not submit for verification instruments in languages which were administered to a very small proportion of the sample. Such countries, however, used instruments that were translated and verified for another country (for example, Egypt used Lebanon's French and English instruments in a few schools).

4.2 Translation of Instruments

The TIMSS 2003 survey translation guidelines called for two independent translations of each test instrument from English to the target language. A translation review team then reviewed and compared the two translations to arrive at a final version of the translated instruments.

The prescribed translation procedure at the National Research Centers included the following steps:

1. Identify the target language(s), i.e. the language(s) of instruction.
2. Identify translators for two independent translations.
3. Translate instruments and adapt as necessary.
4. Confer and reconcile the two independent translations.
5. Document all cultural adaptations.

1 The TIMSS 2003 field test was conducted during April-June 2002, and the main data collection was conducted during September-November 2002 for southern hemisphere countries and February-July 2003 for northern hemisphere countries.

In practice, because of scarcity of resources and/or time allotted for translation, several countries used only one person to translate the instruments, often the NRC, who generally was the person most competent for this task.

4.2.1 Instruments To Be Translated

Each country had to translate the following materials into the language of instruction at each grade:

- 14 blocks of mathematics achievement items and 14 blocks of science achievement items (see next section);
- the student directions for the assessment;
- the background questionnaires – Student Questionnaire, Teacher Questionnaire, and School Questionnaire;[2]
- the School Coordinator Manual;
- the Test Administrator Manual, including the Test Administration Form; and
- the Scoring Guides for the Constructed-Response Items.

Countries testing in English did not have to translate the instruments, but were required to adapt the American-English of the originals to the vernacular, and make whatever adaptations were necessary for cultural reasons. The mathematics and science tests and the background questionnaires underwent the translation verification process, whereas the manuals and scoring guides did not. The International Study Center provided each country with electronic files containing all of the material to be translated to facilitate the translation.

4.2.2 Identification of the Target Language

Each NRC identified the language or languages to be used for testing (see Exhibit 4.1) and the geographical or political areas associated with them. If a single translation was prepared within a country, translators needed to ensure that the translation was acceptable to all of the dialects of the language in which the assessment was to be administered. Professionals from these dialects were to be involved in adapting the instruments and testing materials.

2 At the eighth grade only, there are different versions of the student questionnaire for countries that teach science as a single general/integrated subject and for countries that teach science as separate subjects at the eighth grade, and there are separate versions of the teacher questionnaire for mathematics and science teachers.

Exhibit 4.1 **TIMSS 2003 Translation Verification**

Country	Grade 8	Grade 4	Language(s) of Test	Materials Verified
Argentina	√		Spanish	Adapted Chilean version of test booklets and questionnaires
Armenia	√	√	Armenian	Translated test booklets and questionnaires
Australia	√	√	English	Adapted international English version of full set of instruments
Bahrain	√		Arabic, English	Adapted Egyptian Arabic version of booklets and questionnaires
Belgium (Flemish)	√	√	Dutch	Translated full set of instruments
Botswana	√		English	Translated full set of instruments
Bulgaria	√		Bulgarian	Translated full set of instruments
Chile	√		Spanish	Translated full set of instruments
Chinese Taipei	√	√	Chinese	Translated full set of instruments
Cyprus	√	√	Greek	Translated full set of instruments
Egypt	√		Arabic, English, French	Translated Arabic version of test booklets and questionnaires
England	√	√	English	Adapted international English version of test items and questionnaires
Estonia	√		Estonian, Russian	Translated full set of instruments in both languages
Ghana	√		English	Adapted international English version of full set of instruments
Hong Kong, SAR	√	√	Chinese, English (grade 8 only)	Translated full set of instruments in Chinese and adapted international English version of questionnaires
Hungary	√	√	Hungarian	Translated full set of instruments
Indonesia	√		Indonesian	Translated full set of instruments
Iran, Islamic Rep. of	√	√	Farsi	Translated full set of instruments
Israel	√		Hebrew, Arabic	Translated full set of instruments in Hebrew, translated test blocks, test booklets, and student questionnaire in Arabic (teacher and school questionnaires not administered in Arabic)
Italy	√	√	Italian	Translated full set of instruments
Japan	√	√	Japanese	Translated full set of instruments
Jordan	√		Arabic	Translated full set of instruments
Korea, Rep. of	√		Korean	Translated full set of instruments
Latvia	√	√	Latvian, Russian	Translated full set of instruments in Latvian
Lebanon	√		French, English	Translated French and adapted international English versions of test booklets and questionnaires
Lithuania	√	√	Lithuanian	Translated full set of instruments
Macedonia, Rep. of	√		Macedonian, Albanian	Translated full set of instruments in both languages
Malaysia	√		Malay	Translated full set of instruments
Moldova, Rep. of	√	√	Moldavian, Russian	Adapted Romanian and Russian versions of test booklets and questionnaires

Exhibit 4.1 **TIMSS 2003 Translation Verification** *(...Continued)*

Country	Grade 8	Grade 4	Language(s) of Test	Materials Verified
Morocco	√	√	Arabic	Translated test booklets and questionnaires
Netherlands	√	√	Dutch	Translated full set of instruments
New Zealand	√	√	English, Maori (grade 4 only)	Adapted international English version of full set of instruments, translated test blocks and student questionnaire in Maori (teacher and school questionnaires not administered in Maori)
Norway	√	√	Bokmål, Nynorsk	Translated full set of instruments in both languages
Palestinian Nat'l Auth.	√		Arabic, English	Adapted Jordanian Arabic version of test blocks and questionnaires
Philippines	√	√	English	Adapted international English version of full set of instruments
Romania	√		Romanian, Hungarian	Translated full set of instruments in Romanian
Russian Federation	√	√	Russian	Translated full set of instruments
Saudi Arabia	√		Arabic	Adapted Egyptian version of test booklets and questionnaires
Scotland	√	√	English	Adapted international English version of test items and questionnaires (tests same version as England)
Serbia	√		Serb	Translated full set of instruments
Singapore	√	√	English	Adapted international English version of full set of instruments
Slovak Republic	√		Slovak, Hungarian	Translated full set of instruments in both languages
Slovenia	√	√	Slovene	Translated full set of instruments
South Africa	√		English, Afrikaans	Adapted international English and translated Afrikaans versions of full sets of instruments
Sweden	√		Swedish	Translated full set of instruments
Syrian Arab Republic	√		Arabic	Adapted Egyptian version of test booklets
Tunisia	√	√	Arabic	Translated test booklets and questionnaires
United States	√	√	English	Adapted international English version of test items and questionnaires
Yemen		√	Arabic	Adapted Egyptian version of test booklets and questionnaires
Benchmarking Participants				
Basque Country, Spain	√		Basque, Castilian	Translated full set of instruments in both languages
Indiana State, US	√	√	English	Adapted international English version of test items and questionnaires (same version as United States)
Ontario Province, Can.	√	√	English, French	Adapted international English and translated French versions of full sets of instruments
Quebec Province, Can.	√	√	English, French	Adapted international English and translated French versions of full sets of instruments

Note: Full set of instruments consists of test blocks, test booklets, background questionnaires, and trend items if applicable.

4.2.3 Identification of Translators for Two Independent Translations

Translators were expected to have an excellent knowledge of both English and the target language and experience in the subject matter. For the achievement tests, the translation procedure required four translators for each target language, two with expertise in mathematics education and two in science education. Where subject matter experts were not available to act as translators, the translators were expected to work closely with subject matter specialists to ensure that the content and difficulty of the items did not change as a result of the translation. If a country could not employ all the required translators, the NRC played a major role in translating and/or verifying the translation of the instruments.

Translators of general text materials (student, teacher, and school questionnaires, and procedural manuals) did not need to be subject-matter specialists, so only two translators were necessary for these documents.

4.2.4 Translation and Cultural Adaptation of Instruments

Translators were provided with guidelines and procedures to follow in translating the data collection instruments and adapting them to their national cultural context. The guidelines were designed to yield translations that were as close as possible to the international (English) version of the survey instruments, while allowing for cultural adaptations where necessary. Translators were cautioned not to change the meaning or the difficulty level of an achievement item during the translation process. The primary concern was to convey the same meaning and style of the items as closely as possible to the international version.

The translators' tasks included:

- identifying and minimizing cultural differences;
- finding equivalent words and phrases;
- ensuring that the reading level was the same in the target language as in the original international version;
- ensuring that the essential meaning of the text did not change;
- ensuring that the difficulty level of achievement items did not change; and
- making changes in the instrument layout required due to translation.

As described in Chapter 2, the TIMSS 2003 assessment uses a matrix-sampling technique that involves dividing the entire item pool into a set of unique item blocks, distributing these blocks across a set of test booklets, and rotating the booklets among the students. To facilitate the creation of the student booklets, the items in the assessment pool are first grouped into blocks of items. These then become the building blocks from which the student booklets are assembled. The entire item pool at each grade is divided into 14 blocks of mathematics items and 14 blocks of science items. The 28 blocks of items are distributed across 12 student booklets. To enable linking between booklets, each block appears in two, three, or four different booklets. Each student completes one booklet consisting of six blocks of mathematics and science items. Translation of the assessment was based on blocks rather than booklets. Countries translated each block once and entered the translated text into the electronic file for the appropriate test booklets.

Translators were permitted to adapt the text as necessary to make unfamiliar contextual terms culturally appropriate. Acceptable adaptations included changes in the names of seasons, people, places, animals, plants, currencies, etc. Exhibit 4.2 shows a list provided to translators detailing the types of adaptations that were acceptable.

Exhibit 4.2 **Types of Acceptable Cultural Adaptations**

Type of Change	Specific Change from:	Specific Change to:
Punctuation/Notation	decimal point	decimal comma
	place value comma	space
Units	centimeters	inches
	liters	quarts
	ml	mL
Proper nouns	Ottawa	Oslo
	Mary	Maria
Common nouns	robin	kiwi
	elevator	lift
Spelling	center	centre
Verbs (not related to content)	skiing	sailing
Usage	Bunsen burner	hot plate

Translators were allowed to change terms and expressions that were not familiar in their national culture, as long as the change would not affect the substance of the item. It was important, however, that translators not change any of the following when they modified the text of an item:

- the meaning of the item;
- the reading level of the text;

- the difficulty level of the item; and
- the likelihood of another possible correct answer for the item.

Although item writers and reviewers attempted to write and select items that would readily translate into the languages of the participating countries, occasionally an item proved problematic for translators. In those instances, the International Study Center was to be notified and a corresponding statement was to be included in the NRC Survey Activities Report.

4.2.5 Review of Independent Translations for Consensus

After the two translations were completed, they were compared item by item, and any differences were reconciled. In most cases, by discussing the differences in the translations of a particular item, the translators were able to agree on the version that was most appropriate for the study. A third translation expert was to be contacted if any disagreement in the translation remained.

4.2.6 Documentation of Cultural Adaptations

After a single translation had been agreed upon, the Cultural Adaptation Form was used to record all adaptations made to the achievement and questionnaire items during translation. The description of each adaptation included the international (English) term, the translated term for test items or the adapted term for questionnaire items, and an explanation of why that term was used. Translators also noted if there were any other changes in the translation. This documentation was used during translation verification, and during the achievement item analysis and review where necessary, to evaluate the quality of the translations.

4.3 Verification of Instrument Translations

Each translation went through a rigorous verification process that included verification by an international translation company, review by the International Study Center, verification of the item translations at the national centers and a check by International Quality Control Monitors.

4.3.1 International Verification of the Translations

After the final translated version of each instrument was developed, the translation was checked through an external verification process. The IEA Secretariat developed and managed the translation verification process working closely with two international translating companies with reputations for excellence, Bowne Global Solutions (formerly Berlitz), based in Luton, England, and Capstan, based in Louvain-le-Neuve, Belgium. Bowne and

Capstan staff were to document all errors and omissions and make suggestions for improvements so that National Research Coordinators could revise and improve their instruments.

Translators selected by Bowne and Capstan to serve as translation verifiers for TIMSS were required to have first-language experience in the target language, formal credentials as translators working in English, and to live and work in the target country. When the last condition could not be met, verifiers were expected to maintain close contact with the country and its culture.

4.3.1.1 Submission of Instruments for Verification

NRCs were required to send (no later than six weeks before printing) the following instruments for each grade assessed to the IEA Secretariat in preparation for external translation verification:

- one copy of the test blocks of achievement items (14 blocks of mathematics items and 14 blocks of science items) and the accompanying instructions for students;
- one set of the assembled test booklets (booklets 1 through 12); and
- one copy of the student questionnaire, teacher questionnaire(s),[3] and school questionnaire.

All countries that participated in the TIMSS 2003 data collection submitted national versions of instruments for translation verification (see Exhibit 4.1).

4.3.1.2 The Translation Verification Process

The primary task of translation verifiers was to evaluate the accuracy of the translation and layout of the survey instruments. Verifiers were asked to make recommendations for improvements in the translation, when necessary, and also to alert the national centers to any deviation from the international version in the layout of the translated instruments.

Verifiers were provided with general information about the study and the design of the instruments. They also received materials describing the translation procedures used by the national centers and cultural adaptations deemed acceptable, along with detailed instructions for reviewing the instruments.[4] The verification guidelines emphasized the importance of maintaining the meaning, difficulty level, and format of each item while allowing for cultural adaptations as necessary.

3 As noted above, at fourth grade there is one teacher questionnaire, and at eighth grade there are separate mathematics and science teacher questionnaires.

4 Materials provided to verifiers included *Guidelines for the Translation Verification of the TIMSS 2003 Main Survey Instruments (TIMSS, 2001).*

Each verifier received a package consisting of:

- the international version of each survey instrument (test blocks, test booklets, and background questionnaires);
- a set of the translated national instruments to be verified, along with the Cultural Adaptation Forms;
- a copy of the instructions given to the translators in each country;
- guidelines for translation verification, including instructions for verifying the content and layout of the survey instruments and the instructions to students;
- translation verification control forms to be completed for each instrument; and
- translation verification report forms to be completed for each instrument.

For TIMSS 2003 countries that also participated in prior cycles of the study, verifiers were responsible for ensuring that the translated version of the trend items was identical to that administered in 1995 at fourth grade and 1999 at eighth grade. Accordingly, verifiers reviewing instruments for trend-study countries also received the following:

- the translated trend items used in that country in 1995 for fourth grade and/or 1999 for eighth grade; and
- a trend item verification form.

In addition to receiving detailed written instructions, verifiers had the opportunity to discuss with the IEA coordinator any problems they encountered while performing their task.

4.3.1.3 Translation Verification Reports

Two types of reports were written by the translation verifier to document the verification process. First, the translation verifier completed a translation verification control form for each instrument. This cover sheet served as a checklist indicating which materials had been verified and whether or not deviations were found in the instruments, and including the verifier's opinion about the general quality of the translation. Second, where in the judgment of the verifier the translated version of an achievement or questionnaire item deviated from the international version, the translation verifier completed a translation verification report form with entries made indicating:

- the location of the translation deviation (page and item number);
- the severity of the deviation (using a severity code as defined below);
- a description of the change; and
- a suggested alternative translation.

These records were used to document the quality of the translations and the comparability of the testing materials in each country.

The severity codes ranged from 1 (serious error) to 4 (acceptable adaptation).[5] The severity codes were:

Code 1 – Major Change or Error: Examples include incorrect ordering of choices in a multiple-choice item; omission of a graph; complete omission of an item; incorrect translation of text such that the answer is indicated by the question; an incorrect translation that changes the meaning or difficulty of the question; incorrect ordering of the items or placement of the graphics.

Code 2 – Minor Change or Error: Examples include spelling errors that do not affect comprehension; misalignment of margins or tabs; incorrect font or font size; discrepancies in the headers or footers of the document.

Code 3 – Suggestions for Alternative: The translation may be adequate, but the verifier suggests a different wording for the item.

Code 4 – Acceptable Changes: The verifier identifies changes that are acceptable and appropriate adaptations of the item, e.g., where a reference to winter is changed from January to July for the southern hemisphere.

The layout of the documents was also reviewed during the verification process for any changes or deviations. Exhibit 4.3 details the layout issues that were considered and checked for each survey instrument.

Exhibit 4.3 **Layout Issues Considered in Verification**

Layout Issues	Verification Details
Instructions	Test items should not be visible when the test booklet was opened to the Instructions section.
Items	All items should be included in the same order and location as in the international version.
Response options	Response options should appear in the same order as in the international version.
Graphics	All graphics should be in the same order and modifications should be limited to necessary translation of text or labels.
Font	Font and font size should be consistent with the international version.
Word emphasis	Word emphasis should remain the same as in the international version. If the form of emphasis was not appropriate for the given language, an acceptable alternate form of emphasis should have been used (e.g., italics instead of capital letters).
Shading	Items with shading should be clear and text legible.
Page and item identification	Headers and footers that include booklet, page, and item identification should be present.
Pagination	Page breaks should correspond with the international version of the instruments.

5 When in doubt as to the severity of the deviation, verifiers used code 1.

If the layout of an instrument differed in any way from the international version, an entry was made in the translation verification report form indicating the location of the deviation, the severity of the deviation, and a description of the change in the layout. If necessary and appropriate, a suggestion for improving the layout was included.

For countries that participated in prior cycles of TIMSS, verifiers also completed a trend item verification form, indicating whether there was any difference in translation or format of the trend items between the 2003 version and the 1995 version for fourth grade and 1999 version for eighth grade, with a description of the nature of the change.

The completed translation verification forms were sent to NRCs and an additional copy was sent to the International Study Center at Boston College and the IEA Data Processing Center (DPC) in Hamburg, Germany. The NRCs were responsible for reviewing the reports and revising the instruments, at their own discretion, based on the translation verifiers' suggestions.

Although generally countries complied very well with the requirements for translation verification, a number of countries did not submit for verification instruments in languages that were used . Bahrain did not submit its English version of instruments for review; Egypt did not submit its English and French versions of instruments, which were borrowed from Lebanon, for review; Hong Kong did not submit its English version of achievement tests for review; Latvia did not submit its Russian version of instruments (which were borrowed from the Russian Federation) for review; the Palestinian National Authority did not submit its English version of instruments for review; Romania did not submit its Hungarian version of instruments (which were borrowed from Hungary and not adapted) for review; and Syria did not submit its background questionnaires for review. The following countries submitted test booklets but not blocks or test blocks but not booklets for review: Argentina, Armenia, Bahrain, Cyprus, Egypt, England, Lebanon, Moldova, Morocco, Palestinian National Authority, Saudi Arabia, Scotland, Syria, Tunisia, United States, and Yemen.[6] The following countries did not submit Cultural Adaptations Forms along with their instruments for review: Bahrain, Cyprus (tests), Egypt, Indonesia (questionnaires), Japan, Jordan, Latvia (tests), Lebanon, Lithuania (tests), Morocco, Syria, Tunisia, and Yemen.

6 Due to time limitations, southern hemisphere countries (Australia, Botswana, Chile, Malaysia, New Zealand, Singapore, South Africa) were required to submit only the test blocks and not the test booklets to the IEA Secretariat for review.

4.3.2 International Study Center Review

For a final review, NRCs were required to submit a print-ready copy of the achievement test booklets and questionnaires to the TIMSS & PIRLS International Study Center at Boston College, after implementing the suggestions of the translation verifiers.

For all countries, achievement and questionnaire items were compared with the international version to identify any changes in text, graphics, and format, and the test booklets and questionnaires were reviewed to identify any changes in layout. The text was reviewed for format, and items were checked to ensure that they had identical translations in the stem and options across different booklets.

For trend countries, each trend item was compared to the 1995 translated version for fourth grade and the 1999 translated version for eighth grade to note if any change had been made. When the language of these items was not familiar to the reviewer, the NRC was asked about any apparent changes.

NRCs were provided with a list of any deviations identified by the International Study Center that went beyond those recorded in the translation verification reports. NRCs used these comments to correct errors prior to printing, again at their own discretion. Countries that did not allot enough time for this step of the translation and review process were not required to submit their instruments to the ISC prior to printing, so as not to jeopardize their schedule for administering the assessment. The following countries submitted their instruments to the International Study Center for final review after printing: Armenia, Bahrain, Egypt, Japan, Korea, Lebanon, Morocco, Palestinian National Authority, Slovenia, Syria, Yemen, Ontario, and Quebec. Although the Philippines submitted instruments for review prior to printing, no corrections based on IEA or ISC review were implemented prior to printing.

4.3.3 Verification of Translations at National Centers

The results of statistical item analyses from the TIMSS 2003 field test, conducted during April through June of 2002, were reviewed by each country. Since unusual item statistics could be an indication of errors in translation, each NRC was asked to check the results to identify items that might have been mistranslated.

4.3.4 International Quality Control Monitor Item Review

As part of an ambitious quality control program, International Quality Control Monitors (QCMs) were hired to document the quality of the TIMSS 2003 assessment in each country (see Chapter 7 for a description of the work of the Quality Control Monitors). An important task for the QCMs was to review the translation verification reports for each test language and verify whether the suggested changes were made in the final instruments. The QCM marked on a copy of the translation verification report form whether the change suggested in the report was implemented. This assisted the International Study Center in identifying changes made or not made to the national versions.

4.4 Summary

The rigorous procedures for translation, cultural adaptations, translation verification, and review of the instruments implemented for TIMSS 2003 provided for comparable translations of the instruments across participating countries. The verification process of internal review, external translation verification by bilingual judges, and review by the International Study Center and Quality Control Monitors proved to be a comprehensive program for verification, ensuring accuracy in the analysis and reporting of the TIMSS 2003 data.

References

TIMSS (2002a), *TIMSS 2003 Survey Operations Manual*, prepared by the TIMSS & PIRLS International Study Center, Chestnut Hill, MA: Boston College.

TIMSS (2002b), *Guidelines for the Translation Verification of the TIMSS 2003 Main Survey Instruments*, prepared by the IEA Secretariat, Amsterdam: The Netherlands.

5

Chapter 5

TIMSS 2003 Sampling Design

Pierre Foy and Marc Joncas

5.1 Overview

This chapter describes the TIMSS 2003 international sample design and the procedures developed to ensure effective and efficient sampling of the student populations in each participating country. To be acceptable for TIMSS 2003, national sample designs had to result in probability samples that gave accurate weighted estimates of population parameters such as means and percentages, and for which estimates of sampling variance could be computed. The TIMSS 2003 sample design is similar to that used in TIMSS 1999, with minor refinements. Since sampling for TIMSS was to be implemented by the National Research Coordinator (NRC) in each participating country – often with limited resources – it was essential that the design be simple and easy to implement while yielding accurate and efficient samples of both schools and students. The design that was chosen for TIMSS strikes a good balance, providing accurate sample statistics while keeping the survey simple enough for all participants to implement.

The international project team provided software, manuals, and expert advice to help NRCs adapt the TIMSS sample design to their national system, and to guide them through the phases of sampling. The *School Sampling Manual* (TIMSS, 2001) describes how to implement the international sample design and to select the school sample; and offers advice on initial planning, adapting the design to national situations, establishing appropriate sample selection procedures, and conducting fieldwork. The *Survey Operations Manual* (TIMSS, 2002a) and *School Coordinator Manual* (TIMSS, 2002b) provide information on sampling within schools, assigning assessment booklets and questionnaires to sampled students, and tracking respondents and non-respondents. To automate the rather complex within-school sampling procedures, NRCs were provided with sampling software jointly developed by the IEA Data processing Center (DPC) and Statistics Canada, documented in the *Within School Sampling Software (WinW3S) Manual* (TIMSS, 2002c).

In addition to sampling manuals and software, expert support was made available to help NRCs with their sampling activities. Statistics Canada and the IEA Data Processing Center (in consultation with the TIMSS sampling referee) reviewed and approved the national sampling plans, sampling data, sampling frames, and sample implementation. Statistics Canada and the DPC also provided advice and support to NRCs at all stages of the sampling process, drawing national school samples for nearly all of the TIMSS participants.

Where the local situation required it, NRCs were permitted to adapt the sample design for their educational systems, using more sampling information, and more sophisticated designs and procedures, than the base design required. However, these solutions had to be approved by the TIMSS International Study Center (ISC) at Boston College, and by Statistics Canada.

5.2 TIMSS Target Populations

In IEA studies, the target population for all countries is known as the *international desired population*. TIMSS 2003 chose to study achievement in two target populations, and countries were free to participate in either population, or both. The international desired populations for TIMSS were the following:

- **Population 1:** All students enrolled in the upper of the two adjacent grades that contain the largest proportion of 9-year-olds at the time of testing. This grade level was intended to represent four years of schooling, counting from the first year of primary or elementary schooling, and was the fourth grade in most countries.
- **Population 2:** 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. This grade level was intended to represent eight years of schooling, counting from the first year of primary or elementary schooling, and was the eighth grade in most countries.

To measure trends in student achievement, the TIMSS 2003 eighth- and fourth-grade target populations were intended to correspond to the upper grades of the TIMSS 1995 population definitions, and the TIMSS 2003 eighth-grade target population to the eighth-grade population in TIMSS 1999.

5.2.1 Sampling from the Target Populations

TIMSS expected all participating countries to define their *national desired populations* to correspond as closely as possible to its definition of the international desired populations.

For example, if fourth grade was the upper of the two adjacent grades containing the greatest proportion of 9-year-olds in a particular country, then all fourth grade students in the country should constitute the national desired population for that country.

Although countries were expected to include all students in the target grade in their definition of the populations, sometimes they had to restrict their coverage. Lithuania, for example, collected data only about students in Lithuanian-speaking schools, so their national desired populations fell short of the international desired populations. Appendix A of the TIMSS 2003 international reports in mathematics and science documents such deviations from the international definition of the TIMSS target populations.

Using their national desired populations as a basis, each participating country had to define its populations in operational terms 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. Ideally, the national defined populations should coincide with the *national desired populations*, although in reality there may be some school types or regions that cannot be included. Consequently, the national defined populations are usually a very large subset of the national desired populations. All schools and students in the desired populations not included in the defined populations are referred to as the excluded populations.

TIMSS participants were expected to ensure that the national defined populations included at least 95 percent of the national desired populations. Exclusions (which had to be kept to a minimum) could occur at the school level, within the sampled schools, or both. Because the national desired populations were restricted to schools that contained the required grade, schools not containing the target grade were considered to be outside the scope of the sample, i.e., not part of the target populations.

Although countries were expected to do everything possible to maximize coverage of the populations by the sampling plan, if necessary, schools could be excluded from the sampling frame for the following reasons:

- They were in geographically remote regions.
- They were of extremely small size.

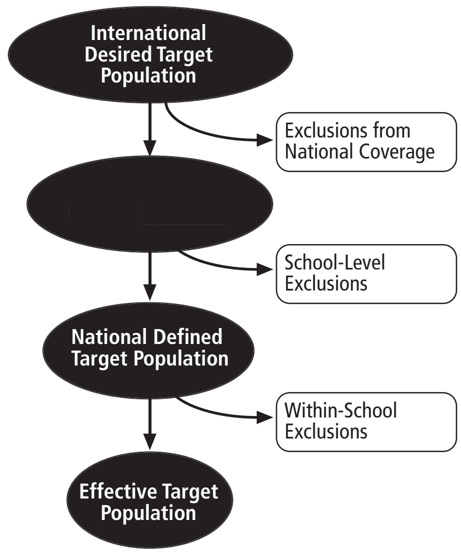
- They offered a curriculum or a school structure that was different from the mainstream education system(s).
- They provided instruction only to students in the categories defined as "within-school exclusions".

Within-school exclusions were limited to students who, because of some disability, were unable to take part in the TIMSS assessment. The general TIMSS rules for defining within-school exclusions included the following three groups:

- **Intellectually disabled students.** These are students who were considered, in the professional opinion of the school principal or other qualified staff members, to be intellectually disabled, or who had been so diagnosed in psychological tests. This category included students who were emotionally or mentally unable to follow even the general instructions of the TIMSS tests. 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 on 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.

Because these categories can vary internationally in the way they are implemented, NRCs were asked to adapt them to local usage. In addition, they were to estimate the size of the target population so that their compliance with the 95 percent rule could be projected. A major objective of TIMSS was that the effective target populations, the populations actually sampled by TIMSS, be as close as possible to the international desired populations. Exhibit 5.1 illustrates the relationship between the desired populations and the excluded populations. Each country had to account for any exclusion of eligible students from the international desired populations. This applied to school-level exclusions, as well as within-school exclusions.

Exhibit 5.1 **Relationship Between the Desired Populations and Exclusions**



5.3 Sample Design

The international sample design for TIMSS is generally referred to as a two-stage[1] stratified cluster sample design. The first stage consists of a sample of schools,[2] which may be stratified; the second stage consists of a sample of one or more classrooms from the target grade in sampled schools.

5.3.1 Units of Analysis and Sampling Units

The TIMSS analytical focus was on the cumulative learning of students, as well as on instructional characteristics related to learning. The sample design, therefore, had to address the measurement both of characteristics thought to influence cumulative learning, and of those specific to the instructional settings. As a consequence, although students were the principal units of analysis, schools and classrooms also were potential units of analysis, and all had to be considered as sampling units in the sample design in order to meet specific requirements for data quality and sampling precision at all levels.

Although the second stage sampling units were generally intact classrooms, the ultimate sampling elements were students – making it important that each student from a target grade be a member of one (and only one) of the classes in a school from which the sampled classes would be selected.

TIMSS prefers to sample intact classrooms because that allows the simplest link between students and teachers. In fourth grade, students in most countries are organized into classrooms that are taught as a unit for all

1 In some countries, it was necessary to include a third stage, where students within large classrooms were sub-sampled (see section 5.6).

2 In the Russian Federation, it was necessary to include an extra preliminary stage, where geographical regions were sampled first, and then schools (see section 5.4.3).

subjects, usually by the same teacher. Sampling intact classrooms is straightforward, therefore, at fourth grade. At eighth grade, however, classrooms are usually organized by subject – mathematics, language, science, etc. – and it is more difficult to arrange classroom sampling. TIMSS has addressed this issue by choosing the mathematics class as the sampling unit, mainly because classes often are organized on the basis of mathematics instruction and because mathematics is a central focus of the study. Although this is the recommended procedure, it can only be implemented where the mathematics classes in a school constitute an exhaustive and mutually exclusive partition of the students in the grade. This is the case when every student in the target grade attends one and only one mathematics class in the school.

5.3.2 Sampling Precision and Sample Size

In planning the sample design for each country, sample sizes for the two stages of the TIMSS sample design had to be specified so as to meet the sampling precision requirements of the study. Since students were the principal units of analysis, the reliability of estimates of student characteristics was paramount. However, TIMSS planned to report extensively on school, teacher, and classroom characteristics, so it was necessary also to have sufficiently large samples of schools and classes. The TIMSS standard for sampling precision requires that all student samples have an effective sample size of at least 400 students for the main criterion variables – mathematics and science achievement. In other words, all student samples should yield sampling errors that are no greater than would be obtained from a simple random sample of 400 students.

An effective sample size of 400 students results in the following approximate 95 percent confidence limits for sample estimates of population means, percentages, and correlation coefficients.

- Means: m ± 0.1s (where *m* is the mean estimate, and *s* is the estimated standard deviation for students)
- Percentages: p ± 5% (where *p* is a percentage estimate)
- Correlations: r ± 0.1 (where *r* is a correlation estimate)

Notwithstanding these precision requirements, TIMSS required a minimum of 4,000 students for each target population. This was necessary to ensure adequate sample sizes for sub-groups of students categorized by school, class, teacher, or student characteristics. Furthermore, since TIMSS planned to conduct analyses at the school and classroom levels, at least 150 schools were to be selected from each target population. Samples of 150 schools yield 95 percent confidence limits for school-level and classroom-level mean estimates that are precise to within 16 percent of their standard deviations. Therefore, to ensure sufficient sample precision for school-level and student-level analyses, some participants had to sample more schools and students than would have been selected otherwise.

5.3.3 Clustering Effect

The precision of multistage cluster sample designs is generally affected by the so-called clustering effect. Students are clustered in schools, and are also clustered in classrooms within the schools. 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 intra-class correlation is a measure of this within-class similarity. Sampling 30 students from a single classroom when the intra-class correlation is high will yield less information than a random sample of 30 students drawn from across all students in the grade level. Consequently, a cluster sample with a positive intra-class correlation will need to have more elements than a random sample of independent elements to achieve the same level of precision. Thus, cluster sample designs are less efficient, in terms of sampling precision, than a simple random sample of the same size. This clustering effect was considered in determining the overall sample sizes for TIMSS.

The size of the cluster (classroom) and the size of the intra-class correlation determine the magnitude of the clustering effect. For planning its sample size, therefore, each country had to identify a value for the intra-class correlation and a value for the expected cluster size (this was known as the minimum cluster size). The intra-class correlation for each country was estimated from previous cycles of TIMSS, from IEA's Progress in International Reading Literacy Study (PIRLS), or from national assessments. In the absence of these sources, an intra-class correlation of 0.3 was assumed. Since participants were generally sampling intact classrooms, the minimum cluster size was in fact the average classroom size.

Sample-design tables, such as the one in Exhibit 5.2, were produced and included in the TIMSS *School Sampling Manual*. These tables illustrate the number of schools necessary to meet the TIMSS sampling precision requirements for a range of values of intra-class correlations and minimum cluster sizes. TIMSS participants could refer to the tables to determine how many schools they should sample. For example, on the basis of Exhibit 5.2, a participant whose intra-class correlation was expected to be 0.6, with an average classroom size of 30, would need to sample a minimum of 262 schools. Whenever the estimated number of schools to sample was less than 150, participants were asked to sample at least 150 schools. Also, if the total expected number of students was less than 4,000, participating countries were asked to select more schools, or more classrooms per school. The sample design tables could also be used to determine sample sizes for more complex designs. For example, geographical regions could be defined as strata, whereby equal numbers of schools would be sampled in each stratum in order to produce equally reliable estimates for all strata, regardless of the relative size of the strata.

Exhibit 5.2 **TIMSS Sample Design Table**

		Intraclass Correlation								
MCS		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
5	a	212	244	276	308	340	372	404	436	468
	n	1,060	1,220	1,380	1,540	1,700	1,860	2,020	2,180	2,340
10	a	150	162	198	234	270	306	342	378	414
	n	1,500	1,620	1,980	2,340	2,700	3,060	3,420	3,780	4,140
15	a	150	150	172	209	247	284	321	359	396
	n	2,250	2,250	2,580	3,135	3,705	4,260	4,815	5,385	5,940
20	a	150	150	159	197	235	273	311	349	387
	n	3,000	3,000	3,180	3,940	4,700	5,460	6,220	6,980	7,740
25	a	150	150	151	190	228	266	305	343	382
	n	3,750	3,750	3,775	4,750	5,700	6,650	7,625	8,575	9,550
30	a	150	150	150	185	223	262	301	339	378
	n	4,500	4,500	4,500	5,550	6,690	7,860	9,030	10,170	11,340
35	a	150	150	150	181	220	259	298	337	375
	n	5,250	5,250	5,250	6,335	7,700	9,065	10,430	11,795	13,125
40	a	150	150	150	179	218	257	296	335	374
	n	6,000	6,000	6,000	7,160	8,720	10,280	11,840	13,400	14,960
45	a	150	150	150	176	216	255	294	333	372
	n	6,750	6,750	6,750	7,920	9,720	11,475	13,230	14,985	16,740
50	a	150	150	150	175	214	253	292	332	371
	n	7,500	7,500	7,500	8,750	10,700	12,650	14,600	16,600	18,550
55	a	150	150	150	173	213	252	291	331	370
	n	8,250	8,250	8,250	9,515	11,715	13,860	16,005	18,205	20,350
60	a	150	150	150	172	212	251	290	330	369
	n	9,000	9,000	9,000	10,320	12,720	15,060	17,400	19,800	22,140

a = Number of sampled schools

n = Number of sampled students in the target grade

Note: The Minimum Cluster Size (MCS) is the number of students selected in each sampled school (generally the average classroom size).

5.3.4 Stratification

Stratification is the grouping of sampling units (e.g., schools) in the sampling frame according to some attribute or variable prior to drawing the sample. It is generally used for the following reasons:

- To improve the efficiency of the sample design, thereby making survey estimates more reliable.
- To apply different sample designs or disproportionate sample-size allocations to specific groups of schools (such as those within certain states or provinces).
- To ensure adequate representation in the sample of specific groups from the target population.

Examples of stratification variables for school samples are: geography (such as states or provinces), school type (such as public and private), and level of urbanization (such as rural and urban). Stratification variables in the TIMSS sample design could be used explicitly, implicitly, or both.

- **Explicit stratification** consists of building separate school lists, or sampling frames, according to the stratification variables under consideration. For example, where geographic regions are an explicit stratification variable, separate school sampling frames would be constructed for each region. Different sample designs, or different sampling fractions, would then be applied to each school sampling frame to select the sample of schools. In TIMSS, the main reason for considering explicit stratification was to ensure disproportionate allocation of the school sample across strata. For example, a country stratifying by school type might require a specific number of schools from each stratum, regardless of the relative sizes of the strata.
- **Implicit stratification** makes use of a single school sampling frame, but sorts the schools in this frame by a set of stratification variables. This type of stratification, combined with the PPS systematic sampling methodology (see section 5.4), is a simple way of ensuring proportional sample allocation without the complexity of explicit stratification. It can also improve the reliability of survey estimates – provided the stratification variables are related to school mean student achievement in either mathematics or science.

5.3.5 Replacement Schools

Although TIMSS participants were expected to make great efforts to secure the participation of sampled schools, it was anticipated that a 100 percent participation rate would not be possible in all countries. To avoid sample size losses, a mechanism was instituted to identify, a priori, replacement schools for each sampled school. For each sampled school, the next school on the ordered school sampling frame was identified as its replacement, and the one after that as a second replacement, should it be needed (see Exhibit 5.3 for an example).

The use of implicit stratification variables and the subsequent ordering of the school sampling frame by size ensured that any sampled school's replacement would have similar characteristics. Although this approach avoids sample size losses, it does not guarantee avoiding response bias. However, it may reduce the potential for bias, and was deemed more acceptable than over-sampling to accommodate a low response rate.

5.4 First Sampling Stage

The sample selection method used for the first sampling stage in TIMSS makes use of a systematic probability-proportional-to-size (PPS) technique. In order to use this method, it is necessary to have some measure of the size (MOS) of the sampling units. Ideally, this should be the number of sampling elements within the unit (e.g., the number of students in the school in the target grade). If this is unavailable, some other highly correlated measure, such as total school enrollment, may be used.

The schools in each explicit stratum are listed in order of the implicit stratification variables, together with the MOS for each school. Schools are further sorted by MOS within the implicit stratification variables. The measures of sizes are accumulated from school to school, and the running total (the cumulative MOS) is listed next to each school (see Exhibit 5.3). The cumulative MOS is an index of the size of the population of sampling elements; dividing it by the number of schools to be sampled gives the sampling interval.

The first school is sampled by choosing a random number in the range between 0 and the sampling interval. The school whose cumulative MOS contains the random number is the sampled school. By adding the sampling interval to that first random number, the second school is identified. This process of consistently adding the sampling interval to the previous selection number results in a PPS sample of schools of the required size.

Among the many benefits of this sample selection method are that it is easy to implement, and that it is easy to verify that it was implemented properly. The latter is critical, since one of the main methodological objectives of TIMSS was to ensure that a sound sampling technique had been used. Exhibit 5.3 illustrates the PPS systematic sampling method applied to a fictitious sampling frame. The first three sampled schools are shown, as well as their pre-selected replacement schools, which may be used should the originally selected schools not participate.

Exhibit 5.3 **Application of the PPS Systematic Sampling Method to TIMSS**

	Total MOS: 392 154 School Sample: 150		Sampling Interval: 2 614.3600 Random Start: 1 135.1551
School Code	**School MOS**	**Cumulative MOS**	**Sample**
939438	532	532	
026825	517	1049	
277618	487	1536	–
228882	461	1997	R1
833389	459	2456	R2
386017	437	2893	
986694	406	3299	
041733	385	3684	
056595	350	4034	–
945801	341	4375	R1
865982	328	4703	R2
700089	311	5014	
656616	299	5313	
647690	275	5588	
381836	266	5854	
510529	247	6101	
729813	215	6316	
294281	195	6511	–
016174	174	6685	R1
292526	152	6837	R2
541397	133	6970	
502014	121	7091	
662598	107	7198	
821732	103	7301	
436600	97	7398	

– = Sampled School R1, R2 = Replacement Schools

5.4.1 Small Schools

Small schools, those with fewer eligible students than are typically found in a classroom, can cause difficulties in PPS sampling because students sampled from them tend to be assigned very large sampling weights, which can increase sampling variance. Also, because such schools supply fewer students than the other schools, the overall student sample size may be reduced. In TIMSS, a school was deemed to be small if the number of students in the target grade was less than the minimum cluster size. For example, if the minimum cluster size was set at 20, then a school with fewer than 20 students in the target grade was considered a small school.

The TIMSS approach for dealing with small schools had two components:

- **Exclude extremely small schools**. Extremely small schools were defined as schools with fewer students than one quarter of the minimum cluster size. For example, if the minimum cluster size was set at 20, schools with fewer than five students in the target grade were considered extremely small schools. If student enrollment in these schools was less than two percent of the eligible population, these schools could be excluded, provided the overall inclusion rate met the 95 percent criterion (see section 5.2.1).
- **Select remaining small schools with equal probabilities**. All remaining small schools were selected with equal probabilities within explicit strata. This was done by calculating, for each explicit stratum, the average size of small schools and setting the MOS of all small schools to this average size. The number of small schools to be sampled within explicit strata would thus remain proportional, and this action would ensure greater stability in the resulting sampling weights.

5.4.2 Very Large Schools

A very large school is a school whose measure of size is larger than the calculated sampling interval. Very large schools can cause operational problems because they stand a chance of being selected more than once under the normal PPS sampling method. This problem was solved in one of two ways:

- **Creating an explicit stratum of very large schools**. All very large schools were put in an explicit stratum and all of them were included in the sample. This was done within the originally defined explicit strata since the sampling intervals were calculated independently for each original explicit stratum. Thus, an explicit stratum would be divided into two parts if it contained any very large schools.
- **Setting their MOS equal to the sampling interval**. All very large schools in an explicit stratum were given a measure of size equal to the sampling interval calculated for that explicit stratum. In this way, very large schools were all included in the sample with probabilities of unity. This approach was simpler to apply and avoided the formation of additional explicit strata.

5.4.3 Optional Preliminary Sampling Stage

In TIMSS, very large countries have the opportunity to introduce a preliminary sampling stage before sampling schools. This consists of first drawing a sample of geographic regions using PPS sampling and then a sample of schools from each sampled region. This design is used mostly as a cost reduction measure, where the construction of a comprehensive list of schools is either impossible or prohibitively expensive. Also, the additional sampling stage reduces the dispersion of the school sample, thereby potentially reducing travel costs. Sampling guidelines ensure that an adequate number of units are sampled from this preliminary stage. The sampling frame has to consist of at least 80 primary sampling units, of which at least 40 must be sampled at this stage. The Russian Federation was the only country to avail of this option in TIMSS 2003.

5.5 Second Sampling Stage

The second sampling stage in the TIMSS international design consisted of selecting classrooms within sampled schools. As a rule, one classroom per school was sampled, although some participants opted to sample two classrooms. Additionally, some participants were required to sample two or more classrooms per school in order to meet the minimum requirement of 4,000 sampled students. Classrooms were generally selected with equal probabilities. For those countries that chose to sub-sample students within classrooms (see section 5.6), classroom sampling was done using PPS sampling within the affected schools.

5.5.1 Small Classrooms

Generally, classrooms in an education system tend to be of roughly equal size. Occasionally, however, small classrooms are devoted to special situations, such as remedial or accelerated programs. These classrooms can become problematic in sampling, since they can lead to a shortfall in sample size, and also introduce some instability in the resulting sampling weights.

In order to avoid these problems, any classroom smaller than half the specified minimum cluster size was combined with another classroom from the same grade and school. For example, if the minimum cluster size was set at 30, any classroom with fewer than 15 students was combined with another. The resulting pseudo-classroom then constituted a sampling unit.

5.6 Sampling Students Within Classes

As a rule, all students in the sampled classrooms were expected to take part in the TIMSS assessment. However, countries where especially large classes were the norm could with permission opt to sub-sample a fixed number of students from each sampled classroom. Where applicable, this was done using a systematic sampling method whereby all students in a sampled classroom were assigned equal selection probabilities. In TIMSS 2003, only Yemen chose this option.

References

TIMSS (2001), *TIMSS 2003 School Sampling Manual*, prepared by P. Foy & M. Joncas, Statistics Canada, Chestnut Hill, MA: Boston College.

TIMSS (2002a), *TIMSS 2003 Survey Operations Manual,* prepared by the International Study Center, Chestnut Hill, MA: Boston College.

TIMSS (2002b), TIMSS 2003 School Coordinator Manual, prepared by the International Study Center, Chestnut Hill, MA: Boston College.

TIMSS (2002c), *TIMSS 2003 Within School Sampling Software (WinW3S) Manual,* prepared by O. Neuschmidt, J. Pickel et al., IEA Data Processing Center, Chestnut Hill, MA: Boston College.

6

Chapter 6

TIMSS 2003 Survey Operations Procedures

Juliane Barth, Eugenio J. Gonzalez, and Oliver Neuschmidt

6.1 Overview

The TIMSS 2003 data collection in each country was a very demanding exercise, with test administration at two grade levels in at least 150 schools, and with questionnaires for students, mathematics and science teachers, and school principals. To conduct the data collection successfully called for close cooperation between the National Research Coordinator (NRC) and school personnel – principals and teachers – and students. The first part of this chapter describes the field operations for collecting the data, including the responsibilities of the NRC, the procedure for sampling classrooms within schools and tracking students and teachers, and the steps involved in administering the achievement tests and background questionnaires. The second part describes the activities involved in preparing the data files at national centers, particularly the procedures for scoring the constructed-response items, creating and checking data files for achievement test and questionnaire responses, and dispatching the completed data files to the IEA Data Processing Center (DPC) in Hamburg, Germany.

6.2 TIMSS 2003 Field Operations

The TIMSS 2003 field operations were developed jointly by the TIMSS & PIRLS International Study Center at Boston College, the IEA Data Processing Center, and Statistics Canada. They were based on procedures used successfully in TIMSS 1995, TIMSS 1999, and other IEA studies, and were refined on the basis of TIMSS 2003 field-test experience.

6.2.1 Responsibilities of the National Research Coordinator

In conducting field operations in each country, the National Research Coordinator was the key person. The NRC had ultimate responsibility for collecting the data for the TIMSS assessment according to internationally agreed-upon procedures and preparing the data according to international specifications. NRC responsibilities in other areas, including sampling schools and translating the achievement tests and questionnaires, have been outlined in earlier chapters of this report.[1] This section focuses on NRC activities with regard to administering the assessment in participating schools. Specifically, it describes the procedures for sampling classes within schools, for tracking classes, teachers, and students in the sampled schools, and for organizing the administration of the achievement tests and questionnaires.

6.2.2 Documentation and Software

NRCs were provided with a comprehensive set of procedural manuals detailing all aspects of the data collection.

- The *TIMSS 2003 Survey Operations Manual* (TIMSS, 2002a) was the essential handbook of the National Research Coordinator, and described in detail all of the activities and responsibilities of the NRC, from the moment the TIMSS instruments arrived at the national center to the moment the checked and verified data files and accompanying documentation were submitted to the IEA Data Processing Center.
- The *TIMSS 2003 School Sampling Manual* (TIMSS, 2001) defined the TIMSS 2003 target populations and sampling goals and described the procedures for the sampling of schools.
- The *TIMSS 2003 School Coordinator Manual* (TIMSS, 2002b) described the activities of the School Coordinator – the person in the school responsible for organizing the TIMSS test administration – from the time the testing materials arrived at the school to the time the completed materials were returned to the national TIMSS center.
- The *TIMSS 2003 Test Administrator Manual* (TIMSS, 2002c) described in detail the procedures for administering the TIMSS tests and questionnaires, from the beginning of the test administration to the return of the testing materials to the School Coordinator.
- The *TIMSS 2003 Scoring Guides for Mathematics and Science Constructed-Response Items* (TIMSS, 2002d; TIMSS, 2002e) contained instructions for scoring the short-answer and extended-response test items.
- The *Manual for Entering the TIMSS 2003 Data* (TIMSS, 2002f) provided the NRCs with instructions for coding, entering, and verifying the data.

1 See Chapter 5 for information about sampling schools, and Chapter 4 for details of the translation task.

- The *TIMSS 2003 National Quality Control Observer's Manual* (TIMSS, 2002g) provided instructions for conducting classroom observations during data collection in a sample of participating schools.

Additionally, six software packages were supplied by the IEA Data Processing Center to assist NRCs with the data collection:

- The within-school sampling software (WinW3S) is a computer program that helps NRCs randomly sample the TIMSS class or classes in each sampled school; prepare the survey tracking forms that keep track of sampled students, classes, and teachers; and assign test booklets to students. The software stores all tracking information in an MS-Access database so that it can be used later in constructing sampling weights and in verifying the integrity of the sampling procedure.
- The DataEntryManager for Windows (WinDEM), is a computer program developed by IEA to enable national center staff to capture all of the TIMSS data through keyboard data entry and to perform a range of validity checks on the keyed data. The WinDEM database includes codebooks for each of the TIMSS 2003 test booklets and questionnaires, providing all information necessary to produce data files for each instrument in a standard international format.
- The WinLink program allows NRCs to check the correspondence between the tracking information stored in the WinW3S database and the student, teacher, and school information keyed into the WinDEM files. Using this program, for example, NRCs can check that each student listed on the student tracking form has a corresponding data record in the student achievement and student questionnaire WinDEM files.
- The Data Correction Software (DCS) is a program that enables national center staff to detect and correct inconsistencies in TIMSS background data files.
- The Trend-Scoring Reliability Software (TSRS) incorporates a database for each country containing a sample of student responses to constructed-response questions administered and scored as part of the TIMSS 1999 data collection. The TSRS software allowed NRCs to have their 2003 scoring staff rescore the 1999 student sample to document the reliability of the scoring process over time. This effort is described in Chapter 8.
- In a related effort, the Cross-Country Scoring Reliability Software (CCSRS) incorporates a database containing a sample of student responses to constructed-response items collected from English-speaking countries participating in TIMSS 2003. The CCSRS software enables every country with English-speaking scoring staff to score these common student responses in order to document the reliability of the scoring across countries participating in 2003. For more information, please refer to Chapter 8.

Each software package was supplied with a detailed manual describing how to install and use the software. In addition to the manuals, NRCs received hands-on training in the use of the WinW3S and WinDEM software from staff at the IEA Data Processing Center and Statistics Canada during a data entry seminar held before the field test.

6.2.3 Within-School Sampling Procedures

The study design anticipated relational analyses between student achievement and teacher-level data at the class level. For field operations, this meant that intact classes had to be sampled, and that for each sampled class the mathematics and science teachers had to be tracked and linked to their students. Although intact classes were the unit to be sampled in each school, the ultimate goal was a nationally representative sample of students. Consequently, in each country a classroom organization had to be chosen that ensured that every student in the school was in one class or another, and that no student was in more than one class. Such an organization is necessary for a random sample of classes to result in a representative sample of students. At the eighth grade in most countries, mathematics classes serve this purpose well, and so were chosen as the sampling units. In countries where students attended different classes for mathematics and science, classrooms were defined on the basis of mathematics instruction for sampling purposes.[2] At fourth grade, most schools use the same class for all subjects, including mathematics and science. Accordingly, the fourth-grade classroom was the sampling unit in these schools.

The TIMSS design required that for each student in each sampled class, all teachers teaching mathematics or science be identified and asked to complete a teacher questionnaire.

Although all students enrolled in the target grade were part of the target population and were eligible to be selected for testing, TIMSS recognized that some students in every school would be unable to take part in the 2003 assessment because of some physical or mental disability. Accordingly, the sampling procedures provide for the exclusion of students with any of several disabilities (see Chapter 5). 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 TIMSS 2003 international reports. It was important that the conditions under which students could be excluded be carefully delineated, because the definition of “disabled” students varied considerably from country to country.

2 For countries where a suitable configuration of classes for sampling purposes could not be identified, TIMSS also provided a procedure for sampling individual students directly from the eighth grade.

Exhibit 6.1 presents the major activities conducted by National Research Coordinators and school personnel while sampling classes within schools. These activities are incorporated in the WinW3S software, which automatically produces all necessary forms, lists, and labels, and assisted NRCs in keeping track of the field operations' status.

Exhibit 6.1 **Procedures for Sampling Classes in Partcipating Schools**

NRC activity	School activity
1. School Tracking • Contact schools participating schools • Prepare Class Listing Forms to be completed by schools.	
	2. **Complete the Class Listing Form listing all mathematics classes in the target grade (4 or 8) along with the names of their mathematics teachers.**
3. Class Tracking and Sampling • Sample a class or classes using the information on the Class Listing Form. • Prepare Student-Teacher Linkage Forms so that schools can list the students in the sampled class(es) and link them to their mathematics and science teachers.	
	4. **Complete Student-Teacher Linkage Forms by listing all of the students in the sampled class(es) (name, birth dates, sex) together with their mathematics and science teachers and course names.**
5. Student/ Teacher Tracking and Student-Teacher Linkage • Prepare a Student Tracking Form for each sampled class listing all students to be tested and their booklet assignments • Prepare a Teacher Tracking Form for each sampled class listing all mathemathematics and science teachers of the students in the class, their questionnaire assignments and their student-teacher link numbers • Send tracking forms, labels and test-instruments to schools.	
TEST ADMINISTRATION	
	6. **After the tests and questionnaires have been administered, record the participation status on Student and Teacher Tracking Forms; complete Test Administrator Forms.**
7. Record Participation Information and Test Administrator Information in Data Files.	

6.2.3.1 Survey Tracking Forms

As may be seen from Exhibit 6.1, TIMSS 2003 relied on a series of "tracking forms" to implement and record the sampling of classes, teachers, and students. It was essential that the tracking forms be completed accurately, since they determine which booklets and questionnaires should be given to which students and teachers, and record what happened as the assessment was administered in each school. In addition to facilitating the data collection, the tracking forms provided essential information for the computation of sampling weights and for evaluating the quality of the sampling procedures. All tracking forms were retained for review by staff of the TIMSS International Study Center and the IEA Data Processing Center.

Survey tracking forms were provided for sampling classes and students; for tracking schools, classes, teachers, and students; for linking students and teachers; and for recording information during test administration.

6.2.3.2 Linking Students, Teachers, and Classes

The Within-School Sampling Software (WinW3S) creates a hierarchical identification numbering system that uniquely identifies the sampled schools, teachers, classes, and students within each country. At the root of the system is a four-digit school identification number unique within each country that is assigned to each sampled school.

A class identification number is assigned to each class in the target grades listed on the class tracking form or entered in WinW3S. The six-digit class identification number consists of the four-digit school number followed by a two-digit number identifying the class within the school.

Each student listed on the student tracking form is assigned a student identification number. This eight-digit number consists of the six-digit class number followed by a two-digit number corresponding to the student's sequential position on the student tracking form. All students listed on the student tracking form, including those marked for exclusion, are assigned a student identification number.

Each mathematics and science teacher of the selected classes (i.e., those listed on the teacher tracking form) is assigned a teacher identification number consisting of the four-digit school number followed by a two-digit teacher number unique within the school. Since a teacher could be teaching both mathematics and science to some or all of the students in a class, it is necessary to have a unique identification number for each teacher/class and teacher/subject combination. This is achieved by adding a two-digit link number to the six digits of the teacher identification number, giving a unique eight-digit teacher/class identification number. Careful implementation of these procedures is necessary so that during data analysis each class may be linked to a teacher, and student outcomes may be analyzed in relation to teacher-level variables.

6.2.4 Assigning Testing Materials to Students and Teachers

At both eighth and fourth grades, the mathematics and science assessment questions were packaged into 12 student test booklets. Each sampled student was required to complete one booklet, as well as the student questionnaire. Booklets were assigned to students by the WinW3S software using a random assignment procedure.

Each teacher listed on the teacher tracking form was assigned a teacher questionnaire. At eighth grade there were separate questionnaires for mathematics and science teachers. Where teachers taught both mathematics and science to the class, every effort was made to collect information about both subjects. However, NRCs had the final decision as to how much response burden to place on such teachers. Where a teacher taught both subjects to a class but completed only one questionnaire, the information from the general background part of the completed questionnaire was copied into the missing questionnaire.

6.2.5 Administering the Test Booklets and Questionnaires

The School Coordinator was the person in the school responsible for administrating the TIMSS 2003 assessment. The coordinator could be the principal, the principal's designee, or an outsider appointed by the NRC with the approval of the principal. The NRC was responsible for ensuring that the School Coordinators were familiar with their responsibilities.

The major responsibilities of the School Coordinators are detailed in the *TIMSS 2003 School Coordinator Manual* (TIMSS, 2002b). Prior to the test administration the tasks for the School Coordinator included:

- providing the NRC with all information necessary to complete the various tracking forms;
- checking the assessment materials when they arrived in the school to ensure that everything was in order;
- ensuring that the assessment materials were kept in a secure place before and after the administration;
- arranging the dates of the assessment administration with the national center;

- arranging for a Test Administrator and giving a briefing on the TIMSS 2003 study, the assessment materials, and the assessment sessions; and
- working with the school principal, the Test Administrator, and the teachers to plan the testing day – this involved arranging rooms, times, classes and materials.

The Test Administrator was responsible for administering the TIMSS tests and student questionnaires. Specific responsibilities were described in the *TIMSS 2003 Test Administrator Manual* (TIMSS, 2002c), and included:

- ensuring that each student received the correct testing materials which were specially prepared for him or her;
- administering the test in accordance with the instructions in the manual;
- ensuring the correct timing of the testing sessions by using a stopwatch and recording the time when the various sessions started and ended on the Test Administration Form; and
- recording student participation on the Student Tracking Form.

The responsibilities of the School Coordinator after the test administration included:

- ensuring that the Test Administrator returned all assessment materials, including the completed Student Tracking Form, the Test Administration Form, and any unused booklets;
- calculating the student response rate and arranging for makeup sessions if it was below 90 percent;
- distributing the teacher questionnaires to the teachers listed on the Teacher Tracking Form, ensuring that the questionnaires were returned completed, and recording teacher participation information on the Teacher Tracking Form;
- preparing a report for the NRC about the test administration in the school; and
- returning both completed and unused test materials and all tracking forms to the NRC.

The NRC prepared two packages for each sampled class. One contained the test booklets for all students listed on the Student Tracking Form and the other the student questionnaires. For each participating school, the test booklets and student questionnaires were bundled together with the Teacher Tracking Form and teacher questionnaires, the school questionnaire, and the materials prepared for briefing School Coordinators and Test Administrators, and were sent to the School Coordinator. A set of labels and prepaid envelopes addressed to the NRC was included to facilitate the return of testing materials.

6.2.6 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 at each grade assessed 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 2003 schools, observed actual testing sessions, and recorded compliance of the test administration with prescribed procedures.

To assist NRCs in conducting their national quality control program, the TIMSS International Study Center prepared the *TIMSS 2003 National Quality Control Observer's Manual* (TIMSS, 2002g) which provided general information about TIMSS 2003 and detailed the role and responsibilities of the National Quality Control Observers.

6.3 Data Preparation

In the period immediately following the administration of the TIMSS 2003 assessment, the major tasks for the NRC included retrieving and collating the materials from the schools; recruiting and training scorers to score the constructed-response items; scoring these items, including double scoring a reliability sample of 1200 booklets; entering the data from the achievement tests and background questionnaires into computer files; checking and editing the data with the software provided by the IEA Data Processing Center; submitting the data files and materials to the IEA Data Processing Center; and preparing a report on survey activities.

When the testing materials were received back from the schools, NRCs had the following tasks:

- check that the appropriate testing materials were received for every student listed on the Student Tracking Form;
- verify all identification numbers on all instruments;
- check that the participation status recorded on the tracking forms matched the information on the test booklets and questionnaires; and
- follow up on schools that did not return the testing materials or for which forms were missing, incomplete, or inconsistent.

NRCs then organized the tests for scoring and data entry. The procedures involved 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 constructed-response coding, and document the reliability of the coding.

6.3.1 Scoring the TIMSS 2003 Constructed-Response Items

Reliable application of the scoring guides to the constructed-response questions, and empirical documentation of the reliability of the scoring process, were critical to the success of TIMSS 2003. The *TIMSS 2003 Survey Operations Manual* (TIMSS, 2002a) provided suggestions about arranging for staff and facilities for the constructed-response scoring effort required for the TIMSS 2003 main data collection; for effective training of the scorers; and for incorporating reliability scoring into the scheme for distributing booklets to scorers and monitoring the scoring. Countries were to double score 1200 booklets to document scoring reliability.

For all countries, the scope of the constructed-response scoring effort was substantial. The assessment contained 130 constructed-response questions at fourth grade and 146 constructed-response questions at eighth grade. These were distributed across 12 student booklets at each grade level.

6.3.1.1 Preparing to Train the Scorers

To ascertain the staff requirements for constructed-response scoring, it was necessary to estimate the amount of scoring to be done and the amount of time available to do it, and also to make provision for staff training and for clerical and quality control throughout the operation. The TIMSS International Study Center recommended at least one half-day of training on each of the 12 booklets, for a total of about a week for training activities.

In scoring the constructed-response items, it was vital that scoring staff apply the scoring rules consistently and in the same way in all participating countries. Hence, in selecting those who were to do the scoring, NRCs took care to arrange for persons who were conscientious and attentive to detail, knowledgeable in mathematics and science, and willing to apply the scoring guides as stated, even if they disagreed with a particular definition or category. Preference was given to individuals with educational backgrounds in the mathematics and science curriculum areas or who had taught at the middle school or primary level. Good candidates for scoring included teachers, retired teachers, college or graduate students, and staff of education agencies or ministries and research centers.

The success of assessments that, like TIMSS, include a large proportion of constructed-response questions is crucially dependent upon reliable scoring of student responses. In TIMSS 2003, scoring reliability was assured through the provision of detailed scoring guides (manuals), extensive training in their use, and continuous monitoring of the quality of the work. To support training in scoring, TIMSS 2003 provided training packets for training in selected questions, and practice papers to help scorers achieve a consistent level of scoring.

At the international scoring training meetings, NRCs received training packets containing example responses and practice papers to help them achieve accuracy and consistency in scoring. For scoring guides that were difficult, example responses were selected to illustrate the scoring categories. The scores on these responses were explained and attached to the scoring guides. Practice sets were created for the more difficult guides. These papers illustrated a range of responses, beginning with several clear-cut examples. About 10 to 15 responses were enough for most guides, but sometimes more practice was necessary.

Each scorer received a copy of the *TIMSS 2003 Main Survey Scoring Guides for Mathematics and Science Constructed-Response Items* (TIMSS, 2002d; TIMSS, 2002e). These manuals explain the TIMSS scoring system, which was designed to produce a rich and varied profile of the range of students' competencies in mathematics and science, and provide detailed scoring guides and example student responses for each constructed-response question in the assessment.[3]

6.3.1.2 Conducting the Constructed-Response Scoring

TIMSS 2003 recommended that scorers be organized into teams of about six, headed by a team leader. The leader's primary responsibility was to monitor scoring reliability by continually checking and rechecking the scores that scorers had assigned. This process, known as back-reading, was essential for identifying scorers who did not understand particular guides or categories. Early detection of any misunderstandings permitted clarification and rectification of mistakes before too many responses had been scored. The back-reading systematically covered the daily work of each scorer. If a particular scorer appeared to have difficulty, however, then the percentage of back-reading for that scorer was increased. Any errors discovered were brought to the attention of the scorer responsible and corrected immediately. If a scorer was found to have been consistently making an error, then all of the booklets scored by that person were checked and any errors corrected.

In order to demonstrate the quality of the TIMSS 2003 data, it was important to document the reliability of the scoring process – within countries, over time across assessments, and across countries.

6.3.1.3 Monitoring Scoring Reliability Within Each Country

To establish the reliability of the scoring within each country, NRCs were required to have a random sample of at least 100 booklets of each of the 12

3 See Chapter 2 for a description of the TIMSS constructed-response scoring system.

student test booklets scored independently by two different scorers. The reliability sample of booklets was selected randomly by the WinW3S software. The degree of agreement between the scores assigned by the two scorers is a measure of the reliability of the scoring process. Since the purpose of the double scoring was to document the consistency of the scoring procedure in each country, 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 procedure recommended by the TIMSS International Study Center was designed to blend the scoring of the reliability sample with the normal scoring activity, to take place at the same time, and to be systematically implemented across student responses and scorers.

In scoring the booklets for the main data set, scorers entered their scores directly into the student booklets. Therefore, in order that the reliability scoring be done "blind" (i.e., so that the two scorers did not know each other's scores), the reliability scoring had to be done before the scoring for the main data, and the reliability scores had to be recorded on a separate scoring sheet, and not in the booklets.

To implement the scoring plan effectively it was necessary that the scorers be divided between two equivalent teams (Team A and Team B), and that booklets be divided into two equivalent sets (Set A and Set B). The scorers in Team A scored around 600 of the booklets in Set B and all the booklets in Set A, while the scorers in Team B scored around 600 of the booklets in Set A and all of the booklets in Set B. Each team, therefore, handled both sets of booklets. For the set it handled first, the team did the reliablity scoring first and recorded the results on a separate answer sheet (this was the reliability sample). In the other set, the team scored all booklets and wrote the scores directly into the booklets.

Periodically during the day, the Team B scorers scored the reliability sample in the Set A batches, while the Team A scorers scored the reliability sample in the Set B batches. It was important that the reliability sample was scored as randomly drawn by the WinW3S software, and not just the top quarter in the set. When the reliability scoring was finished, Team B scorers marked it as completed and forwarded the batch to the Team A scorers. Similarly, the Team A scorers forwarded their scored reliability booklets from Set B to the Team B scorers. Once the booklets from Set A had been distributed to Team A scorers and the Set B booklets to the Team B scorers, all the constructed-response items were scored, and the scores were entered directly into the booklets.

6.3.1.4 Monitoring Scoring Reliability over Time (1999 to 2003)

The double scoring of a sample of the student test booklets provided a measure of the consistency within each country with which constructed-response questions were scored. To measure trends since 1999 and 1995, TIMSS 2003 included items from both of these assessments. TIMSS 2003 took steps to show that those constructed-response items used in 2003 that also had been used in 1999 were scored in the same way in both assessments. To make this possible, countries that participated in TIMSS 1999 sent samples of scored student booklets from the 1999 data collection to the IEA Data Processing Center, where they were digitally scanned and stored for later use. So that the student responses from 1999 could be rescored by 2003 scoring staff as a reliability check, the DPC developed software known as the Trend Scoring Reliability Software (TSRS) that presented the 1999 student responses without their 1999 scores. This enabled 2003 scoring staff to score these 1999 responses without seeing the scores awarded in 1999 and so provide a check on scoring consistency from 1999 to 2003. Those items from 1995 that were used in TIMSS 2003 all were in multiple-choice format, and therefore scoring reliability was not an issue.

6.3.1.5 Monitoring Scoring Reliability Across Countries

Because of the many different languages in use in TIMSS, establishing the reliability of constructed-response scoring across all countries was not feasible. However, TIMSS 2003 did conduct a cross-country study of scoring reliability among northern-hemisphere countries whose scorers were proficient in English. A sample of student responses to a subset of the mathematics and science constructed-response items was provided by the English-speaking southern hemisphere countries. These student responses were digitally scanned and incorporated into customized software known as the Cross-Country Scoring Reliability Software (CCSRS), developed by the DPC. English-speaking scorers in each of the northern-hemisphere countries used this software to independently score the student responses. The degree of agreement between scorers from the various countries may be taken as a measure of cross-country scoring reliability.

6.3.2 Data Entry

As described earlier in this chapter, the IEA Data Processing Center provided an integrated computer program for keyboard data entry and data verification known as DataEntryManager for Windows (WinDEM). This program works on all IBM-PC compatible personal computers running under MicroSoft's Windows operating system (Windows 95, 98, 2000, XP, and NT). WinDem imports student and teacher tracking information directly from the W3S sampling software, facilitating keyboard data entry of responses to test booklets and questionnaires. WinDEM also offered data and file management capabilities, a convenient checking and editing mechanism, interactive error detection, and reporting and quality control procedures. Detailed information and operational instructions were provided in the WinDem manual. Since WinDEM incorporated the international codebooks describing all variables, use of the software ensured that the data files were produced according to the TIMSS 2003 rules and standards for data entry. Although use of WinDEM 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. DPC staff provided training to NRCs and national center personnel at various stages of the project, including prior to the field test and for six countries again prior to the main data collection.

NRCs who chose not to use WinDEM for data entry still had to ensure that all data files delivered to the DPC were in the international format and had passed all of the verification checks built into the WinDEM program. This can be accomplished by running WinDEM in data-checking mode on the data files. The WinDEM data-checking facility identifies a range of problems with identification numbers, out-of-range and otherwise invalid codes, and data file structure that can can be rectified before submitting the files to the DPC. In addition to the data-validation checks incorporated in WinDEM, NRCs were expected to use the WinLINK (or LinkCheck) program supplied by the DPC to verify the integrity of the links between the various student, teacher, and school files. Data files were acceptable at the DPC only if the reports generated by the WinDEM program and WinLINK programs indicated no errors.

During the TIMSS 2003 data collection, data were gathered from several sources, including students, teachers, and principals, as well as from a range of tracking forms. These data were recorded into data files as follows:

- The **school background file** contained information from the school background questionnaire.
- The **mathematics teacher background file** (eighth grade only) contained information from the eighth-grade mathematics teacher questionnaire.
- The **science teacher background file** (eighth grade only) contained information from the eighth-grade science teacher questionnaire.
- The teacher background file (fourth grade only) contained information from the fourth-grade classroom teacher questionnaire.

- The **student background file** contained data from the student background questionnaire.
- The student achievement file contained the achievement test booklet data.
- The **constructed-response scoring reliability file** contained the within-country scoring reliability data for the constructed-response items.

When all data files had passed the WinDEM and WinLINK/LinkCheck quality control checks, they were dispatched to the IEA Data Processing Center in Hamburg for further checking and processing.

6.3.3 Survey Activities Report

NRCs were requested to maintain a record of their experiences during the TIMSS 2003 data collection and to send a report to the TIMSS International Study Center when data-collection activities were completed. The report should describe any problems or unusual occurrences in selecting the sample or securing school participation, translating or preparing the data-collection instruments, administering the tests and questionnaires in the schools, scoring the constructed-response items, or creating and checking the data files.

6.3.4 Data Management Forms

NRCs were requested to document in a series of Data Management Forms any adaptations to the international instruments that they made while producing their national instruments. These forms were sent to the TIMSS International Study Center as well as to the IEA Data Processing Center. The information is used in the data editing and formatting process to recode data wherever possible to a form that allows for international comparisons. Additionally, the information provided in the Data Management Forms is included in a supplement to the *TIMSS 2003 User Guide for the International Database*.

6.4 Summary

This chapter has summarized the design and implementation of the TIMSS 2003 field operations from the point of first contact with the sampled schools to the submission of the ckecked and verified data files to the IEA Data Processing Center. Although the procedures were sometimes complex, each step was clearly documented in the TIMSS operations manuals and supported by training sessions at the NRC meetings. NRC Survey Activities Reports indicated that the field operations generally went well, and that the TIMSS 2003 data were of high quality.

References

TIMSS (2001), *TIMSS 2003 School Sampling Manual*, prepared by P. Foy & M. Joncas, Statistics Canada, Chestnut Hill, MA: Boston College.

TIMSS (2002a), *TIMSS 2003 Survey Operations Manual*, prepared by the International Study Center, Chestnut Hill, MA: Boston College.

TIMSS (2002b), *TIMSS 2003 School Coordinator Manual*, prepared by the International Study Center, Chestnut Hill, MA: Boston College.

TIMSS (2002c), *TIMSS 2003 Test Administrator Manual*, prepared by the International Study Center, Chestnut Hill, MA: Boston College.

TIMSS (2002d), *TIMSS 2003 Scoring Guides for Mathematics and Science Constructed-Response Items – Eighth Grade*, prepared by the International Study Center, Chestnut Hill, MA: Boston College.

TIMSS (2002e), *TIMSS 2003 Scoring Guides for Mathematics and Science Constructed-Response Items – Fourth Grade*, prepared by the International Study Center, Chestnut Hill, MA: Boston College.

TIMSS (2002f), *Manual for Entering the TIMSS 2003 Data*, prepared by the International Study Center in cooperation with the IEA Data Processing Center, Chestnut Hill, MA: Boston College.

TIMSS (2002g), *TIMSS 2003 National Quality Control Observer Manual*, prepared by the International Study Center, Chestnut Hill, MA: Boston College.

7

Chapter 7

Quality Assurance in the TIMSS 2003 Data Collection

Eugenio J. Gonzalez and Dana Diaconu

7.1 Overview

As part of its overall quality assurance efforts, TIMSS conducted an ambitious program of site visits to document the quality of the TIMSS 2003 data collection. Together with the IEA Secretariat and the national centers, the TIMSS & PIRLS International Study Center (ISC) identified and appointed one International Quality Control Monitor (QCM) in each country to observe data collection procedures at both national and classroom levels.

Quality Control Monitors had two major responsibilities: to interview the National Research Coordinator (NRC) about the survey operations and activities, and to conduct site visits to a random sample of 15 schools in the country at each grade assessed during test administration. The QCMs attended a two-day training session conducted by the ISC and the IEA Secretariat,[1] where they were introduced to the TIMSS 2003 survey operations procedures and instructed on how to conduct their interviews and site visit observations. At the training session, QCMs received a copy of the *TIMSS 2003 Manual for International Quality Control Monitors* (TIMSS, 2002a), which explained their duties in detail, and copies of the *Survey Operations Manual* (TIMSS, 2002b), *School Coordinator Manual* (TIMSS, 2002c), *and Test Administrator Manual* (TIMSS, 2002d).

Fifty QCMs were trained across the 49 countries and four Benchmarking participants where the international quality control program was conducted.[2] Where necessary, QCMs who attended the training session were permitted to recruit other QCMs to assist them in covering the territory and

1 Two training sessions were conducted, one for countries in the southern hemisphere and the other for northern hemisphere countries.

2 Iran and Israel were the only countries whose QCMs were not trained; Ontario and Quebec shared the same QCM.

meeting the testing timetable. All together, these monitors and those trained by them observed 1147 testing sessions (755 for grade 8 and 392 for grade 4),[3] and conducted interviews with the National Research Coordinator in each of the participating countries. Exhibit 7.1 indicates the dates of data collection and the number of site visits by QCMs in each country.

7.2 Observing the TIMSS Test Administration

When visiting the school, the QCM had to complete a Classroom Observation Record Form. This form was organized into four sections as follows:

- Preliminary activities of the Test Administrator
- Test session activities
- Summary observations
- Interview with the School Coordinator

7.2.1 Preliminary Activities of the Test Administrator

Section A of the Classroom Observation Record addressed the extent to which the Test Administrator had prepared for the testing session. Monitors were asked to note the following activities of the Test Administrator: checking the testing materials, reading the administration script, organizing space for the session, and arranging for the necessary equipment (e.g., pencils, a watch for timing the testing session).

Exhibit 7.2 summarizes the results for Section A for the eighth grade. In almost all testing sessions, Test Administrators observed the proper preparatory procedures. When deviations occurred, the QCMs provided reasonable explanations for the discrepancies. For example, QCMs noted that the main reason for students receiving booklets with student identifications that did not correspond to the Student Tracking Form was because new students did not appear on the list, as the tracking forms had been created before the students were enrolled. In the few cases where there reportedly was not enough room for students, QCMs indicated that it was due to unavoidable circumstances (e.g., the test was administered in a small classroom, students had to sit two or three at one desk or in groups of five or six around a table).

3 Operational constraints did not permit QCM visits to be conducted in five testing sessions in Japan.

Exhibit 7.1 **TIMSS 2003 International Quality Control Site Visits**

Countries	Eighth Grade		Fourth Grade	
	Date of Data Collection	Number of Site Visits	Date of Data Collection	Number of Site Visits
Argentina	Nov. 2003	16		
Armenia	May 2003	15	May 2003	15
Australia	Oct. - Nov. 2002	15	Nov. 2002	15
Bahrain	Apr.-May 2003	15		
Belgium (Flemish)	May 2003	15	May 2003	15
Botswana	Oct. - Nov. 2002	15		
Bulgaria	Apr.-May 2003	15		
Chile	Nov. 2002	19		
Chinese Taipei	May 2003	15	June 2003	15
Cyprus	May 2003	15	May 2003	14
Egypt	May 2003	15		
England	June 2003	15	May 2003	15
Estonia	June 2003	15		
Ghana	Apr.-May 2003	14		
Hong Kong, SAR	May 2003	15	May - June 2003	15
Hungary	March 2003	15	March – Apr. 2003	15
Indonesia	May 2003	15		
Iran, Islamic Rep. of	Apr.-May 2003	15	Apr.-May 2003	15
Israel	May 2003	15		
Italy	Apr.-May 2003	16	Apr.-May 2003	14
Japan	Feb. 2003	10	Feb. 2003	11
Jordan	May 2003	15		
Korea, Rep. of	Apr. 2003	15		
Latvia	May 2003	15	May 2003	15
Lebanon	Apr. 2003	15		
Lithuania	May 2003	15	May 2003	15
Macedonia, Rep. of	May 2003	15		
Malaysia	Oct. 2002	15		
Moldova, Rep. of	May 2003	15	May 2003	15

Exhibit 7.1 **TIMSS 2003 International Quality Control Site Visits** *(...Continued)*

Countries	Eighth Grade		Fourth Grade	
	Date of Data Collection	Number of Site Visits	Date of Data Collection	Number of Site Visits
Morocco	June 2003	15	May 2003	15
Netherlands	Apr.-May 2003	13	Apr. 2003	14
New Zealand	Nov. 2002	13	Nov. 2002	16
Norway	Apr. 2003	20	Apr. 2003	10
Palestinian Nat'l Auth.	Apr.-May 2003	15		
Philippines	March 2003	16	March 2003	14
Romania	May – June 2003	15		
Russian Federation	Apr.-May 2003	15	Apr.-May 2003	15
Saudi Arabia	May 2003	15		
Scotland	Apr.-May 2003	15	March – May 2003	15
Serbia	May 2003	15		
Singapore	Oct. 2002	15	Oct. 2002	15
Slovak Republic	May 2003	15		
Slovenia	Apr.-May 2003	15	May 2003	15
South Africa	Oct. 2002	15		
Sweden	May 2003	15		
Syria	May 2003	15		
Tunisia	May 2003	14		
United States	Apr.-May 2003	17	Apr.-May 2003	14
Yemen			May 2003	15
Benchmarking Participants				
Basque Country, Spain	May 2003	16		
Indiana State, US[4]	Apr. 2003	15	Apr. 2003	15
Ontario Province, Can.				
Quebec Province, Can.				
TOTAL		**755**		**392**

4 Data collection for Indiana was conducted by Westat, Inc., using the same procedures that it applied in collecting the data for the United States' national sample for TIMSS 2003.

Exhibit 7.2 **Preliminary Activities of the Test Administrator - Eighth Grade**

Question	Yes	No	N/A
Had the Test Administrator verified adequate supplies of the test booklets?	729*	22**	4
Did the student identification information on the test booklets and student questionnaires correspond with the Student Tracking Form?	741	11	3
Had the Test Administrator familiarized himself or herself with the test administration script prior to the testing?	729*	21**	5
Was there adequate seating space for the students to work without distractions?	737	17	1
Was there adequate room for the Test Administrator to move about during the testing to ensure that students were following directions correctly?	738	17	0
Did the Test Administrator have a stopwatch or timer for accurately timing the testing session?	723	24	8
Did the Test Administrator have an adequate supply of pencils and other necessary materials ready for the students?	646	102	7

* Represents the number of respondents answering either "Definitely Yes" or "Probably Yes"
** Represents the number of respondents answering either "Definitely No" or "Probably No"

The absence of a stopwatch was not considered a serious limitation. Test Administrators who did not have a stopwatch had a wristwatch available to monitor the time remaining in the test sessions. In about 14 percent of the testing sessions, the QCMs noted that the Test Administrators did not have an adequate supply of pencils for the students. However, in most of these cases, students provided their own. In general, QCMs observed no procedural deviations in test preparations severe enough to jeopardize the integrity of the test administration.

Exhibit 7.3 summarizes the results for Section A for the fourth grade. Similar to the eighth grade, in almost all testing sessions Test Administrators observed the proper preparatory procedures, and when deviations occurred the QCMs provided reasonable explanations for the discrepancies. As at the eighth grade, QCMs observed no procedural deviations in test preparations severe enough to jeopardize the integrity of the test administration.

7.2.2 Test Session Activities

Section B of the Classroom Observation Record addressed the activities that took place during the actual testing session. These activities included following the Test Administrator script, distributing and collecting test booklets, and making announcements during the testing sessions.

Exhibit 7.3 **Preliminary Activities of the Test Administrator - Fourth Grade**

Question	Yes	No	N/A
Had the Test Administrator verified adequate supplies of the test booklets?	369*	16**	6
Did the student identification information on the test booklets and student questionnaires correspond with the Student Tracking Form?	378	7	6
Had the Test Administrator familiarized himself or herself with the test administration script prior to the testing?	378*	9**	4
Was there adequate seating space for the students to work without distractions?	378	8	5
Was there adequate room for the Test Administrator to move about during the testing to ensure that students were following directions correctly?	382	4	5
Did the Test Administrator have a stopwatch or timer for accurately timing the testing session?	371	13	7
Did the Test Administrator have an adequate supply of pencils and other necessary materials ready for the students?	342	40	9

* Represents the number of respondents answering either "Definitely Yes" or "Probably Yes"

** Represents the number of respondents answering either "Definitely No" or "Probably No"

The achievement test for grade in 8 was administered in two sessions of 45 minutes each, with a short break between. Exhibit 7.4 documents the activities associated with the first testing session and shows that at least 80 percent of the Test Administrators followed their script exactly when preparing the students, distributing the test materials, and beginning testing. In the rare instances when changes were made to the script, these tended to be additions or revisions, rather than deletions.

In only about five percent of the sessions visited, the total testing time for Session 1 was not equal to the time allowed. However, in most of these sessions, this was because all students had completed Session 1 before the allotted time had elapsed, and so the Test Administrator reasonably went on with the next part of the session according to the prescribed procedures. The average testing time for Session 1 was approximately 45 minutes, same as the allocated time.

Exhibit 7.4 also shows that only in about half of the sessions did the Test Administrator collect booklets one at a time at the end of the session, as prescribed in the directions. While this may seem surprising, it turns out that when the booklets were not collected individually from each student, students were instructed to close their test booklets and leave them on their desks during the break. The room was then either secured or supervised during the break.

When asked whether the break between sessions was 20 minutes long, QCMs tended to interpret the question quite literally. As a result, QCMs reported that only about half of classrooms started the test after a break that was "exactly" 20 minutes. The remainder reported having breaks that ranged from no break at all (with all students' agreement) to about one hour.

The achievement test for grade 4 was administered in two sessions of 36 minutes each with a short break in between. Exhibit 7.5 documents the activities associated with the first testing session and shows that about three-quarters of the Test Administrators followed their script exactly when preparing the students, distributing the test materials, and beginning testing. As at grade 8, in the rare instances when changes were made to the script, these tended to be additions, rather than revisions or deletions.

In almost all of the sessions visited the total testing time for Session 1 corresponded to the time allowed. Where it did not, it was because all students had completed Session 1 before the allotted time had elapsed, and the Test Administrator went on with the next session. The average testing time for Session 1 was approximately 36 minutes, identical to the allocated time.

Mirroring grade 8, Exhibit 7.5 also shows that in less than half of the sessions the Test Administrator collected booklets one at a time at the end of the session, as prescribed in the directions. Again, when the booklets were not collected individually from each student, students were instructed to close their test booklets and leave them on their desk during the break. The room was then either secured or supervised during the break, in some instances by the QCM.

Similar to grade 8, when asked whether the break between sessions was 20 minutes long, QCMs tended to interpret the question quite literally. As a result, only 35 percent of Test Administrators reported that the test started after a break that was "exactly" 20 minutes. The total break time across all countries ranged from one to 50 minutes.

Exhibit 7.4 **Testing Session 1 – Eighth Grade**

Question	Yes	No	N/A
Did the Test Administrator follow the Test Administrator's script exactly in each of the following tasks?			
Prepare the students	619	119 (minor changes) 11 (major changes)	6
Distribute the materials	661	70 (minor changes) 13 (major changes)	11
Begin testing	661	69 (minor changes) 12 (major changes)	13
If the Test Administrator made changes to the script, how would you describe them?			
Additions	103	243	409
Revisions	100	245	410
Deletions	58	256	441
Did the Test Administrator distribute test booklets one at a time to each student?	692	52	11
Did the Test Administrator distribute the test booklets according to the booklet assignments on the Student Tracking Form?	738	12	5
Did the Test Administrator record attendance correctly on the Student Tracking Form?	728	11	16
Did the total testing time for Session 1 equal the time allowed?	715	36	4
Did the Test Administrator announce "you have 10 minutes left" prior to the end of Session 1?	721	31	3
Were there any other time remaining announcements made during Session 1?	124	620	11
At the end of Session 1, did the Test Administrator collect the test booklets one at a time from each student?	406	341	8
Was the total time for the break between Session 1 and Session 2 equal to 20 minutes?	344	402	9

Exhibit 7.5 **Testing Session 1 – Fourth Grade**

Question	Yes	No	N/A
Did the Test Administrator follow the Test Administrator's script exactly in each of the following tasks?			
Prepare the students	299	81 (minor changes) 5 (major changes)	5
Distribute the materials	341	40 (minor changes) 2 (major changes)	8
Begin testing	325	54 (minor changes) 2 (major changes)	8
If the Test Administrator made changes to the script, how would you describe them?			
Additions	90	117	184
Revisions	44	147	200
Deletions	17	166	208
Did the Test Administrator distribute test booklets one at a time to each student?	383	3	5
Did the Test Administrator distribute the test booklets according to the booklet assignments on the Student Tracking Form?	383	3	5
Did the Test Administrator record attendance correctly on the Student Tracking Form?	373	9	9
Did the total testing time for Session 1 equal the time allowed?	375	8	8
Did the Test Administrator announce "you have 10 minutes left" prior to the end of Session 1?	374	11	6
Were there any other time remaining announcements made during Session 1?	52	330	9
At the end of Session 1, did the Test Administrator collect the test booklets one at a time from each student?	180	199	12
Was the total time for the break between Session 1 and Session 2 equal to 20 minutes?	139	238	14

Exhibit 7.6 summarizes QCMs' observations from the second testing session for grade 8. In the vast majority of sessions, the Test Administrator kept to the time limits prescribed in the directions. Exhibit 7.6 also reveals that in about 70 percent of the sessions the Test Administrator kept to the testing script for signaling a break. Those who did make changes mostly made additions or other minor changes such as paraphrasing the directions. However, here too, QCMs took the question about time for restarting literally. In more than half of the sessions, the time spent to restart the testing session was the prescribed five minutes. For the rest, the session took up to 10 minutes longer to restart. Finally, this exhibit also shows that in only one-quarter of the sessions did students request additional time to complete the student questionnaire. In most cases, this request was granted.

Exhibit 7.6 **Testing Session 2 – Eighth Grade**

Question	Yes	No	N/A
Was the time spent to restart the testing with Session 2, 5 minutes?	437	314	4
Was the total time for testing in Session 2 correct as indicated in the Administrators' script?	718	27	10
Did the Test Administrator announce "you have 10 minutes left" prior to the end of Session 2?	729	21	5
Were there any other time remaining announcements made during Session 2?	113	631	11
At the end of Session 2, did the Test Administrator collect the test booklets one at a time from each student?	650	91	14
When the Test Administrator read the script to end the testing for Session 2, did the Test Administrator announce a break to be followed by the Student Questionnaire?	610	89	56
How accurately did the Test Administrator read the script to end the testing and signal a break?	531(no changes)	135 (minor changes) 22 (major changes)	67
If there were changes, how would you describe them?			
Additions	45	179	531
Some minor changes	94	139	522
Omissions	41	165	549
At the end of the break, did the Test Administrator distribute the Student Questionnaires and give directions as specified in the script?	585	82	88
Did the students ask for additional time to complete the questionnaire?	192	494	69
At the end of the session, prior to dismissing the students, did the Test Administrator thank the students for participating in the study?	622	68	65

Exhibit 7.7 summarizes QCMs' observations from the second testing session for grade 4. In the large majority of sessions the Test Administrator kept to the time limits prescribed in the directions. About 60 percent of the Test Administrators stuck to the testing script for signaling a break. Of those who did make changes, most made minor changes such as paraphrasing the directions. Similar to grade 8, QCMs here also took the question about time for restarting literally. In about 40 percent of the sessions, the time spent to restart the testing session was the prescribed five minutes. For the rest, the session took up to 10 minutes longer to restart. Only about one-quarter of students requested additional time to complete the student questionnaire. In most cases, this request was granted.

Exhibit 7.7 **Testing Session 2 – Fourth Grade**

Question	Yes	No	N/A
Was the time spent to restart the testing with Session 2, 5 minutes?	169	215	7
Was the total time for testing in Session 2 correct as indicated in the Administrators' script?	372	10	9
Did the Test Administrator announce "you have 10 minutes left" prior to the end of Session 2?	367	15	11
Were there any other time remaining announcements made during Session 2?	48	333	10
At the end of Session 2, did the Test Administrator collect the test booklets one at a time from each student?	322	59	10
When the Test Administrator read the script to end the testing for Session 2, did the Test Administrator announce a break to be followed by the Student Questionnaire?	301	40	50
How accurately did the Test Administrator read the script to end the testing and signal a break?	242	84 (minor changes) 10 (major changes)	53
If there were changes, how would you describe them?			
Additions	29	68	294
Some minor changes	66	54	271
Omissions	25	77	289
At the end of the break, did the Test Administrator distribute the Student Questionnaires and give directions as specified in the script?	288	42	61
Did the students ask for additional time to complete the questionnaire?	96	243	52
At the end of the session, prior to dismissing the students, did the Test Administrator thank the students for participating in the study?	304	35	52

Responses to the remaining questions focusing on the test session activities for eighth grade are summarized in Exhibit 7.8. These questions dealt with topics such as student compliance with instructions and the alignment of the scripted instructions with their implementation. Exhibit 7.8 shows that in almost all of the sessions, the students complied well or very well with the instructions to stop testing. In addition, in at least 70 percent of the sessions, breaks were conducted exactly or nearly exactly as directed in the script. When this was not the case, it was mostly due to differences in the amount of time allocated for the break. It is also notable that in 95 percent of the testing sessions calculators were not allowed for Session 1 – as required in the script – while in 80 percent of cases calculators were allowed for Session 2.

Exhibit 7.8 **Test Session Activities – Eighth Grade**

Question	Very Well	Well	Fairly Well	Not well at all	N/A
When the Test Administrator ended Session 1, how well did the students comply with the instruction to stop work (close their booklets and put their pencils down)?	590	136	16	0	13
When the Test Administrator ended Session 2, how well did the students comply with the instruction to stop work (close their booklets and put their pencils down)?	584	142	21	0	8

	Exactly	Nearly the same	Somewhat differently	Not well at all	N/A
Was the first break conducted as directed in the script?	541	133	56	8	17
Was the second break conducted as directed in the script?	457	72	37	48	141

	Exactly the same	Longer	Shorter	N/A
How did the actual break time compare to the recommended time in the script?	314	113	166	162
How did the total time allocated for the administration of the Student Questionnaire compare to the time specified in the script?	420	150	111	74

	Yes	No	N/A
Were calculators allowed during Session 1?	43	702	10
Were calculators allowed during Session 2?	604	142	9

	Very orderly	Somewhat orderly	Not orderly at all	N/A
In your opinion, how orderly was the dismissal of the students?	502	184	11	58

Exhibit 7.9 presents the results of the remaining questions that focused on the test session activities for grade 4. Similar to the eighth grade, Exhibit 7.9 shows that in almost all the sessions the students complied well or very well with the instructions to stop testing. In addition, in at least two-thirds of the sessions breaks were conducted exactly or nearly exactly as directed in the script. When this was not the case, it was mostly due to differences in the amount of time allocated for the break. It is also notable that calculators were *not* allowed in almost all testing sessions.

Exhibit 7.9 **Test Session Activities – Fourth Grade**

Question	Very Well	Well	Fairly Well	Not well at all	N/A
When the Test Administrator ended Session 1, how well did the students comply with the instruction to stop work (close their booklets and put their pencils down)?	311	56	7	1	16
When the Test Administrator ended Session 2, how well did the students comply with the instruction to stop work (close their booklets and put their pencils down)?	317	58	6	0	10

	Exactly	Nearly the same	Somewhat differently	Not well at all	N/A
Was the first break conducted as directed in the script?	255	74	43	3	16
Was the second break conducted as directed in the script?	213	54	16	1098	

	Yes	No	N/A
Were calculators allowed during Session 1?	1	382	8
Were calculators allowed during Session 2?	21	358	12

	Exactly the same	Longer	Shorter	N/A
How did the actual break time compared to the recommended time in the script?	123	68	71	129
How does the total time allocated for the administration of the Student Questionnaire compare to the time specified in the script?	158	110	68	55

	Very orderly	Somewhat orderly	Not orderly at all	N/A
In your opinion, how orderly was the dismissal of the students?	269	68	1	53

7.2.3 Summary Observations

Section C of the Classroom Observation Record asked QCMs to reflect on their observations. The QCMs reported overall impressions of the test administration, including how well the Test Administrator monitored students' conduct, and any unusual circumstances that arose during the testing session (e.g., student refusal to participate, defective instrumentation, emergency situations, cheating).

The results presented in Exhibit 7.10 for grade 8 show that in almost all sessions the testing took place without any problems. In the few sessions where problems arose due to defective instrumentation, the Test Administrator replaced the instruments appropriately. In less than five percent of sessions, QCMs reported evidence of students attempting to cheat on the exam. However, when asked to explain the situation, QCMs generally indicated that students were merely looking around at their neighbors to see whether their test booklets were indeed different. Because the TIMSS 2003 test design involves 12 different booklets, students were unlikely to have the same booklet as their neighbors. The QCMs reported that on the rare occasions when they observed serious efforts to cheat, the Test Administrator intervened to prevent cheating. Most of the 31 students who were reported to leave the room for an "emergency" during the testing session had already completed the test. When students left the room for an emergency, Test Administrators handled the situation appropriately by ensuring the security of the test booklets until the students returned. Students were permitted to complete the test when they returned to the classroom.

Exhibit 7.10 also indicates that in almost all of the testing sessions at the eighth grade, QCMs found the behavior of students to be orderly and cooperative. The problem cited most often by QCMs as the reason for disorderly behavior was the noise level of students who had completed the test well before the prescribed 45 minutes had passed. In the few cases where students were disruptive, the Test Administrator was able to control the situation. For the great majority of sessions, QCMs reported that the overall quality of the sessions was either excellent or very good.

Exhibit 7.10 **Summary Observations of the QCM – Eighth Grade**

Question	Yes	No	N/A
During the testing sessions did the Test Administrator walk around the room to be sure students were working on the correct section of the test and/or behaving properly?	727	17	11
Did the Test Administrator address students' questions appropriately?	720	26	9
Did you see any evidence of students attempting to cheat on the tests (e.g., by copying from a neighbor)?	39	708	8
Were any defective test booklets detected and replaced before the testing began?	15	726	14
Were any defective test booklets detected and replaced after the testing began?	20	706	29
If any defective test booklets were replaced, did the Test Administrator replace them appropriately?	44	19	692
Did any students refuse to take the test either prior to the testing or during the testing?	17	714	24
If a student refused, did the Test Administrator accurately follow the instructions for excusing the student (collect the test book and record the incident on the Student Tracking Form)?	32	16	707
Did any students leave the room for an "emergency" during the testing?	61	685	9
If a student left the room for an emergency during the testing, did the Test Administrator address the situation appropriately (collect the test booklet, and if re-admitted, return the test booklet)?	56	19	680

	Extremely	Moderately	Somewhat	Hardly	N/A
To what extent would you describe the students as orderly and cooperative?	511	207	26	3	8

	Definitely	Some effort	Hardly any effort	N/A
If the students were not cooperative and orderly, did the Test Administrator make an effort to control the students and the situation?	199	37	2	517

	No, there were no late students	No, they were not admitted	Yes, but before testing began	Yes, after testing began	N/A
Were any late students admitted to the testing room?	659	25	32	27	12

	Excellent	Very good	Good	Fair	Poor	N/A
In general, how would you describe the overall quality of the testing session?	380	278	68	15	6	8

Exhibit 7.11 presents QCMs' summary observations for fourth grade. Similar to eighth grade, in almost all sessions the testing took place without any problems. In the few sessions where problems arose due to defective instrumentation, the Test Administrator replaced the instruments appropriately. In less than four percent of the sessions, QCMs reported evidence of students attempting to cheat on the exam. Like at grade 8, when asked to explain the situation, QCMs indicated that students were merely looking around at their neighbors to see whether their test booklets were indeed different. The QCMs reported that on the rare occasions when they observed serious efforts to cheat, the Test Administrator intervened to prevent cheating. Most of the 58 students who were reported to leave the room for an "emergency" during the testing session had already completed the test. When students left the room for an emergency, Test Administrators handled the situation appropriately by ensuring the security of the test booklets until the students returned. Students were permitted to complete the test when they returned to the classroom.

Exhibit 7.11 also indicates that in almost all of the testing sessions at the fourth grade, QCMs found the behavior of students to be orderly and cooperative. In the few cases where students were disruptive, the Test Administrator was able to control the situation. For the great majority of sessions, QCMs reported that the overall quality of the sessions was either excellent or very good.

Exhibit 7.11 **Summary Observations of the QCM – Fourth Grade**

Question	Yes	No	N/A
During the testing sessions did the Test Administrator walk around the room to be sure students were working on the correct section of the test and/or behaving properly?	371	11	9
Did the Test Administrator address students' questions appropriately?	380	2	9
Did you see any evidence of students attempting to cheat on the tests (e.g., by copying from a neighbor)?	16	368	7
Were any defective test booklets detected and replaced before the testing began?	4	377	10
Were any defective test booklets detected and replaced after the testing began?	11	362	18
If any defective test booklets were replaced, did the Test Administrator replace them appropriately?	19	15	357
Did any students refuse to take the test either prior to the testing or during the testing?	14	355	22
If a student refused, did the Test Administrator accurately follow the instructions for excusing the student (collect the test book and record the incident on the Student Tracking Form)?	13	9	369
Did any students leave the room for an "emergency" during the testing?	31	349	11
If a student left the room for an emergency during the testing, did the Test Administrator address the situation appropriately (collect the test booklet, and if re-admitted, return the test booklet)?	23	13	355

	Extremely	Moderately	Somewhat	Hardly	N/A	
To what extent would you describe the students as orderly and cooperative?	292	86	6	0	7	
	Definitely	**Some effort**	**Hardly any effort**	**N/A**		
If the students were not cooperative and orderly, did the Test Administrator make an effort to control the students and the situation?	103	12	0	276		
	No, there were no late students	**No, they were not admitted**	**Yes, but before testing began**	**Yes, after testing began**	**N/A**	
Were any late students admitted to the testing room?	344	13	5	5	24	
	Excellent	**Very good**	**Good**	**Fair**	**Poor**	**N/A**
In general, how would you describe the overall quality of the testing session?	194	146	37	0	0	355

7.2.4 Interview with the School Coordinator

The QCM recorded details of the interview with the School Coordinator in Section D of the Classroom Observation Record. The interview addressed the shipment of assessment materials, arrangements for the test administration, the responsiveness of the NRC to queries, the necessity for make-up sessions, and, as a validation of within school sampling procedures, the organization of classes in the school.

The results presented in Exhibit 7.12 for the eighth grade show that TIMSS 2003 was an administrative success in the eyes of School Coordinators. In more than 70 percent of the cases, school officials received the correct shipment of the test materials. Mistakes that did occur tended to be minor and could be remedied prior to testing. Furthermore, about 85 percent of School Coordinators reported that the NRCs were responsive to their questions or concerns, and that the relations were cordial and cooperative. About half of the School Coordinators reported that they were able to collect the completed Teacher Questionnaires prior to student testing. Of the rest, the vast majority reported that they were missing only one or two questionnaires and were expecting them to be turned in shortly. It was estimated that the Teacher Questionnaires would take about 60 minutes to complete. About 50 percent of the School Coordinators indicated that the estimate of 60 minutes was about right, while about 10 percent reported that the questionnaires took longer and about 22 percent that they took less time to complete.

In about 50 percent of the cases, School Coordinators indicated that students were given special instructions, motivational talks, or incentives prior to testing. The majority of students received motivational talks either by a school official, classroom teacher, or the TIMSS Test Administrator. Only a few classes received special instructions or practice, such as reading competitions or extra reading assignments prior to the testing session.

Because the sampling of classes requires a complete list of all mathematics classes in the school at the target grade, QCMs were asked to verify that the class list did indeed include all classes. Although a significant number of School Coordinators reported that this was not so, there may have been some misunderstanding since very few (about 3 percent) knew of any students not in the classes listed.

A tribute to the planning and implementation of TIMSS 2003 was the fact that about 85 percent of respondents said they would be willing to serve as a School Coordinator in future international assessments. Furthermore, the majority of School Coordinators believed the testing session went very well, and that school staff members had positive attitudes towards the TIMSS testing.

Exhibit 7.12 **QCM Interviews with the School Coordinator – Eighth Grade**

Question	Yes	No	N/A
Prior to the test day did you have time to check your shipment of materials from your TIMSS National Coordinator?	545	122	88
Did you receive the correct shipment of the following items?			
Test booklets	604	55	96
Test Administrator Manual	566	92	97
School Coordinator Manual	556	94	105
Student Tracking Forms	632	35	88
Student Questionnaires	609	51	95
Teacher Questionnaires	639	50	66
School Questionnaire	655	33	67
Test Administration Form	547	112	96
Teacher Tracking Form	470	176	109
Student-Teacher Linkage Form (if applicable)	264	279	212
Envelopes or boxes addressed to the National Center for the purpose of returning the materials after the assessment	453	189	113
Was the National Coordinator responsive to your questions or concerns?	642	21	92
Were you able to collect completed Teacher Questionnaires prior to the test administration?	356	337	62
Was the estimated time of 60 minutes to complete the Teacher Questionnaire a correct estimate?	373	81 (more time) 166 (less time)	135
Were you satisfied with the accommodations (testing room) you were able to arrange for the testing?	695	22	38
Do you anticipate that makeup sessions will be required at your school?	77	625	53
If yes, do you intend to conduct one?	94	119	542
Did the students receive any special instructions, motivational talk, or incentives to prepare them for the assessment?	378	331	46
Is this a complete list of the mathematics classes in this grade in this school?	561	82	112
To the best of your knowledge, are there any students in this grade level who are not in any of these mathematics classes?	25	606	124
To the best of your knowledge are there any students in this grade level in more than one of these mathematics classes?	16	633	106
If there were another international assessment, would you be willing to serve as a School Coordinator?	647	45	63

Exhibit 7.12 **QCM Interviews with the School Coordinator – Eighth Grade** *(...Continued)*

	Very well, no problems	Satisfactorily, few problems	Unsatisfactorily, many problems	N/A
Overall, how would you say the session went?	616	94	4	41
	Positive	**Neutral**	**Negative**	**N/A**
Overall, how would you rate the attitude of the other school staff members towards the TIMSS testing?	549	159	10	37
	Worked well	**Needs improvement**	**N/A**	
Overall, do you feel the TIMSS School Coordinator Manual worked well or does it need improvement?	584	33	138	

Similar to the eighth grade, the administrative success of TIMSS 2003 at the fourth grade is exemplified by the results of the QCM interviews with School Coordinators, presented in Exhibit 7.13. School Coordinators received the correct shipment of the test materials in most cases. In cases where shipment errors occurred, they tended to be minor and were remedied prior to testing. More than 85 percent of School Coordinators reported that the NRCs were responsive to their questions or concerns.

About half the School Coordinators reported that they were able to collect the completed Teacher Questionnaires prior to student testing. Of those who did not, most reported that teachers completed their questionnaires during the testing sessions. Almost half of the School Coordinators indicated that the estimate of 60 minutes to complete the questionnaire was accurate, while only 11 percent reported that the questionnaires took longer and about 26 percent that they took less time to complete.

In about 35 percent of the cases, School Coordinators indicated that students were given special instructions, motivational talks, or incentives prior to testing. The majority of students received motivational talks either by a school official, classroom teacher, or the TIMSS Test Administrator. Only a few classes received special instructions or practice, such as reading competitions or extra reading assignments prior to the testing session.

As at the eighth grade, a large majority (more than 85 percent) of School Coordinators said they would be willing to serve again in future international assessments. Furthermore, the majority of School Coordinators believed that the testing session went very well, and that school staff members had positive attitudes towards the TIMSS testing.

Exhibit 7.13 **QCM Interviews with the School Coordinator – Fourth Grade**

Question	Yes	No	N/A
Prior to the test day did you have time to check your shipment of materials from your TIMSS National Coordinator?	282	72	37
Did you receive the correct shipment of the following items?			
Test booklets	347	15	29
Test Administrator Manual	321	42	28
School Coordinator Manual	288	43	60
Student Tracking Forms	341	20	30
Student Questionnaires	349	14	28
Teacher Questionnaires	334	15	42
School Questionnaire	362	0	29
Test Administration Form	320	42	29
Teacher Tracking Form	261	84	46
Student-Teacher Linkage Form (if applicable)	113	172	106
Envelopes or boxes addressed to the National Center for the purpose of returning the materials after the assessment	262	74	55
Was the National Coordinator responsive to your questions or concerns?	335	5	51
Were you able to collect completed Teacher Questionnaires prior to the test administration?	162	204	25
Was the estimated time of 60 minutes to complete the Teacher Questionnaire a correct estimate?	165	45 (longer) 102 (less time)	79
Were you satisfied with the accommodations (testing room) you were able to arrange for the testing?	371	6	14
Do you anticipate that makeup sessions will be required at your school?	44	334	13
If yes, do you intend to conduct one?	48	58	285
Did the students receive any special instructions, motivational talk, or incentives to prepare them for the assessment?	138	237	16
Is this a complete list of the mathematics classes in this grade in this school?	320	30	41
To the best of your knowledge, are there any students in this grade level who are not in any of these mathematics classes?	15	325	51
To the best of your knowledge, are there any students in this grade level in more than one of these mathematics classes?	7	342	42
If there were another international assessment, would you be willing to serve as a School Coordinator?	338	35	18

Exhibit 7.13 **QCM Interviews with the School Coordinator – Fourth Grade** *(...Continued)*

Question				
	Very well, no problems	Satisfactorily, few problems	Unsatisfactorily, many problems	N/A
Overall, how would you say the session went?	329	42	0	20
	Positive	Neutral	Negative	N/A
Overall, how would you rate the attitude of the other school staff members towards the TIMSS testing?	287	86	6	12
	Worked well	Needs improvement	N/A	
Overall, do you feel the TIMSS School Coordinator Manual worked well or does it need improvement?	297	11	83	

7.3 Interview with the National Research Coordinator

In addition to observing testing sessions, QCMs conducted face-to-face interviews with the National Research Coordinators for their countries. The QCM who attended the training session was responsible for conducting this interview and for completing an *Interview with the NRC* form.[5]

The interview questions were designed to elicit NRCs' experiences in preparing for and conducting the TIMSS 2003 data collection with a focus on identifying and selecting samples, working with School Coordinators, translating the instruments, assembling and printing the test materials, packing and shipping the test materials, scoring constructed-response questions, entering and verifying data, choosing quality assurance samples, and suggesting improvements in the process.

7.3.1 Sampling

Section A of the NRC interview involved questions about the sampling process. Topics covered in this section included the extent to which the NRCs used the manuals and sampling software provided by Statistics Canada and the IEA Data Processing Center (DPC) and found them helpful, and the difficulties encountered by NRCs as they carried out this task.

Exhibit 7.14 shows that six countries did not use the school sampling manual provided, but that was because Statistics Canada selected their sample. In one case (Bahrain), no school sampling was necessary because the TIMSS sample included the entire school population. Four-fifth of the NRCs used the within-school sampling software provided by the DPC to select classes. In the cases where the sampling software was not used, it was generally because the country had its own software.

5 A total of 50 QCM interviews with the NRCs were conducted. One interview was conducted for Ontario and Quebec combined.

A small number of NRCs encountered organizational constraints in their systems that necessitated deviations from the sample design. In each case, a sampling expert was consulted to ensure that the altered design remained compatible with the TIMSS standards. Sixty percent of NRCs reported that the sampling procedures were not unduly difficult to implement, while nearly 40 percent found the process somewhat difficult. Nevertheless, all but two NRCs managed to deliver school and student samples of high quality for the data collection.[6]

Exhibit 7.14 **Interview with the NRC – Sampling**

Question	Yes	No	N/A	
Were you able to select a sample of schools and students within schools using the Survey Operations Manual and the Sampling Manual provided by the TIMSS International Study Center?	44	6	0	
Did you use the Within-School Sampling Software provided by the TIMSS International Study Center to select classes or students?	40	9	1	
Were there any conditions or organizational constraints that necessitated deviations from the basic sampling TIMSS design?	8	42	0	
	Very difficult	Somewhat difficult	Not difficult at all	N/A
In terms of the complexity of the procedures and number of personnel needed, how would you describe the process of sample selection?	0	19	30	1

7.3.2 Working with School Coordinators

Questions in Section B of the NRC interview asked about cooperation with the School Coordinators, specifically about communication, shipment of materials, and training. A summary of the responses to the questions in Section B is presented in Exhibit 7.15. At the time the interviews were conducted, nearly all NRCs had contacted the School Coordinators in the sampled school, and most had sent the appropriate materials explaining the testing procedures. Where this was not the case, it was often because a meeting had been scheduled but not yet held. Half the NRCs planned to conduct formal training sessions for School Coordinators prior to the test administration.

Exhibit 7.15 **Interview with the NRC – School Coordinator**

Question	Yes	No	N/A
Have all the School Coordinators for your sample been contacted?	45	3	2
If yes, have you sent them materials about the testing procedures?	34	14	2
Did you have formal training sessions for the School Coordinators?	25	24	1

6 See Chapter 9 for information regarding countries' samples.

7.3.3 Translating the Instruments

Section C of the NRC interview dealt with translation and adaptation of the assessment instruments and manuals. Exhibit 7.16 shows that most NRCs were about evenly split between those who reported little difficulty in translating and adapting the test booklets and questionnaires and those who reported that this was somewhat or very difficult. Most NRCs, however, reported little difficulty in translating the Test Administrator and School Coordinator manuals.

In translating the test booklets, NRCs generally reported using their own staff or a combination of staff and outside experts. Almost all NRCs reported that they had submitted the achievement test booklets to the translation verification program at the International Study Center. At the time of the interview, almost all had received a translation verification report back. More than half the NRCs reported that they had already translated or planned to translate the scoring guides for the mathematics and science constructed-response items. Of those who did not translate the scoring guides, two countries reported that translation was not necessary, since all the scorers were proficient in English.

Exhibit 7.16 **Interviews with the NRC – Translation**

Question	Own Staff	Outside Experts	Combination	N/A
Did you use your own staff or outside experts to translate the test booklets for verification?	15	10	25	0
	Very difficult	**Somewhat difficult**	**Not difficult at all**	**N/A**
How difficult was it to translate and/or adapt the test booklets?	4	20	26	0
How difficult was it to adapt the questionnaires?	3	22	25	0
How difficult was it to adapt the Test Administrator Manual?	0	9	40	1
How difficult was it to adapt the School Coordinator Manual?	0	8	35	7
	Yes	No	N/A	
Did you go through the process of submitting test booklets and receiving a translation verification report from the ISC?	48[7]	1	1	
Did you translate or do you plan to translate the scoring guides for mathematics and science constructed-response items?	28	20	2	

7 Contrary to the data reported by the NRCs, all countries went through the translation verification process. See Chapter 4 for details.

7.3.4 Assembling and Printing the Test Materials

Section D of the NRC interview addressed assembling and printing the test materials, as well as issues related to checking the materials and securely storing them. The results in Exhibit 7.17 show that almost all NRCs were able to assemble the test booklets according to the instructions provided, and that nearly all conducted the recommended quality control checks during the printing process. In the cases where the NRCs did not conduct quality assurance procedures during the printing process, it was because of a shortage of time. Thirty percent of the NRCs detected errors during the printing process. Most countries elected to send their test booklets and questionnaires to an external printer, but printed the manuals in-house. Nearly all NRCs reported having followed procedures to protect the security of the tests during assembly and printing. In no instance was there a breach of security reported.

Exhibit 7.17 **Interview with the NRC – Assembling and Printing Test Materials**

Question	Yes	No	N/A
Were you able to assemble the test booklets according to the instructions provided by the International Study Center?	47	3	0
Did you conduct the quality assurance procedures for checking the test booklets during the printing process?	47	3	0
Were any errors detected during the printing process?	16	31	3
If errors were detected, what was the nature of the errors?			
Print quality	10	10	30
Pages missing	5	10	35
Page order	1	14	35
Upside down pages	2	13	35
Did you follow procedures to protect the security of the tests during the assembly and printing process?	49	1	0
Did you discover any potential breaches of security?	0	50	0

	In-House	External	Combination	N/A
Where did you print the test booklets?	8	34	7	1
Where did you print the questionnaires?	10	32	8	0
Where did you print the manuals (TA, SC, Scoring)?	31	11	7	1

7.3.5 Packing and Shipping the Testing Materials

Section E of the NRC interview addressed the extent to which NRCs detected errors in the testing materials as they were packed for shipping to School Coordinators. As shown in Exhibit 7.18, very few errors were found in any of the materials. Errors that were discovered before distribution were remedied. In 18 percent of the cases, the NRCs reported that they had some concerns about confidentiality that restricted their freedom to put student names on the booklet covers. Almost half the NRCs reported having established a procedure to confirm the schools' receipt of the testing materials and for verification of their contents. In most countries, NRCs reported that the deadline for the return of materials from schools was within a day or two of testing. All NRCs reported that the deadline was within two weeks of testing.

Exhibit 7.18 **Interview with the NRC – Packaging Test Materials**

Question	No Errors, or not used	Errors found before distribution	Errors found after distribution	N/A
In packing the assessment materials for shipment to schools, did you detect any errors in any of the following items?				
Supply of test booklets	32	5	1	12
Supply of student questionnaires	37	1	0	12
Student tracking forms	36	0	1	13
Teacher tracking forms	36	0	0	14
Student-Teacher Linkage Form	35	1	0	14
Test Administrator manual	37	1	0	12
School Coordinator manual	33	1	1	15
Supply of Teacher Questionnaires	37	0	1	12
School questionnaire	36	1	1	12
Test book ID labels	36	1	0	13
Sequencing of booklets or questionnaires	37	1	0	12
Return labels	33	1	1	15
Self-addressed post-cards for test dates	35	1	0	14
	Yes	**No**	**N/A**	
Did concerns about confidentiality restrict your freedom to put student names on the booklet covers?	9	32	9	
Do you plan to or have already established a procedure requiring schools to confirm receipt of the testing materials and verification of the contents?	26	12	12	

7.3.6 Scoring Constructed-Response Questions

Section F of the NRC interview focused on the NRC's preparation for scoring the constructed-response items. The scoring process was an ambitious effort, requiring the recruitment and training of scoring staff to score student responses including double scoring to verify scoring reliability. Exhibit 7.19 indicates that, at the time of the NRC interview, about 60 percent of the NRCs had selected their scoring staff, and roughly half of those had already begun the training process. All NRCs reported that they understood the procedures for scoring the reliability sample as explained in the Survey Operations Manual. Two NRCs reported that their own staff would score the constructed-response items, six reported that teachers would do so, six reported that university students would be employed, and 37 reported that a combination of various professionals would score the constructed-response items.

Exhibit 7.19 **Interview with the NRC – Scoring**

Question		Yes		No	N/A
Have you selected your scorers for the constructed-response questions?		29		20	1
If yes, have you trained the scorers?		17		18	15
Have you scheduled the scoring sessions for the constructed-response questions?		40		9	1
Do you understand the procedure for scoring the reliability sample as explained in the Survey Operations Manual?		50		0	0
	Own Staff	Teachers	University Students	Combination	Other
Who will primarily be scoring your constructed-response questions?	2	6	3	37	2

7.3.7 Data Entry and Verification

Section G of the NRC interview addressed preparations for data entry and verification. As shown in Exhibit 7.20, at the time of the interviews about two-thirds of the NRCs had selected their data entry staff and more than half of those selected had participated in training sessions. About two-thirds of the NRCs reported that they planned to enter the data from a percentage of booklets twice, as a verification procedure. The estimated proportion of booklets to be entered twice ranged from five to 50 percent, with two countries reporting that they planned to re-enter 100 percent of the data. Nearly all NRCs established a secure storage area for the returned tests after data entry. Twenty-two NRCs reported that members of their staff would enter the data from test booklets and questionnaires, six reported that an external agency would do so, and 18 reported that a combination of staff and external agency people would enter the data.

7.3.8 Quality Assurance Sample

As part of their national quality assurance activities, NRCs were required to send National Quality Control Observers to a 10 percent sample of the schools to observe the test administration and document compliance with prescribed procedures. These site visits were in addition to the visits to 15 schools conducted by the International Quality Control Monitors.

At the time of the NRC interviews, 64 percent of the NRCs had selected their 10 percent quality assurance sample for site visits. Two NRCs reported that an external agency would conduct the observations, 24 reported that a member of their staff would do so, and 12 reported that a combination of staff and external agency people would conduct the observations. Eight NRCs reported that other professionals, such as inspectors, retired teachers, mathematics and science supervisors or university professors, would be recruited to conduct the on-site observations.

Exhibit 7.20 **Interview with the NRC – Data Entry and Verification**

Question		Yes	No	N/A
Have you selected the data entry staff?		37	11	2
If yes, have you conducted training sessions for the data entry staff?		21	17	12
Do you plan to key enter a percentage of test booklets twice as a verification procedure?		37	10	3
Have you established a secure storage area for the returned tests after coding and until the original documents can be discarded?		48	2	0
	Own Staff	**External Firm**	**Combination**	**Other**
Do you plan to use your own staff or outside experts to enter the data from the achievement test booklets and questionnaires onto computer files?	22	6	18	4

7.3.9 The Survey Activities Report

The final section of the NRC interview asked the NRC for comments on any aspects of the study they felt might improve the assessment process. A major concern expressed by many NRCs was a time constraint for accomplishing all that was required to keep up with the demanding TIMSS 2003 schedule particularly the translation and preparation of the instruments. Some NRCs indicated they did not have ample staff.

References

TIMSS (2002a), *TIMSS 2003 Manual for International Quality Control Monitors*, prepared by the TIMSS & PIRLS International Study Center, Chestnut Hill, MA: Boston College.

TIMSS (2002b), *TIMSS 2003 Survey Operations Manual*, prepared by the TIMSS & PIRLS International Study Center, Chestnut Hill, MA: Boston College.

TIMSS (2002c), *TIMSS 2003 School Coordinator Manual*, prepared by the TIMSS & PIRLS International Study Center, Chestnut Hill, MA: Boston College.

TIMSS (2002d), *TIMSS 2003 Test Administrator Manual*, prepared by the TIMSS & PIRLS International Study Center, Chestnut Hill, MA: Boston College.

8

Chapter 8

Creating and Checking the TIMSS 2003 Database

Juliane Barth, Ralph Carstens, and Oliver Neuschmidt

8.1 Overview

Creating the TIMSS 2003 database and ensuring its integrity was a complex endeavor requiring close coordination and cooperation among the staff at the IEA Data Processing Center (DPC), the TIMSS & PIRLS International Study Center (ISC) at Boston College, Statistics Canada, and the national research centers of participating countries. The overriding concerns were: to ensure that all information in the database conformed to the internationally defined data structure; that national adaptations to questionnaires were reflected appropriately in the codebooks and documentation; and that all variables used for international comparisons were indeed comparable across countries. Quality control measures were applied throughout the process to assure the quality and accuracy of the TIMSS data.

This chapter describes the data entry and verification tasks undertaken by the National Research Coordinators (NRC) and data entry managers of participating countries, the data checking and database creation procedures implemented by the IEA Data Processing Center in collaboration with the International Study Center and Statistics Canada, and the steps taken at all institutions to confirm the integrity of the international database. Section 8.2 describes the quality measures taken in order to document the comparability and consistency of the scoring of constructed-response achievement items within countries, across countries, and over time (from 1999 to 2003).

8.2 Creating and Checking the TIMSS 2003 Database

Database construction began with each national research center using the data-entry software and codebooks provided by the IEA DPC (see Chapter 6) to enter the data collected in the TIMSS 2003 survey into data files following the standard international format. Before sending the files to the DPC, national center staff applied a system of checks specified by the DPC to verify the structure and consistency of the data files. Checking and editing the national data sets was a matter of cooperation between the national centers, the ISC, Statistics Canada, and the DPC team.

On receipt of the data files from each country, the IEA DPC was responsible for checking their integrity, for applying standard cleaning rules to verify the accuracy and consistency of the data, and for documenting electronically any deviation from the international file structure. Any queries were addressed to the national centers and modifications were made to the data files as necessary. After all modifications had been applied, all data were processed and checked again. This process of editing the data, checking the reports, and implementing corrections was repeated as many times as necessary until all data were consistent and comparable within and between countries.

In preparation for creating the international database, the Data Processing Center provided item statistics to the national research centers while the International Study Center provided countries with data almanacs containing international univariate statistics so that National Research Coordinators could examine their data from an international perspective. This was one of the most important checks (in terms of international comparability of the data). While in a national context a particular statistic may seem plausible, it may become apparent in comparing data across countries that it is an outlier in an international context, even with accurate translation. Any such instances were addressed, and the corresponding variables either recoded or removed from the international database.

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 National Research Coordinator from tracking forms and the *WinW3S: Within-School Sampling Software* (IEA, 2002a), was used by Statistics Canada to calculate sampling weights, population coverage, and school and student participation rates.[1]

1 See Chapter 9 for details about TIMSS 2003 sampling design.

When the review of the item statistics was completed and the Data Processing Center had updated the database to include sampling weights, the student achievement files were sent to the International Study Center where the IRT scaling was conducted and proficiency scores in mathematics and science generated for each participating student.[2] 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.

8.3 Data Entry at the National Research Centers

Each TIMSS 2003 national research center was responsible for transcribing the information from the achievement booklets and questionnaires into computer data files. As described in Chapter 6, the IEA DPC supplied national centers with the Windows DataEntryManager (WinDEM) software and manual (IEA, 2002b) to assist with data entry and held a training session on the use of the software. The DPC also provided countries with codebooks describing the structure of the data. The codebooks contained information about the variable names used for each variable in the survey instruments, and about field lengths, field locations, labels, valid ranges, default values, and missing codes. In order to facilitate data entry, the codebooks and data files were structured to match the test instruments and international version of the questionnaires. This meant that for each survey instrument there was a corresponding codebook, which served as a template for creating the corresponding survey instrument data file.

To assist in applying the data-entry software to the TIMSS 2003 data, the International Study Center provided each national research center with a *Manual for Entering the TIMSS 2003 Data* (TIMSS, 2002a) detailing prescribed procedures for data entry and verification. In addition, the *TIMSS 2003 Survey Operation Manual* (TIMSS, 2002b) included general instructions about the test administration and the data entry process.

The data manager at the TIMSS national center in each country gathered data from tracking forms that were used to record information on students selected to participate in the study, as well as their schools, and teachers. Tracking form related information was entered with the help of the WinW3S sampling software distributed by the DPC (see Chapter 6). The responses from the student achievement booklets as well as student, teacher, and school questionnaires were entered into computer data files created from the codebook templates. While strongly encouraged to use the WinDEM software for data entry, a few participating countries elected to use a different data

2 See Chapter 11 for details about scaling procedures.

entry system. However, they were required to conform to all specifications established in the international codebooks and to check their data with all consistency checks provided with the WinDEM software.

For each testing grade the following files were used during data entry:

- The WinW3S database contained sampling information as well as tracking form information (such as student's age, gender, and participation status) from all sampled students, teachers, and schools.
- The student background data file contained data from the *Student Background Questionnaire*. Additionally, these files contained tracking information for those countries, which did not use the WinW3S software.
- The student achievement data file contained the student's responses to whichever of the 12 test booklets was assigned to the student.
- In order to check the reliability of the constructed response item scoring, the constructed-response items were scored independently by a second scorer in a random sample of 100 of each test booklet type. The responses from these booklets were stored in a reliability scoring file.
- Because for eighth grade, separate *Mathematics Teacher* and *Science Teacher Questionnaires* were administered, two data files for the teachers' data were used, one for each questionnaire. For fourth grade, a single *Teacher Questionnaire* was administered, so data were entered into one teacher data file. For all countries not using WinW3S the data files also contained information from the teacher tracking forms.
- The school data file contained data from the *School Questionnaire*.

8.4 Data Checking and Editing at the National Centers

Before sending the data to the DPC for further data processing, countries were responsible for checking the data files with programs specifically prepared for TIMSS and for undertaking corrections as necessary. The first step was to apply the checking programs that are a feature of the WinDEM program. These checks are intended mainly to identify invalid data, but also can check the consistency between some basic variables. For example, an important feature of WinDEM is the ability to check that identification codes (IDs) are unique within a file. The WinDEM checks were mandatory for all countries. Additionally, after each file had been checked, the WinLINK program, which verifies the links between the various files, had to be applied. This software checks that the identification variables (student, teacher, class, and school identification codes) exist and match in related survey files. NRCs were required to resolve any problems identified by the within-country cleaning process before submitting data files to the IEA DPC.

8.5 Submitting Data Documentation to the IEA Data Processing Center

In addition to the data files described above, countries were requested to provide detailed data documentation to the IEA Data Processing Center. This included copies of all original survey tracking forms, copies of the national versions of test booklets and questionnaires, and a report of the survey activities. In order that all national adaptations to the survey instruments be documented, countries were required to submit *Data Management Forms and Cultural Adaptation Forms*.

Countries also were asked to send to the DPC the sample of test booklets selected for double-scoring the constructed-response items (around 1200 booklets altogether). The student responses to constructed-response items in these booklets will be digitally scanned and preserved for use in the next cycle of TIMSS in 2007, when they will be rescored by TIMSS 2007 scoring staff to monitor consistency in scoring practices between 2003 and 2007.

8.6 IEA DPC Quality Assurance Program

The IEA DPC went to great lengths to ensure that the data received from the TIMSS countries were of high quality and were internationally comparable. The foundation for quality assurance was laid before the first data arrived at the DPC through the provision to the TIMSS countries of software designed to standardize a range of operational and data-related tasks.

- The WinW3S software (IEA, 2002a) performed the within-school sampling operations adhering strictly to the sampling rules defined by Statistics Canada and the International Study Center. The software also created all necessary tracking forms and stored student- and teacher- specific tracking form information (such as student's age, gender, and participation status).
- The WinDEM program (IEA, 2002b) enabled key-entry of all TIMSS test and questionnaire data in a standard, internationally-defined format. The software also includes a range of checks for data verification.
- The WinLINK program (and LinkT03M, its DOS version) enabled NRCs to perform consistency checks on the identification variables across the TIMSS survey files.

A study as complex as TIMSS required a complex data cleaning design. To ensure that programs ran in the correct sequence, that no special requirements were overlooked, and that the cleaning process was implemented independently of the persons in charge, the following steps were undertaken:

- Before use with real data, all data-cleaning programs were thoroughly tested using simulated data sets containing all possible problems and inconsistencies.
- All incoming data and documents were registered in a specific database. The date of arrival was recorded, along with any specific issues meriting attention.
- The cleaning was organized following strict rules. Deviations in the cleaning sequence were not possible, and the scope for involuntary changes to the cleaning procedures was minimal.
- All corrections undertaken to country's data files were listed in a country-specific cleaning report.
- Occasionally it was necessary to make changes to a country's data files. Every such "manual" correction was logged using a specially-developed editing program (SAS-ManCorr), which recorded all changes and allowed DPC staff to undo changes, or to redo the whole manual cleaning process automatically at a later stage of the cleaning.
- Data Correction Software (DCS) was developed at the IEA DPC and distributed among the participating countries to assist them in identifying and correcting inconsistencies between variables in the background questionnaire files.
- Once the data-cleaning was completed for a country, all cleaning steps were repeated from the beginning to detect any problems that might have been inadvertently introduced during the cleaning process.
- All national adaptations that countries recorded in their documentation were verified against the structure of the national data files. All deviations from the international data structure that were detected were recorded in a "National Adaptation Database". This database is available for data analysts as an Appendix to the User's Guide to the TIMSS 2003 International Database.

8.7 Data Checking and Editing at the IEA Data Processing Center

Once the data were entered into data files at the National Research Center, the data files were submitted to the IEA Data Processing Center for checking and input into the international database. This process is generally referred to as data cleaning. The main objective of the process was to ensure that the data adhered to international formats, that school, teacher, and student information could be linked between different survey files, and that the data accurately and consistently reflected the information collected within each country.

The program-based data cleaning consisted of the following steps:

- Documentation and structure check
- Identification variable (ID) cleaning
- Linkage check
- Resolving inconsistencies in background questionnaire data

8.7.1 Documentation and Structure Check

For each country, data cleaning began with an exploration of its data file structures and a review of its data documentation: Data Management Forms, Student Tracking Forms, Class Sampling Forms, Teacher Tracking Forms, and Test Administration Forms. Most countries sent all required documentation along with their data, which greatly facilitated the data checking. The DPC contacted those countries for which documentation was incomplete and obtained all forms necessary to complete the documentation.

The first checks implemented at the DPC looked for differences between the international file structure and the national file structures. Some adaptations (such as adding national variables, or omitting or modifying international variables) were made to the background questionnaires in some countries. The extent and nature of such changes differed across the countries: some countries administered the questionnaires without any changes (apart from the translations), whereas other countries inserted items or options within existing international variables or added entirely new national variables. To keep track of any adaptations, NRCs were asked to complete Data Management Forms as they adapted the codebooks. Where necessary, the DPC modified the structure of the country's data to ensure that the resulting data remained comparable between countries.

As part of this standardization process, since direct correspondence between the data-collection instruments and the files was no longer necessary, the file structure was rearranged from a booklet-oriented model designed to facilitate data entry to an item-oriented layout more suited to data analysis. Variables created purely for verification purposes during data entry were dropped at this time, and provision was added for new variables necessary for analysis and reporting (i.e., reporting variables, derived variables, sampling weights, and achievement scores).

After each data file matched the international standard as specified in the international codebooks, a series of standard cleaning rules were applied to the files. This was conducted using software developed at the IEA DPC that could identify and in many cases correct inconsistencies in the data. Each problem was recorded in a database, identified by a unique problem number and with description of the problem and the action taken by the program or by the staff of the DPC.

Where problems could not be rectified automatically, they were reported to the responsible NRC so that original data-collection instruments and tracking forms could be checked to trace the source of the errors. Wherever possible, staff at the IEA Data Processing Center suggested a remedy, and asked the NRCs to either accept it or propose an alternative. Data files then were updated to reflect the solutions agreed on. Where the NRC could not solve problems by inspecting the instruments or forms, a general cleaning rule applying rectified these. After all automatic updates had been applied, remaining corrections to the data files were applied directly by keyboard, using a specially developed editing program (SAS-ManCorr).

8.7.2 Identification Variable (ID) Cleaning

Each record in a data file should have a unique identification number. The existence of records with duplicate ID numbers in a file implies an error of some kind. If two records share the same ID number, and contained exactly the same data, one of the records was deleted and the other remained in the database. If the records contained different data apart from the ID numbers, and it was impossible to identify which record contained the "true data," both records were removed from the database. The DPC tried to keep such losses at a minimum, and in only a few cases were data actually deleted.

The ID cleaning focused on the student background questionnaire file, because most of the critical variables were present in this file. Apart from the unique student ID number, there were variables pertaining to the students' participation and exclusion status – as well as dates of birth and dates of testing used to calculate age at the time of testing. The Student Tracking Forms[3] were essential in resolving any anomalies, as was close cooperation with NRCs (in most cases, the Student Tracking Forms were completed in the country's official language). The information about participation and exclusion was sent to Statistics Canada, where it was used to calculate students' participation rates, exclusion rates, and student sampling weights.

8.7.3 Linkage Check

In TIMSS, data about students and their schools and teachers appear in several files. It was crucial that the records from these files be linked together correctly to obtain meaningful results. The linkage was implemented through a hierarchical ID numbering system incorporating a school, class, and student component,[4] and is cross-checked against the tracking forms. The students'

3 Tracking Forms were used to record the sampling of schools, classes, teachers, and students. (see also chapter 6).

4 The ID number of a higher level is included in the ID number of a lower sampling level: the class ID includes the school ID, and the student ID includes the class ID (e.g., student 1220523 may be described as student 23 of class 05 in school 122).

entries in the achievement file and in the student background file must match one another; the reliability scoring file must represent a specific part of the achievement file; the teachers must be linked to the correct students; and the schools must be linked to the correct teachers and students.

8.7.4 Resolving inconsistencies in background questionnaire data

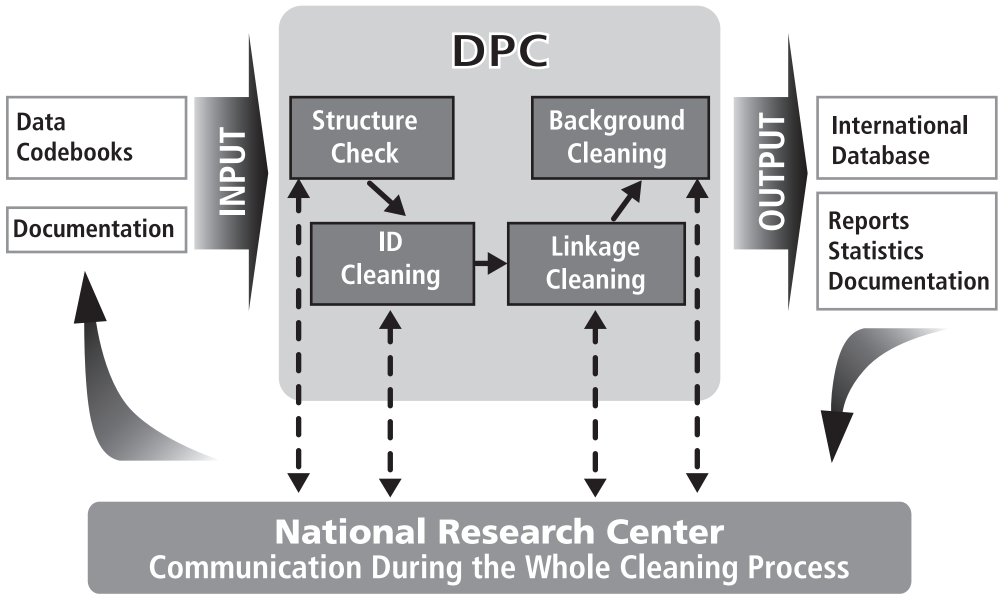
The number of inconsistent and implausible responses in background files varied from country to country, but no country's data were completely free of inconsistent responses. Treatment of these responses was determined on a question-by-question basis, using available documentation to make an informed decision. All background questionnaire data were checked for consistency among the responses given. For example, question number 2(a) in the School Questionnaire asked for the total school enrollment (number of students) in all grades, while 2(b) asked for the enrollment in the fourth grade only. Clearly, the number given for 2(b) should not exceed the number given for 2(a). All such inconsistencies that were detected were flagged and the NRCs asked to investigate. Those cases that could not be corrected or where the data made no sense were recoded to “Omitted”.

Filter questions, which appear in some questionnaires, were used to direct the respondent to a particular section of the questionnaire. Filter questions and the dependent questions that follow were subject to the following cleaning rules: If the answer to the filter question was “No” or “Not applicable” and yet the dependent questions were answered, then the filter question was recoded to "Yes" or “Applicable.”

Split variable checks were applied to questions where the answer was coded into several variables. For example question 5 in the Student Questionnaire listed a number of home possessions and asked the student to check all that applied. Student responses were captured in a series of 16 variables, each one coded as “Yes” if the corresponding possession was checked and “No” if left unchecked. Occasionally, students checked the “Yes” boxes but left the ”No” boxes unchecked, or missing. Since in these cases it was clear that the unchecked boxes actually meant “No,” these were recoded accordingly.

For further details about the standard cleaning procedures, please refer to the TIMSS *General Cleaning Documentation VII* (IEA, 2003).

Exhibit 8.1 **Overview Data Processing at the DPC**



8.7.5 National Cleaning Documentation

National Research Coordinators received a detailed report of all problems identified in their data, and of the steps applied to correct them. These included:

- Documentation of any data problems detected by the cleaning program and the steps applied to resolve them (*General Cleaning Documentation V11* (IEA, 2003))
- A record of all deviations from the international data-collection instruments and the international file structure

Additionally, the IEA DPC provided each NRC with revised data files incorporating all agreed edits, updates, and structural modifications. The revised files included a range of new variables that could be used for analytic purposes. For example, the student files included nationally standardized scores in mathematics and science that could be used in national analyses to be conducted before the international database became available.

8.7.6 Handling of Missing Data

When the TIMSS data were entered using WinDEM, two types of entries were possible: valid data values, and missing data values. Missing data can be assigned a value of omitted, not administered, or invalid during data entry.

At the IEA DPC, additional missing codes were applied to the data to be used for further analyses. In the international database, five missing codes are used:

- Not administered – the respondent was not administered the actual item. He or she had no chance to read and answer the question (assigned both during data entry and data processing).
- Omitted – the respondent had a chance to answer the question, but did not do so (assigned both during data entry and data processing).
- Logically not applicable – the respondent answered a preceding filter question in a way that made the following dependent questions not applicable to him or her (assigned during data processing only).
- Not reached (only used in the achievement files) – this code indicates those items not reached by the students, due to a lack of time (assigned during data processing only).
- Not interpretable (only used in the achievement files) – this code was used for multiple-choice items that were answered, but the chosen answer options were not clear, as well as for constructed-response items where the scorer assigned two or more scores (assigned during data entry and data processing).

8.8 Data Products

Data products sent by the IEA Data Processing Center to NRCs included both data almanacs and data files.

8.8.1 Data Almanacs and Item Statistics

Each country received a set of data almanacs, or summaries, produced by the TIMSS & PIRLS International Study Center. These contained weighted summary statistics for each participating country on each variable included in the survey instruments. The data almanacs were sent to the participating countries for review. When necessary, they were accompanied by specific questions about the data presented in them. They were also used by the International Study Center during the data review and in the production of the reporting exhibits.

Each country also received a set of preliminary national item statistics and reliability statistic reports for review purposes. The item statistics contained summary information about items characteristics, such as the classical item difficulty index, the classical item discrimination index, the Rasch item difficulty, and the Rasch mean square fit index. The reliability statistics contained summary statistics about the percent of agreement between scorers on the score assigned to the items.

8.8.2 Versions of the National Data Files

Building the international database was an iterative process. The IEA Data Processing Center provided NRCs with a new version of their country's data files whenever a major step in data processing was completed. This also guaranteed that the NRCs had a chance to review their data and run their own checks to validate the data files. Before the TIMSS international database was published, three versions of the data files were sent to each country. Each country received its own data only. The first version was sent as soon as the data could be regarded as 'clean' concerning identification codes and linkage issues. These first files contained nationally standardized achievement scores calculated by the Data Processing Center using a Rasch-based scaling method. Documentation, with a list of the cleaning checks and corrections made in the data, was included to enable the NRC to review the cleaning process. A second version of the data files was sent to the NRCs when the weights and the international achievement scores were available and had been merged to the files. A third version was sent together with the data almanacs after all exhibits of the TIMSS International Report have been verified and final updates to the data files had been implemented, to enable the NRCs to validate the results presented in the first international reports.

8.8.3 The International Database

The international database incorporated all national data files. Data processing at the DPC ensured that:

- Information coded in each variable was internationally comparable
- National adaptations were reflected appropriately in all variables
- Questions that are not internationally comparable were removed from the database
- All entries in the database could be linked to the appropriate respondent – student, teacher, or principal.
- Sampling weights and student achievement scores were available for international comparisons

In a joint effort of the IEA DPC and the TIMSS & PIRLS International Study Center at Boston College, a National Adaptations Database containing all adaptations to questionnaires made by individual countries, and documenting how they were handled, was constructed. The meaning of country specific items can also be found in this database, as well as recoding requirements by the International Study Center. Information contained in this database was provided in the User Guide for the international database upon release of the TIMSS 2003 data.

The TIMSS 2003 international database is a unique resource for policy makers and analysts, containing student mathematics and science achievement and background data from representative samples of fourth and eighth grade students from 49 countries and four Benchmarking Particpants.

References:

IEA (2002a). *WinW3S: Within-School Sampling Software*. Hamburg: IEA Data Processing Center.

IEA (2002b). *WinDEM: Software for data entry and verification*. Hamburg: IEA Data Processing Center.

IEA (2003): *General Cleaning Documentation V11*. Hamburg: IEA Data Processing Center

TIMSS (2002a), *Manual for Entering the TIMSS 2003 Data*, prepared by the International Study Center in cooperation with the IEA Data Processing Center, Chestnut Hill, MA: Boston College.

TIMSS (2002b), *TIMSS 2003 Survey Operations Manual*, prepared by the International Study Center, Chestnut Hill, MA: Boston College.

9

Chapter 9

TIMSS 2003 Sampling Weights and Participation Rates

Marc Joncas

9.1 Overview

As described in Chapter 5, TIMSS uses rigorous sampling of schools and students to provide valid and efficient estimates of mathematics and science achievement in the fourth- and eighth- grade student populations of participating countries. The accuracy of these estimates depends to a great extent on the quality of the sampling in each country, which in turn is determined by the quality of the sampling information available in designing the sampling plan and the care with which the sampling activities are conducted. For TIMSS 2003, National Research Coordinators (NRCs) worked on all phases of sampling, in conjunction with staff from Statistics Canada and the IEA Data Processing Centre (DPC). NRCs were trained in how to select the school and student samples, and in how to use the sampling software provided by the IEA Data Processing Centre. This chapter summarizes major characteristics of the national samples, and describes the procedure for computing sampling weights and participation rates for each country. In consultation with the TIMSS 2003 sampling referee[1], staff from Statistics Canada and the IEA DPC reviewed the national sampling plans, sampling data, sampling frames, and sample selection. The TIMSS & PIRLS International Study Centre (ISC) at Boston College, jointly with Statistics Canada, the IEA DPC and the sampling referee, used this information to evaluate the quality of the samples. Summaries of the sample design for each country, including details of population coverage and exclusions, stratification variables, and participation rates, are provided in Appendix B.

1 Keith Rust, Westat.

9.2 Sampling implementation

9.2.1 TIMSS 2003 Target Populations

In IEA studies, the target population for all countries is known as the international desired population. The *international desired populations* for TIMSS 2003 were defined as:

Population 1: All students enrolled in the upper of the two adjacent grades that contain the largest proportion of 9-year-olds at the time of testing. This grade level was intended to represent four years of schooling, counting from the first year of primary or elementary schooling, and was the fourth grade in most countries.

Population 2: 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. This grade level was intended to represent eight years of schooling, counting from the first year of primary or elementary schooling, and was the eighth grade in most countries.

To measure trends in student achievement, the TIMSS 2003 eighth- and fourth-grade target populations were intended to correspond to the upper grades of the TIMSS 1995 population definitions, and the TIMSS 2003 eighth-grade target population to the eighth-grade population in TIMSS 1999.

Exhibits 9.1 and 9.2 summarize the grades identified as the target grades for sampling in all participating countries and Benchmarking entities for the eighth and fourth grades, respectively. For most countries, the target grades did indeed turn out to be the grades with eight and four years of schooling. A number of countries decided to target the eighth or fourth grades even though their students were somewhat older as a result. These included Botswana, Estonia, Ghana, Latvia, Morocco, Romania, and South Africa at the eighth grade and Latvia, Moldova, Morocco, and Yemen at the fourth grade.

Exhibit 9.1 **National Grade Definitions – Eighth Grade**

Country	Country's Name for Grade Tested	Years of Formal Schooling	Mean Age of Students Tested
Armenia	Grade 8	8	14.9
Australia	Year 8	8 or 9	13.9
Bahrain	Second Intermediate	8	14.1
Belgium (Flemish)	2nd Grade of Secondary Education	8	14.1
Botswana	Grade 8 (Form 1)	8	15.1
Bulgaria	Grade 8	8	14.9
Chile	Eighth Grade of Basic Education	8	14.2
Chinese Taipei	2nd Grade Junior High School	8	14.2
Cyprus	2nd Grade Gymnasium	8	13.8
Egypt	Preparatory 3	8	14.4
England	Year 9	9	14.3
Estonia	Grade 8	8	15.2
Ghana	Junior Secondary School II (JSS II)	8	15.5
Hong Kong, SAR	Secondary 2 (S2)	8	14.4
Hungary	Grade 8	8	14.5
Indonesia	2nd Grade Junior Secondary School	8	14.5
Iran, Islamic Rep. of	Third Grade of Guidance School	8	14.4
Israel	Grade 8	8	14.0
Italy	Grade 8 (III Media)	8	13.9
Japan	2nd Grade Lower Secondary School	8	14.4
Jordan	Grade 8	8	13.9
Korea, Rep. of	2nd Grade Middle School	8	14.6
Latvia	Grade 8	8	15.0
Lebanon	Grade 8	8	14.6
Lithuania	Grade 8	8	14.9
Macedonia, Rep. of	Grade 8	8	14.6
Malaysia	Form 2	8	14.3
Moldova, Rep. of	Grade VIII	8	14.9
Morocco	2nd Secondary	8	15.2
Netherlands	Grade 8	8	14.3
New Zealand	Year 9	8.5 - 9.5	14.1
Norway	Grade 8 (these students started in Grade 2)	7	13.8
Palestinian Nat'l Auth.	Grade 8	8	14.1
Philippines	2nd Year High School	8	14.8
Romania	Grade 8	8	15.0
Russian Federation	Grade 8	7 or 8	14.2
Saudi Arabia	2nd Year of Middle School	8	14.1

Exhibit 9.1 **National Grade Definitions – Eighth Grade** *(...Continued)*

Country	Country's Name for Grade Tested	Years of Formal Schooling	Mean Age of Students Tested
Scotland	Secondary 2 (S2)	9	13.7
Serbia	8th grade of Primary School	8	14.9
Singapore	Secondary 2	8	14.3
Slovak Republic	Grade 8	8	14.3
Slovenia	Grade 7 of 8-year elementary school, Grade 8 of 9-year elementary school	7 or 8	13.8
South Africa	Grade 8	8	15.1
Sweden	Year 8	8	14.9
Syrian Arab Republic	Grade 8	8	14.0
Tunisia	8th year of basic school	8	14.8
United States	Grade 8	8	14.2
Benchmarking Participants			
Basque Country, Spain	2nd Course of ESO	8	14.1
Indiana State, US	Grade 8	8	13.5
Ontario Province, Can.	Grade 8	8	13.8
Quebec Province, Can.	Secondary II	8	14.2

9.2.2 Population Coverage and Exclusions

Exhibit 9.3 and 9.4 summarize population coverage and exclusions for the TIMSS 2003 target populations. National coverage of the international desired target population was generally comprehensive. For example, at the eighth grade as shown in Exhibit 9.3, all but Indonesia, Lithuania, Morocco and Serbia sampled from 100% of their international desired population.[2] Since coverage was below 100% of the international desired population, the results for these countries were footnoted in the TIMSS 2003 international reports to reflect this. At fourth grade (Exhibit 9.4), only Lithuania chose a national desired population less than the international desired population[3]. Since coverage was below 100%, the Lithuanian fourth-grade results were footnoted in the international reports.

Within the national desired population, it was possible to exclude certain school types, such as very small or very remote schools, and certain types of students, such as those with a disability that prevented them from participating in the assessment. For most part, school-level exclusions consisted of schools for the disabled and very small schools; however, there were some exceptions that are documented in Appendix B. Within-school exclu-

2 The Indonesian population included Non-Islamic schools only, the Lithuanian population included schools catering to Lithuanian-speaking student only, Morocco included schools from all provinces except Souss Massa Draa, Casablanca and Gharb-Chrarda, and Serbia included schools from all provinces except Kosovo.

3 The Lithuanian population was restricted to schools catering to Lithuanian-speaking student only.

sions generally consisted of disabled students and students who could not be assessed in the language of the test. At fourth grade, the percentage of excluded students was less than 10% in every country, and at eighth grade only in Israel and Macedonia did the level of excluded students exceed this figure. Results for these countries were annotated in the international reports. A few countries had no within-school exclusions.

Exhibit 9. 2 **National Grade Definitions – Fourth Grade**

Country	Country's Name for Grade Tested	Years of Formal Schooling	Mean Age of Students Tested
Armenia	Grade 4	4	10.9
Australia	Year 4	4	9.9
Belgium (Flemish)	Grade 4 primary education	4	10.0
Chinese Taipei	Elementary School, Grade 4	4	10.2
Cyprus	4th grade Primary	4	9.9
England	Year 5	5	10.3
Hong Kong, SAR	Primary 4 (P4)	4	10.2
Hungary	Grade 4	4	10.5
Iran, Islamic Rep. of	4th Grade of Primary School	4	10.4
Italy	Grade 4 (IV Elementare)	4	9.8
Japan	4th Grade at the Elementary School	4	10.4
Latvia	Grade 4	4	11.1
Lithuania	Grade 4	4	10.9
Moldova, Rep. of	Grade IV	4	11.0
Morocco	Grade 4 Primary	4	11.0
Netherlands	Grade 4	4	10.2
New Zealand	Year 5	4.5 - 5.5	10.0
Norway	Grade 4	3	9.8
Philippines	Grade 4	4	10.8
Russian Federation	Fourth grade for 4-year primary school; Third grade for 3-year primary school	3 or 4	10.6
Scotland	Primary 5 (P5)	5	9.7
Singapore	Primary 4	4	10.3
Slovenia	Grade 3 of 8-year elementary school; Grade 4 of 9-year elementary school	3 or 4	9.8
Tunisia	4th year of basic school	4	10.4
United States	Grade 4	4	10.2
Yemen	Grade 4	4	11.0
Benchmarking Participants			
Indiana State, US	Grade 4	4	9.5
Ontario Province, Can.	Grade 4	4	9.8
Quebec Province, Can.	2nd Year of 2nd Cycle	4	10.1

Within the national desired population, it was possible to exclude certain school types, such as very small or very remote schools, and certain types of students, such as those with a disability that prevented them from participating in the assessment. For most part, school-level exclusions consisted of schools for the disabled and very small schools; however, there were some exceptions that are documented in Appendix B. Within-school exclusions generally consisted of disabled students and students who could not be assessed in the language of the test. At fourth grade, the percentage of excluded students was less than 10% in every country, and at eighth grade only in Israel and Macedonia did the level of excluded students exceed this figure. Results for these countries were annotated in the international reports. A few countries had no within-school exclusions.

9.2.3 General Sample design

The basic design of the sample used in TIMSS 2003 was a two-stage stratified cluster design.[4] The first stage consisted of a sample of schools, and the second stage of a sample of intact classrooms (usually mathematics classes) from the target grades in the sampled schools. Countries could, with approval from the sampling consultants, adapt the basic design to their particular situation. For example, the Russian Federation introduced an extra stage where regions were sampled first, and then schools sampled from within the sampled regions, and in Egypt, Morocco, Singapore, South Africa and Yemen, student sub-sampling occurred within sampled classrooms.

The TIMSS 2003 design allowed countries to stratify the school sampling frame in order to improve the precision of survey results. Countries could use an explicit stratification procedure, by which schools were categorized according to some criterion (e.g., regions of the country), ensuring a predetermined number of schools would be selected from each stratum. Countries could also use an implicit stratification procedure, by which schools were sorted according to a set of stratification variables prior to sampling. This approach provided an efficient method of allocating the school sample in proportion to the size of the implicit stratum, when used in conjunction with a systematic probability-proportional-to-size (PPS) sampling method. Stratification variables and procedures for each country are described in Appendix B.

4 The TIMSS 2003 sample design is described in Chapter 5.

Exhibit 9. 3 **National Coverage and Overall Exclusion Rates – Eighth Grade**

Country	International Desired Population		National Desired Population		
	Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
Armenia	100%		2.9%	0.0%	2.9%
Australia	100%		0.4%	0.9%	1.3%
Bahrain	100%		0.0%	0.0%	0.0%
Belgium (Flemish)	100%		3.1%	0.1%	3.2%
Botswana	100%		0.8%	2.2%	3.0%
Bulgaria	100%		0.5%	0.0%	0.5%
Chile	100%		1.6%	0.7%	2.2%
Chinese Taipei	100%		0.2%	4.6%	4.8%
Cyprus	100%		1.1%	1.5%	2.5%
Egypt	100%		3.4%	0.0%	3.4%
England	100%		2.1%	0.0%	2.1%
Estonia	100%		2.6%	0.8%	3.4%
Ghana	100%		0.9%	0.0%	0.9%
Hong Kong, SAR	100%		3.3%	0.1%	3.4%
Hungary	100%		5.5%	3.2%	8.5%
Indonesia	80%	Non-islamic schools	0.1%	0.3%	0.4%
Iran, Islamic Rep. of	100%		5.5%	1.1%	6.5%
Israel	100%		15.2%	8.6%	22.5%
Italy	100%		0.0%	3.6%	3.6%
Japan	100%		0.5%	0.1%	0.6%
Jordan	100%		0.5%	0.8%	1.3%
Korea, Rep. of	100%		1.5%	3.4%	4.9%
Latvia	100%		3.6%	0.1%	3.7%
Lebanon	100%		1.4%	0.0%	1.4%
Lithuania	89%	Students taught in Lithuanian	1.4%	1.2%	2.6%
Macedonia, Rep. of	100%		12.5%	0.0%	12.5%
Malaysia	100%		4.0%	0.0%	4.0%
Moldova, Rep. of	100%		0.7%	0.5%	1.2%
Morocco	69%	All students but Souss Massa Draa, Casablanca, Gharb-Chrarda	1.5%	0.0%	1.5%
Netherlands	100%		3.0%	0.0%	3.0%
New Zealand	100%		1.7%	2.7%	4.4%
Norway	100%		0.9%	1.5%	2.3%
Palestinian Nat'l Auth.	100%		0.2%	0.3%	0.5%

Exhibit 9. 3 **National Coverage and Overall Exclusion Rates – Eighth Grade** *(...Continued)*

Country	International Desired Population		National Desired Population		
	Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
Philippines	100%		1.5%	0.0%	1.5%
Romania	100%		0.4%	0.1%	0.5%
Russian Federation	100%		1.7%	3.9%	5.5%
Saudi Arabia	100%		0.3%	0.2%	0.5%
Scotland	100%		0.0%	0.0%	0.0%
Serbia	81%	Serbia without Kosovo	2.4%	0.6%	2.9%
Singapore	100%		0.0%	0.0%	0.0%
Slovak Republic	100%		5.0%	0.0%	5.0%
Slovenia	100%		1.3%	0.1%	1.4%
South Africa	100%		0.6%	0.0%	0.6%
Sweden	100%		0.3%	2.5%	2.8%
Syrian Arab Republic	100%		18.7%	0.0%	18.8%
Tunisia	100%		1.8%	0.0%	1.8%
United States	100%		0.0%	4.9%	4.9%
Benchmarking Participants					
Basque Country, Spain	100%		2.1%	3.8%	5.8%
Indiana State, US	100%		0.0%	7.8%	7.8%
Ontario Province, Can.	100%		1.0%	5.0%	6.0%
Quebec Province, Can.	100%		1.4%	3.5%	4.8%

Exhibit 9. 4 **National Coverage and Overall Exclusion Rates – Fourth Grade**

Country	International Desired Population		National Desired Population		
	Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
Armenia	100%		2.9%	0.0%	2.9%
Australia	100%		1.2%	1.6%	2.7%
Belgium (Flemish)	100%		5.9%	0.4%	6.3%
Chinese Taipei	100%		0.3%	2.8%	3.1%
Cyprus	100%		1.5%	1.4%	2.9%
England	100%		1.9%	0.0%	1.9%
Hong Kong, SAR	100%		3.7%	0.1%	3.8%
Hungary	100%		4.4%	3.9%	8.1%
Iran, Islamic Rep. of	100%		3.6%	2.1%	5.7%
Italy	100%		0.1%	4.1%	4.2%
Japan	100%		0.4%	0.3%	0.8%
Latvia	100%		4.3%	0.1%	4.4%
Lithuania	92%	Students taught in Lithuanian	2.1%	2.6%	4.6%
Moldova, Rep. of	100%		2.0%	1.6%	3.6%
Morocco	100%		2.2%	0.0%	2.2%
Netherlands	100%		4.1%	1.1%	5.2%
New Zealand	100%		1.5%	2.5%	4.0%
Norway	100%		1.7%	2.7%	4.4%
Philippines	100%		3.8%	0.7%	4.5%
Russian Federation	100%		2.2%	4.7%	6.8%
Scotland	100%		1.5%	0.0%	1.5%
Singapore	100%		0.0%	0.0%	0.0%
Slovenia	100%		0.8%	0.5%	1.3%
Tunisia	100%		0.9%	0.0%	0.9%
United States	100%		0.0%	5.1%	5.1%
Yemen	100%		0.6%	8.9%	9.5%
Benchmarking Participants					
Indiana State, US	100%		0.0%	7.2%	7.2%
Ontario Province, Can.	100%		1.3%	3.5%	4.8%
Quebec Province, Can.	100%		2.7%	0.9%	3.6%

Most countries sampled 150 schools and one intact classroom (i.e., including all of its students) within each school. Classrooms within schools generally were selected with equal probabilities. However, as described above, some countries where large classrooms are the norm sampled students within classrooms was a means of reducing the data collection effort. In these cases, classrooms were sampled with PPS, and then a fixed number of students (with equal probabilities) were sampled from within the sampled classrooms. With the approval of the sampling consultants, several countries chose to sample more than one classroom from each sampled school. Details of the sampling of schools and students for each country are provided in Appendix B

The TIMSS 2003 sample designs were implemented in an acceptable manner by all participating countries except Yemen and the Syrian Arab Republic. Both adopted classroom sampling procedures that did not meet the TIMSS sampling standards and so could not be approved by the International Study Centre. As a result, data for these two countries were summarized in an appendix to the international reports.

9.2.4 Target Population Sizes

Exhibits 9.5 and 9.6 summarize for eighth and fourth grade, respectively, the number of schools and students in each country's target populations, as well as the number of sampled schools and students that participated in the study. The population figures for schools and students were derived from the sampling frames that countries used to draw their TIMSS samples.[5] As a check on the sampling procedure, TIMSS used the sampling weights computed for each country (see Section 9.3) to derive an estimate of the student population size. In most cases, the estimated population size closely matched the actual population size from the sampling frame, as shown in Exhibits 9.5 and 9.6.

5 The school and student population sizes for Russian Federation, however, were not computed from the sampling frame, but were provided by the NRC.

Exhibit 9. 5 **Population and Sample Sizes – Eighth Grade**

Country	Population		Sample			Mean Age of Students Tested
	Schools	Students	Schools	Students	Est. Pop.	
Armenia	1,439	56,841	149	5,726	54,502	14.9
Australia	2,297	253,522	207	4,791	257,407	13.9
Bahrain	67	10,581	67	4,199	10,543	14.1
Belgium (Flemish)	1,084	70,204	148	4,970	70,637	14.1
Botswana	215	37,975	146	5,150	36,142	15.1
Bulgaria	2,360	83,202	164	4,117	87,603	14.9
Chile	5,165	286,050	195	6,377	265,749	14.2
Chinese Taipei	863	318,196	150	5,379	297,842	14.2
Cyprus	59	9,700	59	4,002	9,231	13.8
Egypt	7,586	1,503,480	217	7,095	1,365,244	14.4
England	3,912	615,535	87	2,830	662,049	14.3
Estonia	517	21,419	151	4,040	20,995	15.2
Ghana	6,533	280,912	150	5,100	276,427	15.5
Hong Kong, SAR	423	84,898	125	4,972	82,693	14.4
Hungary	2,563	114,364	155	3,302	100,609	14.5
Indonesia	19,864	2,836,390	150	5,762	2,318,021	14.5
Iran, Islamic Rep. of	22,227	1,639,906	181	4,942	1,369,991	14.4
Israel	816	110,284	146	4,318	85,689	14.0
Italy	5,778	591,400	171	4,278	567,587	13.9
Japan	10,859	1,298,927	146	4,856	1,269,256	14.4
Jordan	1,676	106,875	140	4,489	96,297	13.9
Korea, Rep. of	2,593	610,271	149	5,309	570,771	14.6
Latvia	831	33,255	140	3,630	33,708	15.0
Lebanon	1,567	56,689	152	3,814	57,789	14.6
Lithuania	1,077	54,081	143	4,964	46,940	14.9
Macedonia, Rep. of	338	30,814	149	3,893	25,963	14.6
Malaysia	1,641	435,722	150	5,314	414,259	14.3
Moldova, Rep. of	1,352	61,158	149	4,033	61,669	14.9
Morocco	1,371	387,115	131	2,943	209,164	15.2
Netherlands	1,109	198,171	130	3,065	188,992	14.3
New Zealand	407	57,454	169	3,801	57,392	14.1
Norway	1,076	55,559	138	4,133	61,222	13.8
Palestinian Nat'l Auth.	872	69,210	145	5,357	64,860	14.1
Philippines	7,073	1,393,428	137	6,917	1,395,144	14.8
Romania	7,324	316,441	148	4,104	294,631	15.0

Exhibit 9. 5 **Population and Sample Sizes – Eighth Grade** *(...Continued)*

Country	Population		Sample			Mean Age of Students Tested
	Schools	Students	Schools	Students	Est. Pop.	
Russian Federation	58,595	2,081,919	214	4,667	1,923,173	14.2
Saudi Arabia	6,224	355,676	155	4,295	326,754	14.1
Scotland	425	63,795	128	3,516	58,824	13.7
Serbia	1,100	92,261	149	4,296	87,330	14.9
Singapore	164	53,100	164	6,018	53,292	14.3
Slovak Republic	1,646	85,465	179	4,215	75,718	14.3
Slovenia	444	24,637	174	3,578	22,972	13.8
South Africa	8,926	1,009,215	255	8,952	783,951	15.1
Sweden	1,467	110,121	159	4,256	108,760	14.9
Syrian Arab Republic	1,687	243,356	134	4,895	201,972	14.0
Tunisia	740	196,012	150	4,931	184,104	14.8
United States	45,472	3,911,458	232	8,912	3,447,236	14.2
Benchmarking Participants						
Basque Country, Spain	448	16,803	120	2,514	18,710	14.1
Indiana State, US	937	84,499	54	2,188	76,051	13.5
Ontario Province, Can.	2,919	144,603	186	4,217	145,430	13.8
Quebec Province, Can.	639	91,687	175	4,411	82,209	14.2

Exhibit 9. 6 **Population and Sample Sizes – Fourth Grade**

Country	Population		Sample			Mean Age of Students Tested
	Schools	Students	Schools	Students	Est. Pop.	
Armenia	1,439	56,841	148	5,674	51,844	10.9
Australia	6,779	263,710	204	4,321	257,221	9.9
Belgium (Flemish)	2,154	73,232	149	4,712	66,236	10.0
Chinese Taipei	2,436	318,173	150	4,661	311,390	10.2
Cyprus	256	10,322	150	4,328	9,946	9.9
England	15,341	646,863	123	3,585	588,366	10.3
Hong Kong, SAR	756	85,364	132	4,608	79,039	10.2
Hungary	2,563	116,580	157	3,319	101,631	10.5
Iran, Islamic Rep. of	47,274	1,668,358	171	4,352	1,322,801	10.4
Italy	7,504	555,270	171	4,282	513,655	9.8
Japan	20,256	1,185,936	150	4,535	1,172,766	10.4
Latvia	890	34,775	140	3,687	29,607	11.1
Lithuania	1,554	52,679	153	4,422	45,123	10.9
Moldova, Rep. of	1,425	58,467	151	3,981	56,649	11.0
Morocco	14,219	567,743	197	4,264	632,376	11.0
Netherlands	6,668	198,775	130	2,937	170,068	10.2
New Zealand	1,944	60,410	220	4,308	59,301	10.0
Norway	2,330	62,344	139	4,342	60,354	9.8
Philippines	34,127	2,040,230	135	4,572	1,805,303	10.8
Russian Federation	63,641	1,312,450	205	3,963	1,138,069	10.6
Scotland	1,870	63,879	125	3,936	56,191	9.7
Singapore	182	49,900	182	6,668	49,994	10.3
Slovenia	444	19,826	174	3,126	18,750	9.8
Tunisia	3,944	222,537	150	4,334	216,491	10.4
United States	71,863	4,143,117	248	9,829	3,518,039	10.2
Yemen	5,748	526,954	150	4,205	445,965	11.0
Benchmarking Participants						
Indiana State, US	1,675	88,487	56	2,233	80,151	9.5
Ontario Province, Can.	3,770	153,625	189	4,362	142,180	9.8
Quebec Province, Can.	1,879	98,326	193	4,350	85,895	10.1

9.3 Calculating Sampling Weights

While the TIMSS 2003 multistage stratified cluster design provided very economical and effective data collection in a school environment, it resulted in differential probabilities of selection of the students. Individual country designs could be quite complex, as may be seen from Appendix B showing how the design was implemented in each country. To adjust for these differential selection probabilities and ensure accurate survey estimates, TIMSS 2003 computed a sampling weight for each participant student. Because appropriate sampling weights were essential for the computation of accurate survey results, the capacity to provide proper sampling weights was an essential requirement of an acceptable sample design. This section describes the procedures for calculating sampling weights for the TIMSS 2003 data.

Sampling weights were calculated according to a three-step procedure involving selection probabilities for schools, classrooms, and students. The first step consisted of calculating a school weight, which also incorporated weighting factors from any additional front-end sampling stages such regions. A school-level participation adjustment was then made in the school weight to compensate for any sampled schools that did not participate. That adjustment was calculated independently for each explicit stratum.

In the second step, a classroom weight reflecting the probability of the sampled classroom(s) being selected from among all the classrooms in the school at the target grade level was calculated. This classroom weight was calculated independently for each school. A classroom-level participation adjustment was then made in the class weight to compensate for any sampled classrooms that did not participate, or for classrooms where the participation rate among students fell below 50 percent. This participation adjustment was set to unity in cases where a single classroom was sampled in each school. If a school agreed to take part in the study but the classroom (i.e., the classroom teacher) refused to participate, adjustment for non-participation was made at the school level. If one of two (or more) selected classrooms in a school did not participate, the classroom participation adjustment was calculated for that school, independently for each explicit stratum.

The third and final step consisted of calculating a student weight. For most countries, because intact classrooms were sampled, each student in the sampled classrooms was certain of selection, and so the student weight was 1.0. When students were further sampled within classrooms, a student weight reflecting the probability of being sampled from the classroom was calculated. A non-participation adjustment was then made to compensate for students who did not take part in the testing. This was calculated independently for each sampled classroom.

The basic sampling weight attached to each student record was the product of the three weights described above: the first stage (school) weight, the second stage (classroom) weight, and the third stage (student) weight. The overall student sampling weight was the product of the three weights including non-participation adjustments.

9.3.1 The First Stage (School) Weight

Essentially, the first stage weight represented the inverse of the probability of a school being sampled at the first stage. The TIMSS 2003 sample design required that school selection probabilities be proportional to the school size, generally defined as enrolment in the target grade. The basic first stage weight for the i^{th} sampled school was thus defined as:

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

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

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

where N was the total number of schools in the explicit stratum.

For countries such as the Russian Federation that included region as a preliminary sampling step, the basic first stage weight also incorporated the probability of selection in this stage. The first stage weight in this case was simply the product of the "region" weight and the first stage weight, as described above.

In some countries, schools were selected with equal probabilities. This generally occurred when a large sampling ratio was used. In some countries also, explicit or implicit strata were defined to deal with very large schools or small schools. Equal probability sampling was necessary in these strata.

Under equal probability sampling, the basic first stage weight for the i^{th} sampled school was defined as

$$BW_{sc}^{i} = \frac{N}{n}$$

where n was the number of sampled schools and N was the total number of schools in the explicit stratum. The basic weight for all sampled schools in a stratum was identical in this context.

9.3.2 School Non-Participation Adjustment

First stage weights were calculated for all sampled and replacement schools that participated. A school-level participation adjustment was applied to compensate for schools that were sampled but did not participate, and were not replaced. Sampled schools that were found to be ineligible[6] were removed from the calculation of this adjustment. The school-level participation adjustment was calculated separately for each explicit stratum for all participants except England at the eighth grade.[7]

The adjustment was calculated as follows:

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

where n_s was the number of originally sampled schools that participated, n_{r1} and n_{r2} the number of first and second replacement schools, respectively, that participated, and n_{nr} the number of schools that did not participate.

The final first stage weight for the i^{th} school, corrected for non-participating schools, thus became:

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

9.3.3 The Second Stage (Classroom) Weight

The second stage weight represented the inverse of the probability of a classroom within a sampled school being selected. Although most countries sampled classrooms within schools with equal probability, when student subsampling was involved, countries had to sample classrooms using PPS techniques. Procedures for calculating sampling weights are presented below for both approaches.

Equal Probability Weighting: For the i^{th} school, let C^i be the total number of classrooms and c^i the number of sampled classrooms in the study. Using equal probability sampling, the basic second stage weight assigned to all sampled classrooms in the i^{th} school was:

$$BW_{cl1}^{i} = \frac{C^i}{c^i}$$

For most countries, c^i took the values 1, 2 or 3. Some countries sampled all classrooms in a selected school.

6 A sampled school was ineligible if it was found to contain no eligible (i.e. eighth- or fourth-grade students). Such schools usually were in the sampling frame by mistake, or schools that had recently closed.

7 The sampling plan for England included implicit stratification of schools by a measure of school academic performance. Because the school participation rate even after including replacement schools was relatively low (54%), it was decided to apply the school non-participation adjustment separately for each implicit stratum. Since the measure of academic performance used for stratification was strongly related to average school mathematics and science achievement on TIMSS, this served to reduce the potential for bias introduced by low school participation.

Probability Proportional to Size Weighting: For the i^{th} school, let $k^{i,j}$ be the size of the j^{th} classroom. Using PPS sampling, the final second stage weight assigned to the j^{th} sampled classroom in the i^{th} school was

$$BW_{cl2}^{i,j} = \frac{K^i}{c^i \cdot k^{i,j}}$$

where c^i was the number of sampled classrooms in the i^{th} school, as defined earlier, and

$$K^i = \sum_{j=1}^{c^i} k^{i,j}$$

For most countries, c^i took the values 1 or 2. Some countries sampled all classrooms in a selected school.

9.3.4 Classroom Non-Participation Adjustment

Second stage weights were calculated for all sampled classrooms in the sampled schools and replacement schools that participated. A classroom-level participation adjustment was applied to compensate for classrooms that did not participate or where student participation rate was below 50 percent. Sampled classrooms with student participation below 50 percent were given a weight of zero and considered to be non-participating. The classroom-level participation adjustment was calculated separately for each explicit stratum.

The adjustment was calculated as follows:

$$A_{cl} = \frac{\sum_{i}^{s+r1+r2} c^i}{\sum_{i}^{s+r1+r2} c_*^i}$$

where c^i was the number of sampled classrooms in the i^{th} school, as defined earlier, and c_*^i was the number of sampled classrooms in the i^{th} school that participated.

When no subsampling of classrooms was involved, the final second stage weight assigned to all sampled classrooms in the i^{th} school became:

$$FW_{cl1}^{i} = A_{cl} \cdot BW_{cl1}^{i}$$

When classrooms were subsampled within schools, the final second stage weight assigned to the j^{th} sampled classroom in the i^{th} school became:

$$FW_{cl2}^{i,j} = A_{cl} \cdot BW_{cl2}^{i,j}$$

9.3.5 The Third Stage (Student) Weight

The third stage weight represented the inverse of the probability of a student in a sampled class being selected. Where intact classrooms that included all students were sampled, as was the case in most participating countries, this probability was unity. However, the probability of selection varied when students were sampled within classrooms. Procedures for calculating weights are presented below for both sampling approaches. The third stage weight is calculated independently for each sampled classroom.

Sampling Intact Classrooms: The basic third stage weight for the j^{th} classroom in the i^{th} school was simply:

$$BW_{st1}^{i,j} = 1.0$$

Subsampling Students: The basic third stage weight for the j^{th} classroom in the i^{th} school was :

$$BW_{st2}^{i,j} = \frac{k^{i,j}}{s^{i,j}}$$

where $k^{i,j}$ was the size of the j^{th} classroom in the i^{th} school, as defined earlier, and $s^{i,j}$ was the number of sampled students per sampled classroom. The latter number usually remained constant for all sampled classrooms.

9.3.6 Adjustment for Student Non-Participation

The student non-participation adjustment was calculated for each participating classroom as follows:

$$A_{st}^{i,j} = \frac{s_{rs}^{i,j} + s_{nr}^{i,j}}{s_{rs}^{i,j}}$$

where $s_{rs}^{i,j}$ was the number of eligible students that participated in the j^{th} classroom of the i^{th} school and $s_{nr}^{i,j}$ was the number of eligible students that did not participate in the j^{th} classroom of the i^{th} school.

The third and final stage weight for students the j^{th} classroom in the i^{th} school thus became

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

where Δ equals one when there was no student subsampling and 2 when students were subsampled within classrooms.

9.3.7 Overall Sampling Weight

The overall sampling weight was simply the product of the final first stage weight, the final second stage weight, and the final third stage weight. For example, when no subsampling of classrooms was involved, this product is given by

$$W^{i,j} = FW_{sc}^{i} \cdot FW_{cl1}^{i} \cdot FW_{st\Delta}^{i,j}$$

or

$$W^{i,j} = A_{sc} \cdot BW_{sc}^{i} \cdot FW_{cl1}^{i} \cdot A_{st}^{i,j} BW_{st\Delta}^{i,j}$$

When classrooms were subsampled within schools, the overall sampling weight was

$$W^{i,j} = FW_{sc}^{i} \cdot FW_{cl2}^{i,j} \cdot FW_{st\Delta}^{i,j}$$

or

$$W^{i,j} = A_{sc} \cdot BW_{sc}^{i} \cdot FW_{cl2}^{i,j} \cdot A_{st}^{i,j} BW_{st\Delta}^{i,j}$$

It is important to note that sampling weights vary by school and classroom, but that students within the same classroom have the same sampling weights. It is also important to note that sampling weights were calculated separately by explicit strata.[8]

9.4 Calculating School and Student Participation Rates

Since non-participation by sampled schools or students can lead to bias in the study results, a variety of participation rates were computed to show the level of success each country achieved in securing participation from their sampled schools and students. To monitor school participation, three school participation rates were computed: one based on originally sampled schools only; one based on originally sampled and first replacement schools; and one based on originally sampled and both first and second replacement schools. Classroom and student participation rates were also computed, as were overall participation rates.

9.4.1 Unweighted School Participation Rates

The three unweighted school participation rates that were computed were the following:

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

R_{unw}^{sc-r1} = unweighted school participation rate, including sampled and first replacement schools,

8 Overall sampling weights for Malaysia were modified to allow sampling estimate of national gender ratio to equal the ratio observed on the sampling frame. This was accomplished by multiplying all male (female) student weights by the desired constant.

R_{unw}^{sc-r2} = unweighted school participation rate, including sampled, first and second replacement schools.

Each unweighted school participation rate was defined as the ratio of the number of participating schools to the number of originally sampled schools, excluding any ineligible schools. A school was labelled as "participating school" if at least one of its sampled classrooms had at least a 50 percent student participation rate. The rates were calculated as follows:

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

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

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

9.4.2 Unweighted Classroom Participation Rates

The unweighted classroom participation rate was computed as follows (see section 9.3.4 for a complete definition of A_{cl}):

$$R_{unw}^{cl} = \frac{1}{A_{cl}}$$

9.4.3 Unweighted Student Participation Rates

The unweighted student participation rate was computed as follows where summations are done over all participating schools and over all classrooms with at least 50 percent of its students participating in the study:

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

9.4.4 Unweighted Overall Participation Rates

Three unweighted overall participation rates were computed for each country. They were as follows:

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

R_{unw}^{ov-r1} = unweighted overall participation rate, including sampled and first replacement schools,

R_{unw}^{ov-r2} = unweighted overall participation rate, including sampled, first and second replacement schools.

For each country, the overall participation rate was defined as the product of the unweighted school participation rate, unweighted classroom participation rate and the unweighted student participation rate. They were calculated as follows:

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

$$R_{unw}^{ov-r1} = R_{unw}^{sc-r1} \cdot R_{unw}^{cl} \cdot R_{unw}^{st}$$

$$R_{unw}^{ov-r2} = R_{unw}^{sc-r2} \cdot R_{unw}^{cl} \cdot R_{unw}^{st}$$

9.4.5 Weighted School Participation Rates

Three weighted school-level participation rates were computed for each country. They were as follows:

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

R_{wtd}^{sc-r1} = weighted school participation rate, including sampled and first replacement schools,

R_{wtd}^{sc-r2} = weighted school participation rate, including sampled, first and second replacement schools.

The weighted school participation rates were calculated as follows:

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

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

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

where both the numerator and denominator were summations over all responding students and the appropriate classroom-level and student-level sampling weights were used. Ω and Δ take the value one when no sub-sampling was involved and two otherwise. Note that the basic school-level weight appears in the numerator, whereas the final school-level weight appears in the denominator.

The denominator remains unchanged in all three equations and is the weighted estimate of the total enrolment in the target population. The numerator, however, changes from one equation to the next. Only students from originally sampled schools and from classrooms with at least 50 percent of their students participating in the study were included in the first equation. Students from first replacement schools were added in the second equation, and students from first and second replacement schools were added in the third equation.

9.4.6 Weighted Classroom Participation Rates

The weighted classroom participation rate was computed as follows:

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

where both the numerator and denominator were summations over all responding students from classrooms with at least 50 percent of their students participating in the study, and the appropriate student-level sampling weights were used. Note that the basic classroom-level weight appears in the numerator, whereas the final classroom-level weight appears in the denominator. Furthermore, the denominator in this formula was the same quantity that appears in the numerator of the weighted school-level participation rate for all participating schools, sampled and replacement.

9.4.7 Weighted Student Participation Rates

The weighted student participation rate was computed as follows:

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

where both the numerator and denominator were summations over all responding students from participating schools. Note that the basic student-level weight appears in the numerator, whereas the final student-level weight appears in the denominator. Furthermore, the denominator in this formula was the same quantity that appears in the numerator of the weighted classroom-level participation rate for all participating schools, sampled and replacement.

9.4.8 Weighted Overall Participation Rates

Three weighted overall participation rates were computed. They were as follows:

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

R_{wtd}^{ov-r1} = weighted overall participation rate, including sampled and first replacement schools,

R_{wtd}^{ov-r2} = weighted overall participation rate, including sampled, first and second replacement schools.

Each weighted overall participation rate was defined as the product of the appropriate weighted school participation rate, weighted classroom participation rate and the weighted student participation rate. They were computed as follows:

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

$$R_{wtd}^{ov-r1} = R_{wtd}^{sc-r1} \cdot R_{wtd}^{cl} \cdot R_{wtd}^{st}$$

$$R_{wtd}^{ov-r2} = R_{wtd}^{sc-r2} \cdot R_{wtd}^{cl} \cdot R_{wtd}^{st}$$

Weighted school, classroom, student, and overall participation rates were computed for each participating country using these procedures.

9.5 Meeting TIMSS' Standards for Sampling Participation

Countries understood that the goal for sampling participation was 100 percent for all sampled schools and students. Guidelines for reporting achievement data for countries securing less than full participation were modelled after IEA's TIMSS previous studies. As summarized in Exhibit 9.7, countries were assigned to one of three categories on the basis of their sampling participation. Countries in Category 1 were considered to have met the TIMSS sampling requirement, and to have an acceptable participation rate. Countries in Category 2 met the sampling requirements only after including replacement schools. Countries that failed to meet the participation requirements even with the use of replacement schools were assigned to Category 3. One of the main goals for quality data in TIMSS 2003 was to have as many countries as possible achieve Category 1 status.

Exhibits 9.8 through 9.15 present the school, classroom, student, and overall participation rates (weighted and unweighted) and achieved sample sizes for each participating country. At the eighth grade, most countries had excellent participation rates and belong in Category 1. However, Hong Kong, the Netherlands, and Scotland met the sampling requirements only after including replacement schools, and therefore belong in Category 2. Although the United States and Morocco had overall participation rates after including replacement schools of just below 75 percent (73 percent and 71 percent, respectively) it was decided during the sampling adjudication that this rate did not warrant placement in Category 3. Instead, results for the two countries in the international reports were annotated with a double-obelisk indicating that they nearly satisfied the guidelines for sample participation rates after including replacement schools. Despite extraordinary efforts to secure full participation, England's participation fell below the minimum requirement of 50 percent, so its results were annotated accordingly and placed below a line in exhibits in the International Reports. As described earlier in this chapter, a special school-level participation adjustment that capitalized on the unique implicit stratification variables used by England was applied to England's data to reduce the risk of bias.

At the fourth grade, all participants achieved the minimum acceptable participation rates, although Australia, England, Hong Kong SAR, the Netherlands, Scotland and the United States did so only after including replacement schools, and so their results were annotated with an obelisk in the achievement exhibits in the international report.

Exhibit 9. 7 **Categories of Sampling Participation**

Category 1	Acceptable sampling participation rate **without** the use of replacement schools. In order to be placed in this category, a country had to have: • An **unweighted** school response rate **without** replacement of at least 85% (after rounding to nearest whole percent) AND an **unweighted** student response rate (after rounding) of at least 85% OR • A **weighted** school response rate **without** replacement of at least 85% (after rounding to nearest whole percent) AND a **weighted** student response rate (after rounding) of at least 85% OR • The product of the (unrounded) **weighted** school response rate **without** replacement and the (unrounded) **weighted** student response rate of at least 75% (after rounding to the nearest whole percent). Countries in this category would appear in the tables and figures in international reports without annotation, and will be ordered by achievement as appropriate.
Category 2	Acceptable sampling participation rate **only when replacement schools are included**. A country would be placed in this category 2 if: • It failed to meet the requirements for Category 1 but had a weighted school response rate **without** replacement of at least 50% (after rounding to the nearest percent) AND EITHER • A **weighted** school response rate **with** replacement of at least 85% (after rounding to nearest whole percent) AND a **weighted** student response rate (after rounding) of at least 85% OR • The product of the (unrounded) **weighted** school response rate **with** replacement and the (unrounded) **weighted** student response rate of at least 75% (after rounding to the nearest whole percent). Countries in this category would be annotated with a "dagger" in the tables and figures in international reports, and ordered by achievement as appropriate.
Category 3	Unacceptable sampling response rate even when replacement schools are included. Countries that could provide documentation to show that they complied with TIMSS sampling procedures and requirements but did not meet the requirements for Category 1 or Category 2 would be placed in Category 3. Countries in this category would appear in a separate section of the achievement tables, below the other countries, in international reports. These countries would be presented in alphabetical order.

Exhibit 9. 8 **School Participation Rates & Sample Sizes – Eighth Grade**

Country	School Participation Before Replacement (Weighted Percentage)	School Participation After Replacement (Weighted Percentage)	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample That Participated	Number of Replacement Schools That Participated	Total Number of Schools That Participated
Armenia	99.3%	99.3%	150	150	149	0	149
Australia	80.7%	90.1%	230	226	186	21	207
Bahrain	100.0%	100.0%	67	67	67	0	67
Belgium (Flemish)	81.5%	98.7%	150	150	122	26	148
Botswana	97.6%	97.6%	152	150	146	0	146
Bulgaria	96.7%	97.0%	170	169	163	1	164
Chile	98.1%	100.0%	195	195	191	4	195
Chinese Taipei	100.0%	100.0%	150	150	150	0	150
Cyprus	100.0%	100.0%	59	59	59	0	59
Egypt	99.3%	100.0%	217	217	215	2	217
England	39.6%	54.1%	160	160	62	25	87
Estonia	99.3%	99.3%	154	152	151	0	151
Ghana	100.0%	100.0%	150	150	150	0	150
Hong Kong, SAR	74.5%	83.3%	150	150	112	13	125
Hungary	98.2%	98.7%	160	157	154	1	155
Indonesia	98.1%	100.0%	150	150	148	2	150
Iran, Islamic Rep. of	100.0%	100.0%	188	181	181	0	181
Israel	97.7%	99.4%	150	147	143	3	146
Italy	95.9%	100.0%	172	171	164	7	171
Japan	97.3%	97.3%	150	150	146	0	146
Jordan	100.0%	100.0%	150	140	140	0	140
Korea, Rep. of	99.3%	99.3%	151	150	149	0	149
Latvia	91.6%	93.9%	150	149	137	3	140
Lebanon	93.2%	95.0%	160	160	148	4	152
Lithuania	91.5%	95.3%	150	150	137	6	143
Macedonia, Rep. of	93.9%	99.4%	150	150	142	7	149
Malaysia	100.0%	100.0%	150	150	150	0	150
Moldova, Rep. of	98.8%	100.0%	150	149	147	2	149
Morocco	78.5%	78.5%	227	165	131	0	131
Netherlands	78.7%	86.7%	150	150	118	12	130
New Zealand	85.9%	97.1%	175	174	149	20	169
Norway	91.9%	91.9%	150	150	138	0	138
Palestinian Nat'l Auth.	100.0%	100.0%	150	145	145	0	145

Note: Some percentages may appear inconsistent because of rounding.

Exhibit 9. 8 **School Participation Rates & Sample Sizes – Eighth Grade** *(...Continued)*

Country	School Participation Before Replacement (Weighted Percentage)	School Participation After Replacement (Weighted Percentage)	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample That Participated	Number of Replacement Schools That Participated	Total Number of Schools That Participated
Philippines	81.4%	85.5%	160	160	132	5	137
Romania	99.3%	99.3%	150	149	148	0	148
Russian Federation	99.3%	99.3%	216	216	214	0	214
Saudi Arabia	95.1%	96.9%	160	160	154	1	155
Scotland	76.2%	85.3%	150	150	115	13	128
Serbia	99.3%	99.3%	150	150	149	0	149
Singapore	100.0%	100.0%	164	164	164	0	164
Slovak Republic	95.8%	100.0%	180	179	170	9	179
Slovenia	94.3%	98.7%	177	177	169	5	174
South Africa	89.4%	95.7%	265	265	241	14	255
Sweden	96.8%	99.4%	160	160	155	4	159
Syrian Arab Republic	81.0%	89.0%	150	150	121	13	134
Tunisia	100.0%	100.0%	150	150	150	0	150
United States	70.8%	78.4%	301	296	211	21	232
Benchmarking Participants							
Basque Country, Spain	99.6%	100.0%	120	120	119	1	120
Indiana State, US	96.6%	96.6%	56	56	54	0	54
Ontario Province, Can.	84.4%	93.4%	200	196	171	15	186
Quebec Province, Can.	91.2%	92.8%	199	185	173	2	175

Note: Some percentages may appear inconsistent because of rounding.

Exhibit 9. 9 **School Participation Rates & Sample Sizes – Fourth Grade**

Country	School Participation Before Replacement (Weighted Percentage)	School Participation After Replacement (Weighted Percentage)	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample That Participated	Number of Replacement Schools That Participated	Total Number of Schools That Participated
Armenia	98.7%	98.7%	150	150	148	0	148
Australia	77.9%	90.3%	230	227	178	26	204
Belgium (Flemish)	88.9%	99.3%	150	150	133	16	149
Chinese Taipei	100.0%	100.0%	150	150	150	0	150
Cyprus	100.0%	100.0%	150	150	150	0	150
England	54.3%	82.0%	150	150	79	44	123
Hong Kong, SAR	77.3%	88.0%	150	150	116	16	132
Hungary	98.2%	98.7%	160	159	156	1	157
Iran, Islamic Rep. of	100.0%	100.0%	176	171	171	0	171
Italy	96.6%	100.0%	172	171	165	6	171
Japan	100.0%	100.0%	150	150	150	0	150
Latvia	91.2%	94.0%	150	149	137	3	140
Lithuania	91.6%	95.6%	160	160	147	6	153
Moldova, Rep. of	97.4%	100.0%	153	151	147	4	151
Morocco	86.8%	86.8%	227	225	197	0	197
Netherlands	51.7%	87.2%	150	149	77	53	130
New Zealand	87.0%	97.7%	228	228	194	26	220
Norway	89.3%	92.6%	150	150	134	5	139
Philippines	78.4%	85.0%	160	160	122	13	135
Russian Federation	99.4%	100.0%	206	205	204	1	205
Scotland	63.6%	83.3%	150	150	94	31	125
Singapore	100.0%	100.0%	182	182	182	0	182
Slovenia	94.6%	98.8%	177	177	169	5	174
Tunisia	100.0%	100.0%	150	150	150	0	150
United States	69.9%	82.1%	310	300	212	36	248
Yemen	100.0%	100.0%	150	150	150	0	150
Benchmarking Participants							
Indiana State, US	100.0%	100.0%	56	56	56	0	56
Ontario Province, Can.	88.9%	94.5%	200	196	179	10	189
Quebec Province, Can.	99.0%	99.9%	198	194	192	1	193

Note: Some percentages may appear inconsistent because of rounding.

Exhibit 9. 10 **Student Participation Rates & Sample Sizes - Eighth Grade**

Country	Within School Student Participation (Weighted Percentage)	Number of Sampled Students in Participating Schools	Number of Students Withdrawn from Class/ School	Number of Students Excluded	Number of Students Eligible	Number of Students Absent	Number of Students Assessed
Armenia	90.1%	6,388	56	0	6,332	606	5,726
Australia	92.6%	5,286	60	16	5,210	419	4,791
Bahrain	97.9%	4,351	64	0	4,287	88	4,199
Belgium (Flemish)	96.7%	5,161	19	7	5,135	165	4,970
Botswana	98.0%	5,388	70	70	5,248	98	5,150
Bulgaria	95.7%	4,489	167	0	4,322	205	4,117
Chile	98.5%	6,528	15	39	6,474	97	6,377
Chinese Taipei	99.0%	5,525	54	37	5,434	55	5,379
Cyprus	96.0%	4,314	79	66	4,169	167	4,002
Egypt	97.5%	7,259	0	0	7,259	164	7,095
England	86.1%	3,360	34	0	3,326	496	2,830
Estonia	96.1%	4,242	28	5	4,209	169	4,040
Ghana	93.0%	5,690	189	0	5,501	401	5,100
Hong Kong, SAR	96.8%	5,204	33	4	5,167	195	4,972
Hungary	95.4%	3,506	7	34	3,465	163	3,302
Indonesia	99.0%	5,884	61	0	5,823	61	5,762
Iran, Islamic Rep. of	97.9%	5,215	118	52	5,045	103	4,942
Israel	94.7%	4,880	2	319	4,559	241	4,318
Italy	96.9%	4,628	35	173	4,420	142	4,278
Japan	95.9%	5,121	51	5	5,065	209	4,856
Jordan	96.5%	4,871	176	41	4,654	165	4,489
Korea, Rep. of	98.6%	5,451	18	50	5,383	74	5,309
Latvia	89.0%	4,146	23	5	4,118	488	3,630
Lebanon	95.9%	4,030	64	0	3,966	152	3,814
Lithuania	88.9%	6,619	58	955	5,606	642	4,964
Macedonia, Rep. of	96.7%	4,028	0	0	4,028	135	3,893
Malaysia	98.2%	5,464	46	0	5,418	104	5,314
Moldova, Rep. of	96.2%	4,262	58	0	4,204	171	4,033
Morocco	90.8%	3,243	25	0	3,218	275	2,943
Netherlands	93.6%	3,283	2	0	3,281	216	3,065
New Zealand	92.8%	4,343	170	65	4,108	307	3,801
Norway	92.4%	4,569	24	61	4,484	351	4,133
Palestinian Nat'l Auth.	99.0%	5,543	117	14	5,412	55	5,357
Philippines	95.9%	7,498	288	0	7,210	293	6,917

Note: Some percentages may appear inconsistent because of rounding.

Exhibit 9. 10 **Student Participation Rates & Sample Sizes - Eighth Grade** *(...Continued)*

Country	Within School Student Participation (Weighted Percentage)	Number of Sampled Students in Participating Schools	Number of Students Withdrawn from Class/ School	Number of Students Excluded	Number of Students Eligible	Number of Students Absent	Number of Students Assessed
Romania	98.2%	4,249	53	4	4,192	88	4,104
Russian Federation	97.0%	4,926	50	62	4,814	147	4,667
Saudi Arabia	97.5%	4,553	115	5	4,433	138	4,295
Scotland	89.5%	3,962	24	0	3,938	422	3,516
Serbia	96.3%	4,514	52	2	4,460	164	4,296
Singapore	96.7%	6,236	5	0	6,231	213	6,018
Slovak Republic	95.4%	4,428	16	0	4,412	197	4,215
Slovenia	92.5%	3,883	19	2	3,862	284	3,578
South Africa	92.1%	9,905	320	0	9,585	633	8,952
Sweden	89.0%	4,941	58	93	4,790	534	4,256
Syrian Arab Republic	98.0%	5,001	0	1	5,000	105	4,895
Tunisia	98.0%	5,106	74	0	5,032	101	4,931
United States	94.0%	9,891	90	279	9,522	610	8,912
Benchmarking Participants							
Basque Country, Spain	97.6%	2,736	41	113	2,582	68	2,514
Indiana State, US	97.1%	2,402	43	107	2,252	64	2,188
Ontario Province, Can.	95.1%	4,693	59	208	4,426	209	4,217
Quebec Province, Can.	91.8%	4,919	78	46	4,795	384	4,411

Note: Some percentages may appear inconsistent because of rounding.

Exhibit 9. 11 **Student Participation Rates & Sample Sizes - Fourth Grade**

Country	Within School Student Participation (Weighted Percentage)	Number of Sampled Students in Participating Schools	Number of Students Withdrawn from Class/ School	Number of Students Excluded	Number of Students Eligible	Number of Students Absent	Number of Students Assessed
Armenia	91.4%	6,275	57	0	6,218	544	5,674
Australia	94.2%	4,675	69	39	4,567	246	4,321
Belgium (Flemish)	97.7%	4,866	17	20	4,829	117	4,712
Chinese Taipei	99.3%	4,793	11	88	4,694	33	4,661
Cyprus	97.2%	4,536	27	60	4,449	121	4,328
England	92.8%	3,917	45	0	3,872	287	3,585
Hong Kong, SAR	94.9%	4,901	23	4	4,874	266	4,608
Hungary	94.0%	3,603	11	67	3,525	206	3,319
Iran, Islamic Rep. of	98.4%	4,587	83	80	4,424	72	4,352
Italy	96.7%	4,641	23	185	4,433	151	4,282
Japan	97.4%	4,690	16	16	4,658	123	4,535
Latvia	93.7%	3,980	16	4	3,960	273	3,687
Lithuania	92.0%	5,701	35	852	4,814	392	4,422
Moldova, Rep. of	97.0%	4,162	46	0	4,116	135	3,981
Morocco	93.0%	4,546	0	0	4,546	282	4,264
Netherlands	96.4%	3,080	0	30	3,050	113	2,937
New Zealand	94.8%	4,785	145	107	4,533	225	4,308
Norway	95.2%	4,706	22	107	4,577	235	4,342
Philippines	95.0%	5,225	40	31	5,154	582	4,572
Russian Federation	96.8%	4,229	54	66	4,109	146	3,963
Scotland	92.0%	4,283	34	0	4,249	313	3,936
Singapore	97.6%	6,851	16	0	6,835	167	6,668
Slovenia	91.7%	3,410	13	17	3,380	254	3,126
Tunisia	98.9%	4,408	23	0	4,385	51	4,334
United States	95.5%	10,795	49	429	10,317	488	9,829
Yemen	92.6%	4,550	0	0	4,550	345	4,205
Benchmarking Participants							
Indiana State, US	98.2%	2,472	44	151	2,277	44	2,233
Ontario Province, Can.	95.6%	4,813	91	158	4,564	202	4,362
Quebec Province, Can.	91.2%	4,864	51	73	4,740	390	4,350

Note: Some percentages may appear inconsistent because of rounding.

Exhibit 9. 12 **Unweighted School, Class, and Student Participation Rates – Eighth Grade**

Country	School Participation Before Replacement	School Participation After Replacement	Class Participation	Student Participation	Overall Participation Before Replacement	Overall Participation After Replacement
Armenia	99%	99%	99%	90%	89%	89%
Australia	82%	92%	100%	92%	76%	84%
Bahrain	100%	100%	100%	98%	98%	98%
Belgium (Flemish)	81%	99%	98%	97%	77%	94%
Botswana	97%	97%	100%	98%	96%	96%
Bulgaria	96%	97%	99%	95%	91%	92%
Chile	98%	100%	100%	99%	96%	99%
Chinese Taipei	100%	100%	100%	99%	99%	99%
Cyprus	100%	100%	100%	96%	96%	96%
Egypt	99%	100%	100%	98%	97%	98%
England	39%	54%	99%	85%	33%	46%
Estonia	99%	99%	100%	96%	95%	95%
Ghana	100%	100%	100%	93%	93%	93%
Hong Kong, SAR	75%	83%	99%	96%	71%	80%
Hungary	98%	99%	100%	95%	93%	94%
Indonesia	99%	100%	100%	99%	98%	99%
Iran, Islamic Rep. of	100%	100%	100%	98%	98%	98%
Israel	97%	99%	100%	95%	92%	94%
Italy	96%	100%	100%	97%	93%	97%
Japan	97%	97%	100%	96%	93%	93%
Jordan	100%	100%	100%	96%	96%	96%
Korea, Rep. of	99%	99%	100%	99%	98%	98%
Latvia	92%	94%	99%	88%	81%	82%
Lebanon	93%	95%	100%	96%	89%	91%
Lithuania	91%	95%	100%	89%	81%	84%
Macedonia, Rep. of	95%	99%	100%	97%	91%	96%
Malaysia	100%	100%	100%	98%	98%	98%
Moldova, Rep. of	99%	100%	100%	96%	95%	96%
Morocco	79%	79%	100%	91%	73%	73%
Netherlands	79%	87%	100%	93%	73%	81%
New Zealand	86%	97%	100%	93%	79%	90%
Norway	92%	92%	100%	92%	85%	85%
Palestinian Nat'l Auth.	100%	100%	100%	99%	99%	99%
Philippines	83%	86%	100%	96%	79%	82%
Romania	99%	99%	100%	98%	97%	97%
Russian Federation	99%	99%	100%	97%	96%	96%

Note: Some percentages may appear inconsistent because of rounding.

Exhibit 9. 12 **Unweighted School, Class, and Student Participation Rates – Eighth Grade** *(...Continued)*

Country	School Participation Before Replacement	School Participation After Replacement	Class Participation	Student Participation	Overall Participation Before Replacement	Overall Participation After Replacement
Saudi Arabia	96%	97%	100%	97%	93%	94%
Scotland	77%	85%	100%	89%	68%	76%
Serbia	99%	99%	100%	96%	96%	96%
Singapore	100%	100%	100%	97%	97%	97%
Slovak Republic	95%	100%	100%	96%	91%	96%
Slovenia	95%	98%	100%	93%	88%	91%
South Africa	91%	96%	100%	93%	85%	90%
Sweden	97%	99%	99%	89%	85%	87%
Syrian Arab Republic	81%	89%	100%	98%	79%	87%
Tunisia	100%	100%	100%	98%	98%	98%
United States	71%	78%	99%	94%	66%	73%
Benchmarking Participants						
Basque Country, Spain	99%	100%	100%	97%	97%	97%
Indiana State, US	96%	96%	100%	97%	94%	94%
Ontario Province, Can.	87%	95%	100%	95%	83%	90%
Quebec Province, Can.	94%	95%	100%	92%	86%	87%

Note: Some percentages may appear inconsistent because of rounding.

Exhibit 9. 13 **Unweighted School, Class, and Student Participation Rates – Fourth Grade**

Country	School Participation Before Replacement	School Participation After Replacement	Class Participation	Student Participation	Overall Participation Before Replacement	Overall Participation After Replacement
Armenia	99%	99%	100.0%	91%	90%	90%
Australia	78%	90%	100%	95%	74%	85%
Belgium (Flemish)	89%	99%	100%	98%	87%	97%
Chinese Taipei	100%	100%	100%	99%	99%	99%
Cyprus	100%	100%	100%	97%	97%	97%
England	53%	82%	100%	93%	49%	76%
Hong Kong, SAR	77%	88%	99%	95%	73%	83%
Hungary	98%	99%	100%	94%	92%	93%
Iran, Islamic Rep. of	100%	100%	100%	98%	98%	98%
Italy	96%	100%	100%	97%	93%	97%
Japan	100%	100%	100%	97%	97%	97%
Latvia	92%	94%	100%	93%	86%	87%
Lithuania	92%	96%	99%	92%	84%	87%
Moldova, Rep. of	97%	100%	100%	97%	94%	97%
Morocco	88%	88%	100%	94%	82%	82%
Netherlands	52%	87%	100%	96%	50%	84%
New Zealand	85%	96%	100%	95%	81%	92%
Norway	89%	93%	100%	95%	85%	88%
Philippines	76%	84%	100%	89%	68%	75%
Russian Federation	100%	100%	100%	96%	96%	96%
Scotland	63%	83%	100%	93%	58%	77%
Singapore	100%	100%	100%	98%	98%	98%
Slovenia	95%	98%	100%	92%	88%	91%
Tunisia	100%	100%	100%	99%	99%	99%
United States	71%	83%	99%	95%	67%	78%
Yemen	100%	100%	100%	92%	92%	92%
Benchmarking Participants						
Indiana State, US	100%	100%	100%	98%	98%	98%
Ontario Province, Can.	91%	96%	100%	96%	87%	92%
Quebec Province, Can.	99%	99%	100%	92%	91%	91%

Note: Some percentages may appear inconsistent because of rounding.

Exhibit 9. 14 **Weighted School, Class, and Student Participation Rates – Eighth Grade**

Country	School Participation Before Replacement	School Participation After Replacement	Class Participation	Student Participation	Overall Participation Before Replacement	Overall Participation After Replacement
Armenia	99%	99%	99%	90%	89%	89%
Australia	81%	90%	100%	93%	75%	83%
Bahrain	100%	100%	100%	98%	98%	98%
Belgium (Flemish)	82%	99%	98%	97%	77%	94%
Botswana	98%	98%	100%	98%	96%	96%
Bulgaria	97%	97%	99%	96%	92%	92%
Chile	98%	100%	100%	99%	97%	99%
Chinese Taipei	100%	100%	100%	99%	99%	99%
Cyprus	100%	100%	100%	96%	96%	96%
Egypt	99%	100%	100%	97%	97%	97%
England	40%	54%	99%	86%	34%	46%
Estonia	99%	99%	100%	96%	95%	95%
Ghana	100%	100%	100%	93%	93%	93%
Hong Kong, SAR	74%	83%	99%	97%	72%	80%
Hungary	98%	99%	100%	95%	94%	94%
Indonesia	98%	100%	100%	99%	97%	99%
Iran, Islamic Rep. of	100%	100%	100%	98%	98%	98%
Israel	98%	99%	100%	95%	93%	94%
Italy	96%	100%	100%	97%	93%	97%
Japan	97%	97%	100%	96%	93%	93%
Jordan	100%	100%	100%	96%	96%	96%
Korea, Rep. of	99%	99%	100%	99%	98%	98%
Latvia	92%	94%	100%	89%	81%	83%
Lebanon	93%	95%	100%	96%	89%	91%
Lithuania	92%	95%	100%	89%	81%	84%
Macedonia, Rep. of	94%	99%	100%	97%	91%	96%
Malaysia	100%	100%	100%	98%	98%	98%
Moldova, Rep. of	99%	100%	100%	96%	95%	96%
Morocco	79%	79%	100%	91%	71%	71%
Netherlands	79%	87%	100%	94%	74%	81%
New Zealand	86%	97%	100%	93%	80%	90%
Norway	92%	92%	100%	92%	85%	85%
Palestinian Nat'l Auth.	100%	100%	100%	99%	99%	99%
Philippines	81%	86%	100%	96%	78%	82%

Note: Some percentages may appear inconsistent because of rounding.

Exhibit 9. 14 **Weighted School, Class, and Student Participation Rates – Eighth Grade** *(...Continued)*

Country	School Participation Before Replacement	School Participation After Replacement	Class Participation	Student Participation	Overall Participation Before Replacement	Overall Participation After Replacement
Romania	99%	99%	100%	98%	98%	98%
Russian Federation	99%	99%	100%	97%	96%	96%
Saudi Arabia	95%	97%	100%	97%	93%	94%
Scotland	76%	85%	100%	89%	68%	76%
Serbia	99%	99%	100%	96%	96%	96%
Singapore	100%	100%	100%	97%	97%	97%
Slovak Republic	96%	100%	100%	95%	91%	95%
Slovenia	94%	99%	100%	93%	87%	91%
South Africa	89%	96%	100%	92%	82%	88%
Sweden	97%	99%	99%	89%	85%	87%
Syrian Arab Republic	81%	89%	100%	98%	79%	87%
Tunisia	100%	100%	100%	98%	98%	98%
United States	71%	78%	99%	94%	66%	73%
Benchmarking Participants						
Basque Country, Spain	100%	100%	100%	98%	97%	98%
Indiana State, US	97%	97%	100%	97%	94%	94%
Ontario Province, Can.	84%	93%	100%	95%	80%	89%
Quebec Province, Can.	91%	93%	100%	92%	84%	85%

Note: Some percentages may appear inconsistent because of rounding.

Exhibit 9. 15 **Weighted School, Class, and Student Participation Rates – Fourth Grade**

Country	School Participation Before Replacement	School Participation After Replacement	Class Participation	Student Participation	Overall Participation Before Replacement	Overall Participation After Replacement
Armenia	99%	99%	100%	91%	90%	90%
Australia	78%	90%	100%	94%	73%	85%
Belgium (Flemish)	89%	99%	100%	98%	87%	97%
Chinese Taipei	100%	100%	100%	99%	99%	99%
Cyprus	100%	100%	100%	97%	97%	97%
England	54%	82%	100%	93%	50%	76%
Hong Kong, SAR	77%	88%	99%	95%	73%	83%
Hungary	98%	99%	100%	94%	92%	93%
Iran, Islamic Rep. of	100%	100%	100%	98%	98%	98%
Italy	97%	100%	100%	97%	93%	97%
Japan	100%	100%	100%	97%	97%	97%
Latvia	91%	94%	100%	94%	85%	88%
Lithuania	92%	96%	99%	92%	84%	87%
Moldova, Rep. of	97%	100%	100%	97%	94%	97%
Morocco	87%	87%	100%	93%	81%	81%
Netherlands	52%	87%	100%	96%	50%	84%
New Zealand	87%	98%	100%	95%	82%	93%
Norway	89%	93%	100%	95%	85%	88%
Philippines	78%	85%	100%	95%	75%	81%
Russian Federation	99%	100%	100%	97%	96%	97%
Scotland	64%	83%	100%	92%	59%	77%
Singapore	100%	100%	100%	98%	98%	98%
Slovenia	95%	99%	100%	92%	87%	91%
Tunisia	100%	100%	100%	99%	99%	99%
United States	70%	82%	99%	95%	66%	78%
Yemen	100%	100%	100%	93%	93%	93%
Benchmarking Participants						
Indiana State, US	100%	100%	100%	98%	98%	98%
Ontario Province, Can.	89%	94%	100%	96%	85%	90%
Quebec Province, Can.	99%	100%	100%	91%	90%	91%

Note: Some percentages may appear inconsistent because of rounding.

10

Chapter 10

Item Analysis and Review

Ina V.S. Mullis, Michael O. Martin, and Dana Diaconu

10.1 Overview

Before applying item response theory (IRT) scaling to the TIMSS 2003 achievement data to derive student mathematics and science achievement scores for analysis and reporting, the TIMSS & PIRLS International Study Center conducted a review of a range of diagnostic statistics to examine and evaluate the psychometric characteristics of each achievement item in the 49 countries and four Benchmarking participants in TIMSS 2003. This review played a crucial role in the quality assurance of the TIMSS 2003 data, enabling the detection of unusual item properties that could signal a problem or error for a particular country. For example, an item that was uncharacteristically easy or difficult, or had an unusually low discriminating power, could indicate a potential problem with either translation or printing. Similarly, a constructed-response item with unusually low scoring reliability could indicate a problem with a scoring rubric in a particular country. In the rare instances where such items were found, the country's translation verification documents and printed booklets were examined for flaws or inaccuracies and, if necessary, the item was removed from the international database for that country. This chapter describes the basic item statistics that were consulted and the review criteria that were applied, and provides examples from the assessment to illustrate the review process.

10.2 Statistics for Item Analysis

To begin the review process, the International Study Center computed item analysis statistics for each mathematics and science achievement item, showing the properties of the item in each of the 49 countries and four Benchmarking entities participating in TIMSS 2003. Exhibits 10.1 and 10.2 show examples of the statistics calculated for a multiple-choice and a constructed-response item, respectively. Statistics for each item were displayed alphabetically by country, with the international average for each statistic in the bottom row. For those countries that tested in more than one language, statistics were presented separately by language group. For all items, regardless of item format, statistics included the number of students that responded in each country, the difficulty level (the percentage of students that answered the item correctly), and the discrimination index (the point-biserial correlation between success on the item and a total score).[1] Also provided was an estimate of the item's difficulty using a Rasch one-parameter IRT model. The international means of the item difficulties and item discriminations served as guides to the overall statistical properties of the items.

Statistics displayed for multiple-choice items included the percentage of students that chose each option, as well as the percentage of students that omitted or did not reach the item, and the point-biserial correlation between the response to each option and the total score. Statistics displayed for constructed-response items (which could have one or two score levels) included the difficulty and discrimination of each score level. Constructed-response item displays also provided information about the reliability with which the item was scored in each country, with the total number of double-scored cases and the percent exact agreement between the scorers.

Detailed descriptions of the statistics provided in Exhibits 10.1 and 10.2 are listed below in order of appearance in the displays:

N: This is the number of students to whom the item was administered. If a student did not reach an item in the achievement booklet, the item was considered not administered for the purpose of the item analysis.[2]

Diff: Item difficulty is the percentage of students providing a fully correct response to the item. In the case of constructed-response items worth more than one point, this was the percentage of students receiving the maximum score. For the computation of this statistic, not reached items were treated as not administered.

Disc: Item discrimination was computed as the correlation between a correct response to the item and the total score on all of the items in the test booklet.[3] Items exhibiting good measurement properties should have a moderately positive correlation.

PCT_A, PCT_B, PCT_C, PCT_D, and PCT_E: Used for multiple-choice items only (see Exhibit 10.1), each column indicates the percentage of students choosing the particular response option for the item (A, B, C, D, or E). Not reached items were excluded from the denominator for these calculations.

1 For the purpose of computing the discrimination index, the total score was the percentage of items a student answered correctly.

2 In TIMSS, for the purposes of item analysis and item parameter estimation in scaling, items not reached by a student were treated as if they had not been administered. For purposes of estimating student proficiency, however, not reached items were treated as incorrectly answered.

3 For constructed-response items, the discrimination is the correlation between the number of score points and total score.

Exhibit 10.1 International Item Statistics for Item M012040

Trends in International Mathematics and Science Study - TIMSS 2003 Main Survey   July 4, 2004   10
International Item Statistics (Unweighted) - 8th Grade For Internal Review Only: DO NOT CITE OR CIRCULATE

Mathematics : Algebra (M012040 - M01_10)
Label: Find n from proportionality equation Type : MC Key: B

Country	N	Diff	Disc	Pct_A	Pct_B	Pct_C	Pct_D	Pct_E	Pct_In	Pct_OM	Pct_NR	PB_A	PB_B	PB_C	PB_D	PB_E	PB_In	PB_OM	RDIFF	Flags
Armenia	931	55.3	0.60	9.1	55.3	9.6	6.3	.	0.1	19.5	0.0	-0.21	0.60	-0.21	-0.17	.	0.00	-0.34	-1.28	____F.
Australia	800	74.5	0.44	8.6	74.4	8.9	6.4	.	0.0	1.8	0.2	-0.19	0.44	-0.29	-0.15	.	0.00	-0.14	-1.58	____F.
Bahrain	666	47.3	0.47	11.0	47.3	20.6	13.4	.	0.3	7.5	2.8	-0.12	0.47	-0.23	-0.17	.	0.00	-0.16	-1.23	_____.
Belgium (Flemish)	848	87.9	0.36	3.3	88.0	2.9	5.0	.	0.0	0.8	0.1	-0.21	0.36	-0.17	-0.19	.	0.00	-0.12	-1.95	__E_F.
Botswana	876	41.3	0.36	27.2	41.6	14.0	13.4	.	0.0	3.9	1.2	-0.15	0.36	-0.19	-0.13	.	0.00	-0.02	-1.34	_____.
Bulgaria	666	61.1	0.48	11.1	61.1	14.3	9.2	.	0.2	4.2	0.4	-0.21	0.48	-0.27	-0.15	.	-0.03	-0.13	-1.07	____F.
Chile	995	50.6	0.51	8.8	50.6	16.7	10.7	.	0.0	13.3	5.2	-0.11	0.51	-0.28	-0.09	.	0.00	-0.22	-1.30	____F.
Chinese Taipei	904	83.3	0.53	4.0	83.3	5.0	7.5	.	0.1	0.1	0.0	-0.30	0.53	-0.35	-0.23	.	-0.05	-0.03	-1.18	____F.
Cyprus	666	65.5	0.47	8.6	65.5	12.2	10.8	.	0.0	3.0	0.3	-0.20	0.47	-0.29	-0.21	.	0.00	-0.03	-1.57	____F.
Egypt	1167	68.7	0.51	6.7	68.7	15.7	6.2	.	0.0	2.7	0.5	-0.20	0.51	-0.36	-0.19	.	0.00	-0.07	-1.88	__E_F.
England	505	76.0	0.37	8.5	75.6	6.5	7.1	.	0.0	2.2	0.2	-0.15	0.37	-0.26	-0.14	.	0.00	-0.11	-1.54	____F.
Estonia	695	70.5	0.39	10.6	70.6	4.9	8.5	.	0.3	5.0	0.0	-0.27	0.39	-0.18	-0.09	.	-0.06	-0.12	-0.80	__H_F.
Ghana	845	29.0	0.27	26.3	29.0	24.1	15.1	.	0.2	5.2	2.3	-0.06	0.27	-0.12	-0.05	.	-0.06	-0.07	-1.22	_____.
Hong Kong, SAR	830	88.5	0.34	2.0	88.4	3.3	5.7	.	0.0	0.6	0.0	-0.19	0.34	-0.27	-0.13	.	0.00	-0.06	-1.46	____F.
Hungary	550	80.6	0.46	6.0	80.2	6.2	4.7	.	0.0	2.9	0.5	-0.25	0.46	-0.29	-0.13	.	0.00	-0.17	-1.48	____F.
Indonesia	957	61.0	0.45	17.1	61.0	11.0	7.8	.	0.0	3.0	1.4	-0.20	0.45	-0.26	-0.14	.	0.00	-0.10	-1.71	____F.
Iran, Islamic Rep. of	845	66.7	0.40	11.4	67.2	9.3	9.3	.	0.2	2.5	0.9	-0.20	0.40	-0.24	-0.17	.	0.01	-0.03	-2.10	__E_F.
Israel	717	71.7	0.48	7.8	71.5	8.5	7.0	.	0.0	5.2	0.4	-0.18	0.48	-0.30	-0.19	.	0.00	-0.16	-1.40	____F.
Italy	708	66.4	0.42	9.6	66.2	8.8	10.3	.	0.1	4.9	1.4	-0.19	0.42	-0.31	-0.08	.	-0.05	-0.11	-1.25	____F.
Japan	804	79.0	0.44	4.7	79.0	5.3	10.3	.	0.0	0.6	0.2	-0.21	0.44	-0.26	-0.21	.	0.00	-0.11	-0.96	__H_F.
Jordan	758	53.3	0.51	9.0	53.3	24.0	10.9	.	0.1	2.6	0.1	-0.13	0.51	-0.33	-0.17	.	-0.03	-0.13	-1.38	____F.
Korea, Rep. of	890	89.4	0.51	3.7	89.3	2.8	4.2	.	0.0	0.0	0.0	-0.27	0.51	-0.29	-0.30	.	0.00	0.00	-1.72	____F.
Latvia	603	73.7	0.48	9.1	73.8	9.8	4.8	.	0.0	2.5	0.2	-0.28	0.48	-0.27	-0.13	.	0.00	-0.14	-1.25	____F.
Lebanon	645	71.6	0.45	10.2	71.6	9.6	4.7	.	0.0	3.9	0.5	-0.24	0.45	-0.26	-0.12	.	0.00	-0.14	-2.13	__E_F.
Lithuania	848	66.0	0.48	13.0	66.0	8.8	7.4	.	0.1	4.6	0.1	-0.24	0.48	-0.25	-0.15	.	-0.04	-0.19	-0.95	__H_F.
Macedonia, Rep. of	650	49.5	0.54	12.3	49.5	17.5	9.2	.	0.6	10.8	0.3	-0.14	0.54	-0.29	-0.13	.	-0.05	-0.22	-0.95	__H_F.
Malaysia	873	83.3	0.36	7.0	83.3	4.5	4.5	.	0.0	0.8	0.5	-0.20	0.36	-0.22	-0.14	.	0.00	-0.09	-2.07	__E_F.
Moldova, Rep. of	662	62.4	0.47	9.5	62.4	16.3	6.5	.	0.0	5.3	0.5	-0.19	0.47	-0.26	-0.14	.	0.00	-0.15	-1.33	____F.
Morocco	521	55.5	0.43	9.6	55.1	18.8	9.4	.	0.2	6.9	1.9	-0.09	0.43	-0.29	-0.13	.	-0.05	-0.11	-1.77	____F.
Netherlands	517	86.1	0.32	7.9	86.1	2.1	2.9	.	0.0	1.0	0.2	-0.25	0.32	-0.14	-0.10	.	0.00	-0.05	-1.76	____F.
New Zealand	629	65.2	0.47	13.4	65.2	11.4	8.3	.	0.0	1.7	0.0	-0.18	0.47	-0.29	-0.18	.	0.00	-0.15	-1.17	____F.
Norway	677	60.2	0.46	11.1	60.3	16.2	6.8	.	0.1	5.5	1.0	-0.10	0.46	-0.32	-0.15	.	-0.02	-0.17	-1.28	____F.
Palestinian Nat'l Aut	911	53.1	0.50	11.6	52.4	23.3	9.4	.	0.3	3.0	0.5	-0.12	0.50	-0.34	-0.16	.	-0.07	-0.08	-1.53	____F.
Philippines	1264	52.9	0.44	17.7	53.5	17.3	10.1	.	0.1	1.3	0.7	-0.19	0.44	-0.27	-0.13	.	0.00	-0.02	-1.65	_____.
Romania	678	63.6	0.57	5.5	63.6	17.3	9.6	.	0.0	4.1	0.3	-0.20	0.56	-0.36	-0.18	.	0.00	-0.18	-1.20	____F.
Russian Federation	773	71.4	0.44	9.3	71.4	8.5	6.1	.	0.3	4.4	0.8	-0.23	0.44	-0.22	-0.14	.	-0.02	-0.18	-1.16	____F.
Saudi Arabia	699	33.6	0.42	10.4	33.6	31.8	16.2	.	0.0	8.0	1.7	-0.05	0.43	-0.25	-0.11	.	0.00	-0.11	-1.17	_____.
Scotland	587	78.9	0.38	10.1	78.9	4.9	4.8	.	0.0	1.4	0.0	-0.18	0.38	-0.23	-0.17	.	0.00	-0.13	-1.82	____F.
Serbia	718	64.3	0.53	12.1	64.3	10.0	6.7	.	0.0	6.8	0.6	-0.25	0.53	-0.24	-0.13	.	0.00	-0.26	-1.23	____F.
Singapore	993	93.2	0.30	1.7	93.2	1.1	3.7	.	0.0	0.3	0.0	-0.16	0.30	-0.18	-0.19	.	0.00	-0.01	-2.04	__E_F.
Slovak Republic	683	77.9	0.43	8.9	77.9	4.4	6.0	.	0.1	2.6	0.7	-0.24	0.43	-0.24	-0.14	.	-0.01	-0.14	-1.45	____F.
Slovenia	623	71.9	0.47	12.8	71.9	7.5	4.7	.	0.0	3.0	0.0	-0.26	0.47	-0.25	-0.14	.	0.00	-0.18	-1.55	____F.
South Africa	1415	27.0	0.47	14.6	27.0	33.1	19.3	.	0.7	5.3	4.4	-0.05	0.47	-0.20	-0.10	.	-0.09	-0.14	-1.01	__H__.
Sweden	707	65.7	0.40	15.4	65.3	8.2	6.5	.	0.0	4.5	1.1	-0.14	0.40	-0.23	-0.16	.	0.00	-0.15	-1.02	__H_F.
Syrian Arab Republic	811	47.7	0.41	13.3	47.7	23.2	11.7	.	0.0	4.1	0.7	-0.13	0.41	-0.24	-0.10	.	0.00	-0.10	-1.48	_____.
Tunisia	788	67.5	0.34	11.9	67.5	10.0	4.2	.	0.5	5.8	4.8	-0.14	0.34	-0.23	-0.01	.	-0.06	-0.14	-2.21	__E_F.
United States	1503	79.9	0.41	8.3	79.9	5.0	5.8	.	0.0	1.0	0.5	-0.23	0.41	-0.22	-0.18	.	0.00	-0.05	-1.86	__E_F.
International Avg.	.	65.7	0.44	10.3	65.7	11.8	8.1	.	0.1	4.0	0.9	-0.18	0.44	-0.26	-0.15	.	-0.01	-0.12	-1.46	__E_F.
Basque Country, Spain	420	76.0	0.42	4.3	76.0	8.8	8.1	.	0.0	2.9	0.7	-0.11	0.42	-0.31	-0.22	.	0.00	-0.04	-1.77	____F.
Ontario Province, Can	717	85.4	0.39	6.8	85.6	3.8	3.5	.	0.0	0.3	0.4	-0.26	0.39	-0.23	-0.13	.	0.00	-0.06	-2.07	__E_F.
Quebec Province, Can.	736	86.4	0.33	5.2	86.4	3.3	4.1	.	0.0	1.1	0.3	-0.18	0.33	-0.17	-0.18	.	0.00	-0.07	-1.83	____F.

Keys: Diff: Percent obtaining correct score; Disc: Discrimination; Pct_A...D: Percent choosing each option;
Pct_In, OM, NR: Percent Invalid, Omitted, and Not Reached Responses; PB_A...D: Point Biserial for each option;
PB_OM: Point Biserial for omittted. RDIFF= Difficulty (1-PL).
Flags: A= Ability not ordered/ Attractive distractor; C= Difficulty less than chance; D= Negative/low discrimination; E= Easier than average;
F= Distractor chosen by less than 10%; H= Harder than average; R= Scoring reliability < 80%; V= Difficulty greater than 95.

Exhibit 10.2 International Item Statistics for Item S032680

Trends in International Mathematics and Science Study - TIMSS 2003 Main Survey August 12, 2004 95
International Item Statistics (Unweighted) - 8th Grade For Internal Review Only: DO NOT CITE OR CIRCULATE

Science : Chemistry (S032680 - S07_04)
Label: Identify iron, water and oxygen Type : CR Key: X

Country	N	Diff	Disc	Pct_0	Pct_1	Pct_2	Pct_3	Pct_OM	Pct_NR	PB_0	PB_1	PB_2	PB_3	PB_OM	RDIFF	Reliability Cases	Reliability Score	Reliability Code	Flags
Armenia	462	37.2	0.66	7.4	7.4	37.2	.	48.1	0.0	-0.16	0.06	0.63	.	-0.55	0.05	410	99.3	99.3	__H_F.
Australia	407	52.3	0.61	16.0	25.6	52.3	.	6.1	0.0	-0.29	-0.25	0.60	.	-0.35	-0.33	193	100.0	100.0	_____.
Bahrain	337	25.5	0.52	23.7	35.9	25.5	.	14.8	0.0	-0.27	0.21	0.35	.	-0.40	-0.04	174	100.0	99.4	_____.
Belgium (Flemish)	434	58.5	0.60	11.3	27.0	58.5	.	3.2	0.0	-0.38	-0.26	0.56	.	-0.25	-0.57	227	100.0	99.6	_____.
Botswana	446	32.3	0.62	39.9	23.3	32.3	.	4.5	0.0	-0.48	0.03	0.56	.	-0.17	-0.74	219	95.4	90.4	__E__.
Bulgaria	335	40.3	0.58	16.7	19.7	40.3	.	23.3	0.0	-0.18	0.09	0.49	.	-0.50	0.02	175	92.6	91.4	__H__.
Chile	521	17.5	0.50	38.8	29.8	17.5	.	14.0	0.0	-0.19	0.06	0.46	.	-0.31	0.53	256	96.5	94.5	__H__.
Chinese Taipei	445	69.0	0.68	6.1	17.8	69.0	.	7.2	0.0	-0.30	-0.29	0.65	.	-0.46	-0.27	172	98.8	98.8	____F.
Cyprus	326	34.0	0.61	18.1	30.7	34.0	.	17.2	0.0	-0.24	-0.04	0.56	.	-0.40	-0.34	179	98.3	97.2	_____.
Egypt	591	42.1	0.70	27.6	22.7	42.1	.	7.6	0.0	-0.56	-0.01	0.63	.	-0.20	-0.45	66	100.0	100.0	_____.
England	251	68.9	0.62	5.6	21.1	68.9	.	4.4	0.0	-0.30	-0.30	0.58	.	-0.37	-0.61	59	98.3	98.3	____F.
Estonia	335	74.9	0.60	6.3	15.5	74.9	.	3.3	0.0	-0.37	-0.31	0.58	.	-0.28	-0.66	217	100.0	99.5	____F.
Ghana	427	11.0	0.58	52.5	15.2	11.0	.	21.3	0.0	-0.30	0.28	0.46	.	-0.23	-0.18	327	98.8	96.3	_____.
Hong Kong, SAR	424	67.9	0.50	9.2	19.8	67.9	.	3.1	0.0	-0.30	-0.14	0.44	.	-0.35	-0.32	205	100.0	99.0	____F.
Hungary	278	66.9	0.58	10.4	20.5	66.9	.	2.2	0.0	-0.36	-0.34	0.58	.	-0.18	-0.54	223	93.7	92.8	_____.
Indonesia	485	23.3	0.62	32.2	24.9	23.3	.	19.6	0.0	-0.36	0.13	0.53	.	-0.28	0.06	233	95.3	93.1	__H__.
Iran, Islamic Rep. of	431	40.6	0.54	16.9	31.3	40.6	.	11.1	0.0	-0.29	-0.16	0.53	.	-0.23	-0.53	193	95.9	93.8	_____.
Israel	358	51.7	0.59	8.9	31.8	51.7	.	7.5	0.0	-0.27	-0.26	0.57	.	-0.32	-0.71	183	92.9	92.3	__E_F.
Italy	363	36.6	0.49	15.7	35.3	36.6	.	12.4	0.0	-0.17	-0.16	0.48	.	-0.29	-0.07	206	99.0	99.0	_____.
Japan	399	70.2	0.59	9.0	17.3	70.2	.	3.5	0.0	-0.35	-0.29	0.57	.	-0.28	-0.55	203	98.5	98.0	____F.
Jordan	383	31.3	0.65	29.8	31.9	31.3	.	7.0	0.0	-0.48	0.05	0.55	.	-0.24	0.03	179	99.4	98.3	__H__.
Korea, Rep. of	444	66.9	0.66	10.8	18.2	66.9	.	4.1	0.0	-0.42	-0.23	0.61	.	-0.36	-0.35	205	100.0	99.5	_____.
Latvia	304	60.9	0.62	17.8	13.2	60.9	.	8.2	0.3	-0.42	-0.08	0.57	.	-0.32	-0.25	176	98.3	97.7	_____.
Lebanon	326	40.2	0.66	22.7	20.9	40.2	.	16.3	0.0	-0.37	-0.07	0.63	.	-0.34	-0.68	174	98.3	97.1	_____.
Lithuania	420	55.0	0.53	10.7	28.8	55.0	.	5.5	0.0	-0.27	-0.26	0.52	.	-0.25	-0.41	190	95.3	94.7	_____.
Macedonia, Rep. of	322	36.6	0.65	14.9	18.9	36.6	.	29.5	0.0	-0.26	0.07	0.58	.	-0.47	-0.09	142	100.0	100.0	_____.
Malaysia	432	68.1	0.55	9.3	19.2	68.1	.	3.5	0.0	-0.30	-0.28	0.54	.	-0.27	-0.90	235	99.6	99.1	__E_F.
Moldova, Rep. of	333	36.0	0.50	3.9	41.7	36.0	.	18.3	0.0	-0.06	-0.06	0.42	.	-0.42	-0.40	180	100.0	100.0	____F.
Morocco	254	14.2	0.44	28.7	35.0	14.2	.	22.0	0.0	-0.16	0.22	0.30	.	-0.33	0.18	104	90.4	87.5	__H__.
Netherlands	259	65.3	0.58	10.4	20.8	65.3	.	3.5	0.0	-0.33	-0.31	0.58	.	-0.26	-0.58	218	94.5	94.5	_____.
New Zealand	315	45.7	0.61	21.0	28.9	45.7	.	4.4	0.0	-0.35	-0.23	0.61	.	-0.26	-0.20	167	98.8	98.2	_____.
Norway	339	64.3	0.56	11.2	20.1	64.3	.	4.4	0.0	-0.36	-0.20	0.52	.	-0.28	-0.92	202	97.5	97.5	__E__.
Palestinian Nat'l Aut	449	26.7	0.60	34.1	33.9	26.7	.	5.3	0.0	-0.45	0.13	0.47	.	-0.25	-0.16	212	97.6	97.2	_____.
Philippines	642	17.4	0.62	39.9	21.0	17.4	.	21.7	0.0	-0.26	0.18	0.53	.	-0.36	0.02	352	96.3	93.8	__H__.
Romania	332	44.0	0.61	13.9	29.2	44.0	.	13.0	0.0	-0.25	-0.18	0.59	.	-0.37	-0.43	151	98.0	97.4	_____.
Russian Federation	395	62.0	0.56	13.2	18.0	62.0	.	6.8	0.0	-0.35	-0.24	0.56	.	-0.24	-0.40	194	100.0	100.0	_____.
Saudi Arabia	361	13.3	0.40	30.5	41.6	13.3	.	14.7	0.0	-0.16	0.03	0.38	.	-0.21	0.00	208	97.6	95.2	_____.
Scotland	291	66.7	0.53	9.6	20.3	66.7	.	3.4	0.0	-0.34	-0.25	0.51	.	-0.22	-0.82	76	98.7	96.1	__E_F.
Serbia	352	49.4	0.57	11.9	22.4	49.4	.	16.2	0.0	-0.21	-0.13	0.54	.	-0.40	-0.50	196	99.5	99.0	_____.
Singapore	497	71.0	0.60	6.2	20.9	71.0	.	1.8	0.0	-0.36	-0.33	0.56	.	-0.25	-0.82	200	100.0	100.0	__E_F.
Slovak Republic	342	61.7	0.56	12.6	19.3	61.7	.	6.4	0.0	-0.37	-0.11	0.50	.	-0.31	-0.28	196	99.0	99.0	_____.
Slovenia	312	58.7	0.56	11.2	22.8	58.7	.	7.4	0.0	-0.28	-0.24	0.55	.	-0.30	-0.44	99	98.0	98.0	_____.
South Africa	743	11.8	0.52	52.1	26.9	11.8	.	9.2	0.0	-0.38	0.22	0.41	.	-0.14	-0.34	396	99.0	97.2	_____.
Sweden	355	75.5	0.54	7.9	13.5	75.5	.	3.1	0.0	-0.34	-0.27	0.53	.	-0.24	-0.97	178	94.4	93.8	__E_F.
Syrian Arab Republic	404	26.5	0.57	30.2	30.7	26.5	.	12.6	0.0	-0.40	0.17	0.43	.	-0.25	-0.07	172	91.9	89.5	_____.
Tunisia	409	16.9	0.49	25.7	41.3	16.9	.	16.1	0.0	-0.22	0.21	0.34	.	-0.36	-0.19	196	100.0	100.0	_____.
United States	756	41.4	0.53	19.4	35.6	41.4	.	3.6	0.0	-0.32	-0.16	0.50	.	-0.23	0.00	370	97.8	97.0	__H__.
International Avg.	.	45.7	0.58	18.8	24.8	45.7	.	10.7	0.0	-0.31	-0.09	0.52	.	-0.30	-0.35	.	97.7	96.8	__E__.
Basque Country, Spain	216	42.1	0.61	16.7	32.9	42.1	.	8.3	0.0	-0.34	-0.24	0.61	.	-0.23	-0.34	229	97.8	97.4	_____.
Indiana State, US	195	41.0	0.46	17.4	39.5	41.0	.	2.1	0.0	-0.29	-0.23	0.46	.	-0.05	-0.12	90	97.8	97.8	_____.
Ontario Province, Can	361	56.8	0.60	15.8	24.7	56.8	.	2.8	0.0	-0.39	-0.26	0.59	.	-0.22	-0.42	175	92.6	91.4	_____.
Quebec Province, Can.	368	60.3	0.57	14.4	21.5	60.3	.	3.8	0.0	-0.36	-0.21	0.54	.	-0.26	-0.32	109	92.7	90.8	_____.

Keys: Diff: Percent obtaining maximum score; Disc: Discrimination; Pct_0...3: Percent obtaining score level; Pct_OM, NR: Percent Omitted, and Not Reached Responses; PB_0...3: Point Biserial for each score level; PB_OM: Point Biserial for omittted; RDIFF= Difficulty (1-PL); Reliability (Cases): Responses Double Scored; Reliability (Score): Percent Agreement on Score; Reliability (Code): Percent Agreement on Code.
Flags: A= Ability not ordered/ Attractive distractor; C= Difficulty less than chance; D= Negative/low discrimination; E= Easier than average; F= Score obtained by less than 10%; H= Harder than average; R= Scoring reliability < 80%; V= Difficulty greater than 95.

PCT_0, PCT_1, PCT_2, and PCT_3: Used for constructed-response items only (see Exhibit 10.2), each column indicates the percentage of students scoring at the particular score level, up to and including the maximum score level for the item. Not reached items were excluded from the denominator for these calculations.

PCT_IN: Used for multiple-choice items only, this was the percentage of students that provided an invalid response to a multiple-choice item. Typically, invalid responses were the result of students selecting more than one response option for the same item.

PCT_OM: This is the percentage of students who, having reached the item, did not provide a response. Not reached items were excluded from the denominator when calculating this statistic.

PCT_NR: This is the percentage of student that did not reach the item. An item was coded as not reached when there was no evidence of a response to any subsequent items in the booklet and the response to the item preceding it was omitted.

PB_A, PB_B, PB_C, PB_D, and PB_E: Used for multiple-choice items only, these present the correlation between choosing each of the response options A, B, C, D, or E and the score on the test booklet. Items with good psychometric properties have near-zero or negative correlations for the distracter options (the incorrect options) and moderately positive correlations for the correct option.

PB_0, PB_1, PB_2, and PB_3: Used for constructed-response items only, these present the correlation between the score levels on the item (0, 1, 2, or 3) and the score on the test booklet. For items with good measurement properties the correlation coefficients should change from negative to positive as the score on the item increases.

PB_OM: This is the correlation between a binary variable - indicating an omitted response to the item - and the score on the test booklet. This correlation should be negative or near zero.

PB_IN: Used for multiple-choice items only, this presents the correlation between an invalid response to the item (usually caused by selecting more than one response option) and the score on the test booklet. This correlation also should be negative or near zero.

RDIFF: This is an estimate of the item's difficulty based on a Rasch one-parameter IRT model. The difficulty estimate is expressed in the logit metric (with a positive logit indicating a difficult item) and was scaled so that the average Rasch item difficulty was zero within each country.

Reliability - Cases: To provide a measure of the reliability of the scoring of the constructed-response items, those items in approximately 25 percent of the test booklets in each country were scored by two independent scorers. This column indicates the number of times each item was double-scored in a country.

Reliability - Score: This column contains the percentage of exact agreement between two independent scorers.

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

- Item difficulty exceeds 95 percent in the sample as a whole
- Item difficulty is less than 25 percent for 4-option multiple-choice items in the sample as a whole
- One or more of the distracter percentages is less than 10 percent
- One or more of the distracter percentages is greater than the percentage for the correct answer, or the point-biserial correlation for one or more of the distracters exceeds zero
- Item discrimination (i.e., the point-biserial for the correct answer) is less than 0.2
- Item discrimination does not increase with each score level (for constructed-response items with more than one score level)
- The Rasch difficulty estimate is harder than the average across all items
- The Rasch difficulty estimate is easier than the average across all items
- Difficulty levels on the item differ significantly for males and females
- Difference in item difficulty levels between males and females diverge significantly
- Scoring reliability is less than 70 percent (for constructed-response items only)

Although not all of these conditions necessarily indicate a problem, the flags are a useful way to draw attention to potential sources of concern.

In order to measure trends, TIMSS 2003 included items from TIMSS 1999 and TIMSS 1995 at the eighth grade and from TIMSS 1995 at the fourth grade.[4] For these trend items, the review included an examination of changes in item statistics between 1999 and 2003 at eighth grade and between 1995 and 2003 at fourth grade. An example item statistics display for an eighth-grade trend item is shown in Exhibit 10.3. Different from the item statistics presented in Exhibits 10.1 and 10.2, this display includes countries' statistics from both the TIMSS 1999 and TIMSS 2003 assessments. In review-

4 For more information on trend items, see Chapter 2.

ing these item statistics, the aim was to detect any unusual changes in item properties between assessments, which might indicate a problem in using the item to measure change.

10.2.1 Item-by-Country Interaction

Although countries are expected to exhibit some variation in performance across items, in general, countries with high average performance on the achievement test as a whole should perform relatively well on each of the items, and low-scoring countries should do less well on each of items. When this does not occur, i.e., when a high-scoring country has low performance on an item on which other countries are doing well, there is said to be an item-by-country interaction. When large, such item-by-country interactions may be a sign of an item that is flawed in some way and measures should be taken to address the problem.

To assist in detecting sizeable item-by-country interactions, the International Study Center produced a graphical display for each item showing the average probability across all countries of a correct response for a student of average international proficiency, compared with the probability of a correct response by a student of average proficiency in each country. Exhibit 10.4 provides an example of a TIMSS item-by-country interaction display. The probability for each country is presented as a 95 percent confidence interval, which includes a built-in Bonferroni correction for multiple comparisons. The limits for the confidence interval are computed as follows:

$$\text{Upper Limit} = \left(1 - \frac{e^{RDIFF_{ik} + SE_{RDIFF_{ik}} \times Z_b}}{1 + e^{RDIFF_{ik} + SE_{RDIFF_{ik}} \times Z_b}}\right)$$

$$\text{Lower Limit} = \left(1 - \frac{e^{RDIFF_{ik} - SE_{RDIFF_{ik}} \times Z_b}}{1 + e^{RDIFF_{ik} - SE_{RDIFF_{ik}} \times Z_b}}\right)$$

where $RDIFF_{ik}$ is the Rasch difficulty of item k within country i; $SE_{RDIFFik}$ is the standard error of the difficulty of item k in country i; and Z_b is the critical value from the Z distribution, corrected for multiple comparisons using the Bonferroni procedure.

The International Study Center also produced item-by-country interaction displays for each item in the trend study, showing for eighth grade the results from 1999 and 2003 separately in each display, and for fourth grade, the results from 1995 and 2003. An example of an item-by-country interaction display for a trend item is presented in Exhibit 10.5. Confidence intervals for 1999 and 2003 within a country appear side-by-side in the display

Exhibit 10.3 **International Item Statistics for Trend Item M012001**

Trends in International Mathematics and Science Study - TIMSS 2003 Main Survey 15:49 Wednesday, August 25, 2004
Percent of responses by Item Category - 8th Grade
For Internal Review Only: DO NOT CITE OR CIRCULATE

Mathematics: Number (M012001 - M01_01) Label: Number of squares in shaded fraction Item Type = MC Key = A

COUNTRY	Year	N	A	B	C	D	E	OTHER INCOR RECT	DIFF	INVALID	NOT REACHED	OMIT	1.GIRL % Right	2.BOY % Right
Belgium (Flemish)	1999	5256	85.5	2.1	1.9	3.8	6.4	0.3	85.5	0.0	0.0	0.3	84.4	86.5
	2003	849	85.9	2.0	2.0	3.7	4.9	1.5	85.9	0.2	0.0	1.3	84.3	87.6
Bulgaria	1999	3270	66.0	5.4	6.4	9.0	10.4	2.8	66.0	0.2	0.1	2.6	64.7	67.3
	2003	669	46.6	8.8	11.4	13.3	13.0	6.9	46.6	0.0	0.0	6.9	45.2	48.1
Chile	1999	5862	18.3	3.8	7.0	23.8	42.0	5.1	18.3	0.0	0.0	5.0	16.3	20.3
	2003	1050	25.0	4.3	6.4	22.5	32.3	9.6	25.0	0.1	0.9	8.7	23.6	26.3
Chinese Taipei	1999	5772	78.4	2.4	5.8	4.9	8.5	0.1	78.4	0.0	0.0	0.0	78.9	77.8
	2003	904	78.7	2.2	6.0	4.6	8.5	0.0	78.7	0.0	0.0	0.0	77.4	79.9
Cyprus	1999	3109	60.2	7.9	7.5	10.5	13.0	0.9	60.2	0.2	0.0	0.8	58.5	61.8
	2003	668	56.1	8.1	11.1	8.2	12.0	4.5	56.1	0.0	0.0	4.5	55.5	56.8
England	1999	2946	58.7	2.7	4.7	12.7	20.7	0.4	58.7	0.1	0.0	0.3	54.3	62.9
	2003	506	66.8	2.4	6.5	8.7	15.2	0.4	66.8	0.0	0.0	0.4	64.5	69.6
Hong Kong, SAR	1999	5176	86.3	2.0	3.0	2.7	5.9	0.2	86.3	0.1	0.0	0.1	86.0	86.5
	2003	830	87.8	1.8	3.9	2.5	3.7	0.2	87.8	0.1	0.0	0.1	86.2	89.5
Hungary	1999	3178	67.2	2.8	4.2	10.9	12.7	2.3	67.2	0.2	0.0	2.1	65.9	68.4
	2003	548	66.8	3.6	8.6	8.6	10.2	2.2	66.8	0.0	0.0	2.2	68.0	65.7
Indonesia	1999	5847	24.6	7.8	20.8	17.9	27.4	1.5	24.6	0.1	0.0	1.4	24.3	24.9
	2003	970	22.6	10.7	22.9	15.8	22.8	5.3	22.6	0.0	0.1	5.2	19.9	25.5
Iran, Islamic Rep. o	1999	5291	47.2	4.4	5.5	15.3	25.5	2.1	47.2	0.0	0.1	2.0	41.3	51.3
	2003	853	42.0	7.0	7.7	15.2	25.9	2.1	42.0	0.1	0.1	1.9	41.8	42.1
Israel	1999	4191	49.1	8.5	7.2	13.5	17.8	3.9	49.1	0.1	0.4	3.4	44.7	53.6
	2003	720	65.1	6.1	5.8	7.9	13.1	1.9	65.1	0.1	0.0	1.8	61.6	69.3
Italy	1999	3328	48.5	5.0	6.7	12.5	25.2	2.0	48.5	0.0	0.0	2.0	44.4	52.8
	2003	718	50.6	4.6	5.0	10.6	24.2	5.0	50.6	0.7	0.3	4.0	46.5	54.9
Japan	1999	4735	79.6	2.5	4.8	5.3	7.7	0.2	79.6	0.0	0.0	0.1	80.4	78.9
	2003	806	78.9	3.1	4.2	6.1	6.7	1.0	78.9	0.0	0.1	0.9	79.5	78.4
Jordan	1999	5040	38.5	7.4	10.2	18.0	24.4	1.5	38.5	0.4	0.0	1.1	35.6	41.1
	2003	759	31.9	10.0	15.3	17.0	22.4	3.4	31.9	0.1	0.0	3.3	34.3	29.4
Korea, Rep. of	1999	6113	80.4	1.6	3.3	4.7	9.8	0.1	80.4	0.0	0.0	0.1	79.7	81.2
	2003	890	82.4	1.2	3.5	3.7	9.2	0.0	82.4	0.0	0.0	0.0	83.1	81.7
Latvia	1999	2870	56.2	4.6	7.0	12.8	17.4	2.0	56.2	0.3	0.0	1.7	52.0	60.7
	2003	604	61.6	4.5	7.0	9.1	15.1	2.8	61.6	0.0	0.0	2.8	56.3	66.8
Lithuania	1999	2359	40.4	8.4	8.0	16.9	21.5	4.8	40.4	0.0	0.0	4.8	39.3	41.7
	2003	849	48.3	8.0	7.3	12.8	20.3	3.3	48.3	0.4	0.0	2.9	44.3	52.0
Macedonia, Rep. of	1999	4022	42.2	3.7	6.8	19.8	24.0	3.5	42.2	0.3	0.0	3.2	41.2	43.2
	2003	652	37.0	5.4	7.8	19.6	18.6	11.7	37.0	1.1	0.0	10.6	36.2	37.6
Malaysia	1999	5577	72.5	3.6	4.1	8.1	11.3	0.5	72.5	0.0	0.0	0.5	73.6	71.1
	2003	879	67.8	4.4	4.9	10.2	11.3	1.4	67.8	0.2	0.0	1.1	72.5	61.4
Moldova, Rep. of	1999	3711	53.3	7.1	9.9	14.0	14.4	1.4	53.3	0.1	0.0	1.3	52.2	54.6
	2003	665	51.3	8.6	11.6	10.2	10.5	7.8	51.3	0.0	0.2	7.7	53.7	48.6
Morocco	1999	5384	19.7	17.3	20.2	19.1	16.9	6.8	19.7	1.1	0.0	5.7	20.2	19.2
	2003	516	21.1	11.8	16.7	17.4	19.4	13.6	21.1	0.2	0.2	13.2	22.3	19.6
Netherlands	1999	2957	75.1	2.6	3.3	5.8	12.6	0.7	75.1	0.0	0.0	0.7	72.0	78.4
	2003	518	81.7	2.5	2.1	4.4	8.7	0.6	81.7	0.0	0.0	0.6	81.3	81.6
New Zealand	1999	3603	57.3	2.9	6.3	12.5	20.4	0.5	57.3	0.2	0.0	0.3	57.4	57.2
	2003	629	57.4	4.5	8.3	11.0	17.8	1.1	57.4	0.0	0.0	1.1	59.3	55.1
Philippines	1999	6599	22.3	6.2	37.8	10.4	22.5	0.7	22.3	0.0	0.0	0.7	23.5	21.0
	2003	1273	23.0	8.5	35.3	10.4	21.8	1.0	23.0	0.1	0.0	0.9	22.7	23.5
Romania	1999	3425	53.8	6.0	8.8	13.4	16.3	1.6	53.8	0.5	0.0	1.1	53.0	54.6
	2003	680	45.3	8.1	12.1	14.1	15.1	5.3	45.3	0.1	0.1	5.0	44.7	45.8
Russian Federation	1999	4331	62.2	3.9	5.3	12.2	14.2	2.1	62.2	0.1	0.0	2.0	61.5	63.0
	2003	779	52.8	6.2	8.1	14.4	14.2	4.4	52.8	0.3	0.1	4.0	51.8	53.6
Singapore	1999	4966	87.5	1.4	2.1	3.2	5.5	0.3	87.5	0.0	0.0	0.3	87.7	87.4
	2003	993	86.9	2.2	2.8	3.9	3.9	0.2	86.9	0.0	0.0	0.2	86.7	87.1
Slovak Republic	1999	3490	59.1	4.4	6.7	12.3	15.5	2.0	59.1	0.0	0.0	2.0	57.5	60.8
	2003	688	56.7	5.4	8.0	12.1	13.2	4.7	56.7	0.1	0.1	4.4	52.8	60.6
South Africa	1999	8124	13.7	6.2	44.4	10.0	23.6	2.1	13.7	0.6	0.2	1.4	13.1	14.4
	2003	1480	15.5	9.3	37.2	9.8	19.4	8.9	15.5	2.2	0.8	5.9	14.7	15.9
Tunisia	1999	5037	35.1	7.6	10.8	17.8	23.1	5.6	35.1	1.3	0.0	4.2	30.9	39.5
	2003	827	26.5	10.3	10.6	14.5	21.6	16.4	26.5	1.9	0.2	14.3	22.9	30.5
United States	1999	8985	57.2	3.3	6.6	11.4	21.2	0.3	57.2	0.0	0.0	0.3	53.2	61.2
	2003	1510	60.6	3.7	7.0	10.3	17.9	0.5	60.6	0.0	0.1	0.5	55.2	66.6
International Avg.	1999	.	54.7	5.0	9.3	11.8	17.4	1.9	54.7	0.2	0.0	1.7	53.2	56.2

DIFF = Item Difficulty Other Incorrect = Sum of invalid, not reached and omitted
Because of missing gender information, some totals may appear inconsistent

to compare performance from one administration to the next. At the same time, the display can be used to detect item-by-country interactions across all countries.

10.3 Scoring Reliability

About one-third of the items in the TIMSS 2003 assessment were constructed-response items, comprising nearly half of the score points for the assessment.[5] An essential requirement for use of such items is that they be reliably scored by all participants. That is, a particular student response should receive the same score, regardless of the scorer. In conducting TIMSS 2003, measures taken to ensure that the constructed-response items were scored reliably in all countries included developing scoring guides for each constructed-response question (which provided descriptions of acceptable responses for each score point value),[6] and providing extensive training in the application of the scoring guides. Scoring procedures for organizing and monitoring the scoring sessions were outlined in the *TIMSS 2003 Survey Operations Manual* (TIMSS, 2002).

10.3.1 Within-Country Scoring Reliability

To gather and document information about the within-country agreement among scorers, a random sample of at least 200 student responses to each item was selected to be scored independently by two scorers.[7] The inter-rater agreement for each item in each country was examined as part of the item review process. The average and range of the within-country exact percent of agreement across all items is presented in Exhibit 10.6 for mathematics and Exhibit 10.7 for science at both grades. Agreement across items was high – on average across countries, exact percent agreement was 99 percent at both grades in mathematics and 97 percent at the eighth grade and 96 at the fourth grade in science. All countries had an average exact percent agreement above 96 percent at the eighth grade and 97 at the fourth grade in mathematics and above 90 percent at the eighth grade and 91 at the fourth grade in science.

10.3.2 Trend Item Scoring Reliability

The double scoring of a sample of the student test booklets provided a measure of the consistency within each country with which constructed-response questions were scored. TIMSS 2003 also took steps to show that those constructed-response items from 1999 that were used in 2003 were scored in the same way in both assessments. In anticipation of this, countries that participated in TIMSS 1999 sent samples of scored student booklets from the 1999

5 For details on the development of the TIMSS 2003 assessment items, see Chapter 2.

6 Discussion of the development of the scoring guides for constructed-response items is provided in Chapter 2.

7 Since individual items appear in at least two booklets, 100 of each of the 12 booklets were chosen randomly for double-scoring. For a sample of 4500, this amounts to almost 25% of the total sample.

Exhibit 10.4 **Example Item-by-Country Interaction Display for Item M012040**

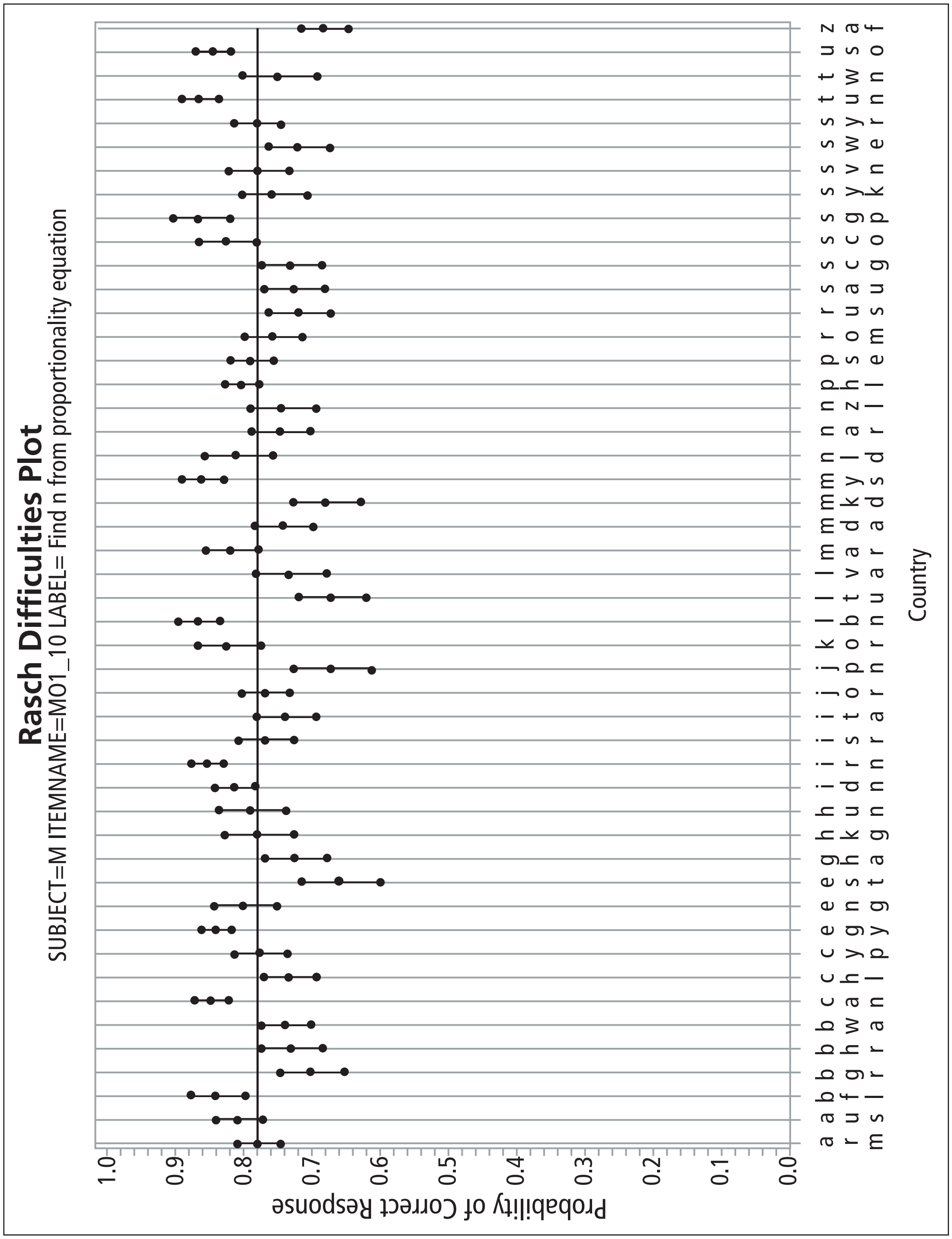
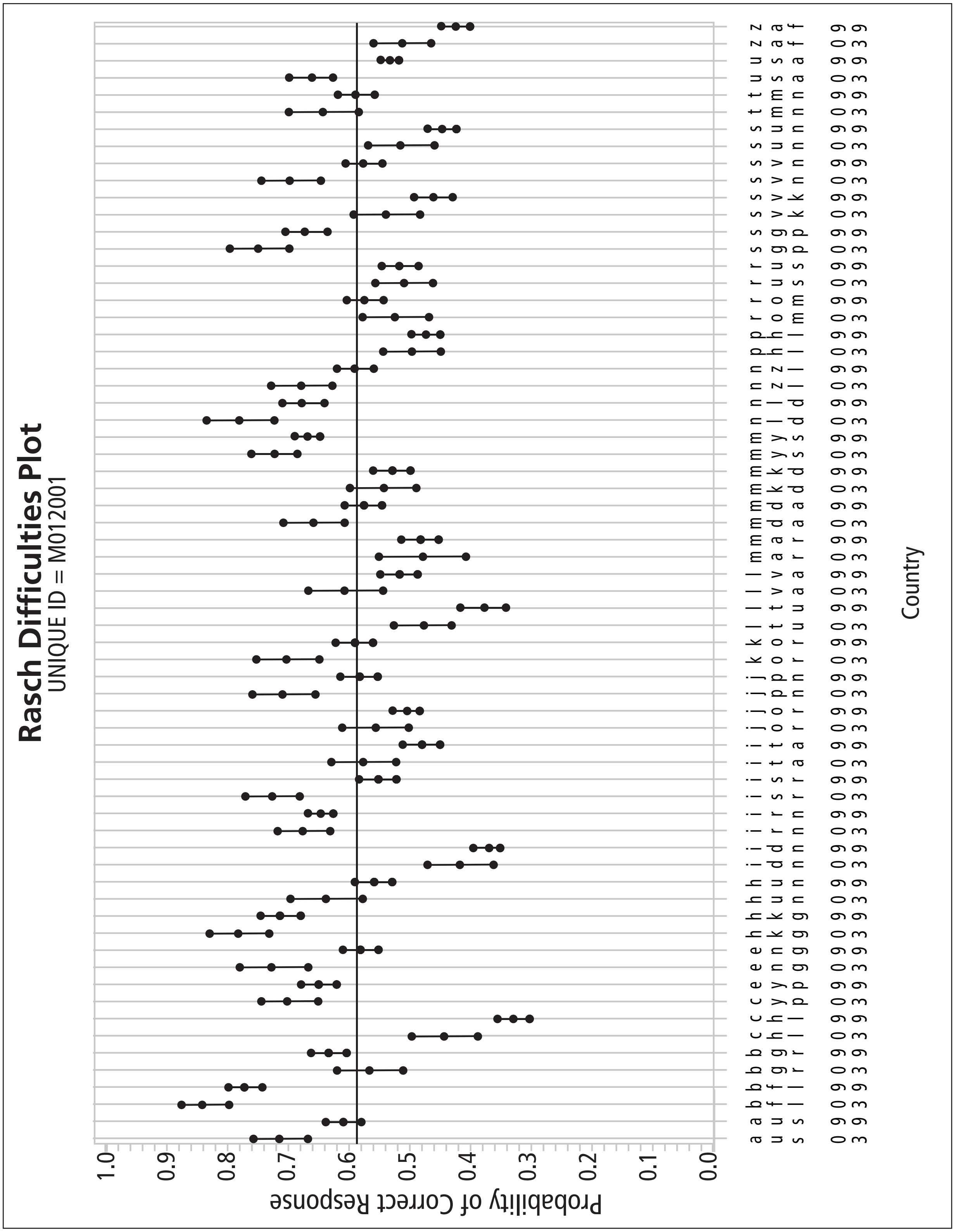


Exhibit 10.5 **Example Item-by-Country Interaction Display for Trend Item M012001**



eighth-grade data collection to the IEA Data Processing Center, where they were digitally scanned and stored in presentation software for later use. As a check on scoring consistency from 1999 to 2003, staff members working in each country on scoring the 2003 eighth-grade data were asked also to score these 1999 responses using the DPC software. The items from 1995 that were used in TIMSS 2003 all were in multiple-choice format, and therefore scoring reliability was not an issue. As shown in Exhibit 10.8 for mathematics and Exhibit 10.9 for science, there was a very high degree of scoring consistency, with 98 percent exact agreement in mathematics, on average, internationally, between the scores awarded in 1999 and those given by the 2003 scorers and 92 percent in science. There also was high agreement at the diagnostic score level, with 93 percent exact agreement, on average, in mathematics and somewhat less, 81 percent, in science.

10.3.3 Cross-Country Scoring Reliability Study

Although because of the many different languages in use in TIMSS, establishing the reliability of constructed-response scoring across all countries was not feasible, TIMSS 2003 did conduct a cross-country study of scoring reliability among northern-hemisphere countries whose scorers were proficient in English.[8] A sample of student responses to a subset of the eighth-grade mathematics and science constructed-response items was provided by the English-speaking southern hemisphere countries.

A sample of 150 student responses to each of 20 mathematics items and 21 science items (41 in total, representing about one-quarter of constructed-response items at eighth grade) was collected from Australia, Botswana, New Zealand, and Singapore. This set of 6,150 student responses in English was scored independently in each country that had at least one but preferably two scorers proficient in English. In all, 37 scorers from 20 countries participated in the study. Scoring for this study took place shortly after the within-country scoring reliability activities were completed. Making all possible comparisons among scorers gave 666 comparisons for each student response to each item, and 99,900 total comparisons when aggregated across all 150 student responses to that item. Agreement across countries was defined in terms of the percentage of these comparisons that were in exact agreement. Exhibits 10.10 and 10.11 show that scorer reliability across countries was high, with the percent exact agreement averaging 96 percent across the 20 mathematics items for the correctness score and 92 percent for the diagnostic score, and averaging 87 percent across the 21 science items for the correctness score and 76 percent for the diagnostic score.

8 See Chapter 6 for further details.

10.4 Item Review Procedures

The International Study Center thoroughly reviewed the item statistics for all participating countries to ensure that items were performing comparably across countries. In particular, items with the following problems were considered for possible deletion from the international database:

- An error was detected during TIMSS 2003 translation verification but was not corrected before test administration.
- Data checking revealed a multiple-choice item with more or fewer options than in the international version.
- The item analysis showed the item to have a negative biserial, or, for an item with more than one score point, a nonmonotonic relationship between score level and total score.
- The item-by-country interaction results showed a very large negative interaction for a particular country.
- For constructed-response items, the within-country scoring reliability data showed an agreement of less than 70 percent.
- For trend items, an item performed substantially differently in 1999 compared to 2003, or an item was not included in the 1999 assessment for a particular country.

When the item statistics indicated a problem with an item, the documentation from the translation verification[9] was used as an aid in checking the test booklets. If a question remained about potential translation or cultural issues, however, then the National Research Coordinator (NRC) was consulted before deciding how the item should be treated. If a problem could be detected by the International Study Center (such as a negative point-biserial for a correct answer or too few options for a multiple-choice item), the item was deleted from the international scaling.

The checking of the TIMSS 2003 achievement data involved 696 items for 49 countries and four Benchmarking participating at both grades (approximately 37,000 item-country combinations), and resulted in the detection of very few items that were inappropriate for international comparisons. Among the few items singled out in the review process were mostly items with differences attributable to either translation or printing problems. Appendix C provides a list of deleted items as well as a list of recodes made to constructed-response item codes.

9 See Chapter 4 for a description of the process for translating and verifying the TIMSS 2003 data-collection instruments.

Exhibit 10.6 **TIMSS 2003 Within-Country Constructed-Response Scoring Reliability Mathematics Items – Eighth Grade**

Countries	Correctness Score Agreement			Diagnostic Score Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement	
		Min	Max		Min	Max
Armenia	99	94	100	98	92	100
Australia	100	97	100	99	95	100
Bahrain	99	98	100	98	91	100
Belgium (Flemish)	99	96	100	98	91	100
Botswana	99	91	100	94	81	100
Bulgaria	96	70	100	92	64	99
Chile	99	95	100	97	91	100
Chinese Taipei	100	91	100	99	91	100
Cyprus	98	86	100	96	79	100
Egypt	100	97	100	99	97	100
England	99	93	100	98	91	100
Estonia	100	98	100	99	96	100
Ghana	99	97	100	95	90	99
Hong Kong, SAR	100	98	100	99	98	100
Hungary	98	90	100	96	80	100
Indonesia	98	90	100	94	82	100
Iran, Islamic Rep. of	99	94	100	96	90	100
Israel	98	93	100	93	83	99
Italy	99	95	100	98	92	100
Japan	99	94	100	98	91	100
Jordan	99	98	100	98	92	100
Korea, Rep. of	99	87	100	98	87	100
Latvia	98	90	100	96	79	100
Lebanon	100	94	100	99	91	100
Lithuania	97	71	100	95	62	100
Macedonia, Rep. of	100	97	100	99	95	100
Malaysia	100	98	100	99	97	100
Moldova, Rep. of	100	99	100	100	99	100
Morocco	97	89	100	92	82	99
Netherlands	97	84	100	95	78	100
New Zealand	99	96	100	97	88	100
Norway	98	91	100	96	86	100
Palestinian Nat'l Auth.	99	94	100	97	88	100

Exhibit 10.6 **TIMSS 2003 Within-Country Constructed-Response Scoring Reliability Mathematics Items – Eighth Grade** *(...Continued)*

Countries	Correctness Score Agreement			Diagnostic Score Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement	
		Min	Max		Min	Max
Philippines	99	97	100	97	92	100
Romania	100	98	100	99	94	100
Russian Federation	99	95	100	97	89	100
Saudi Arabia	99	94	100	95	81	99
Scotland	99	95	100	98	92	100
Serbia	99	96	100	98	94	100
Singapore	100	98	100	100	98	100
Slovak Republic	100	98	100	99	96	100
Slovenia	97	86	100	94	75	100
South Africa	99	95	100	97	90	99
Sweden	98	89	100	95	84	99
Tunisia	98	89	100	95	78	99
United States	97	86	100	94	75	99
International Avg.	**99**	**92**	**100**	**97**	**87**	**100**
Benchmarking Participants						
Basque Country, Spain	98	87	100	96	83	100
Indiana State, US	98	88	100	95	76	100
Ontario Province, Can.	97	80	100	93	72	100
Quebec Province, Can.	97	81	100	94	79	100

Exhibit 10.6 **TIMSS 2003 Within-Country Constructed-Response Scoring Reliability Mathematics Items – Fourth Grade**

Countries	Correctness Score Agreement			Diagnostic Score Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement	
		Min	Max		Min	Max
Armenia	99	98	100	98	95	100
Australia	100	98	100	99	97	100
Belgium (Flemish)	100	96	100	98	87	100
Chinese Taipei	99	83	100	97	76	100
Cyprus	98	91	100	95	82	100
England	99	91	100	98	90	100
Hong Kong, SAR	100	98	100	99	87	100
Hungary	98	91	100	95	78	100
Iran, Islamic Rep. of	100	98	100	99	96	100
Italy	98	92	100	96	81	100
Japan	99	95	100	98	94	100
Latvia	98	87	100	96	78	100
Lithuania	97	77	100	94	69	100
Moldova, Rep. of	100	100	100	100	100	100
Morocco	98	93	100	94	86	98
Netherlands	97	86	100	94	73	100
New Zealand	99	94	100	96	85	100
Norway	99	95	100	97	92	100
Philippines	99	96	100	97	91	100
Russian Federation	100	97	100	99	96	100
Scotland	99	98	100	98	93	100
Singapore	100	99	100	100	99	100
Slovenia	98	84	100	96	73	100
Tunisia	97	89	100	91	77	98
United States	97	88	100	95	82	100
International Avg.	**99**	**92**	**100**	**97**	**86**	**100**
Benchmarking Participants						
Indiana State, US	99	92	100	96	83	100
Ontario Province, Can.	98	87	100	96	84	100
Quebec Province, Can.	98	92	100	96	86	100

Exhibit 10.7 **TIMSS 2003 Within-Country Constructed-Response Scoring Reliability Science Items – Eighth Grade**

Countries	Correctness Score Agreement			Diagnostic Score Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement	
		Min	Max		Min	Max
Armenia	98	92	100	97	90	100
Australia	99	94	100	97	89	100
Bahrain	98	94	100	95	85	100
Belgium (Flemish)	97	89	100	93	83	100
Botswana	95	74	100	87	74	97
Bulgaria	91	72	99	84	64	99
Chile	97	91	100	94	89	99
Chinese Taipei	99	97	100	98	86	100
Cyprus	96	87	100	91	80	99
Egypt	100	98	100	100	97	100
England	98	92	100	96	85	100
Estonia	99	97	100	98	88	100
Ghana	98	93	100	93	83	99
Hong Kong, SAR	99	97	100	97	92	100
Hungary	96	87	100	92	83	100
Indonesia	96	87	100	86	68	99
Iran, Islamic Rep. of	98	87	100	95	84	100
Israel	95	89	100	84	66	98
Italy	98	91	100	96	90	100
Japan	97	81	100	93	80	100
Jordan	99	97	100	96	91	100
Korea, Rep. of	98	84	100	95	74	100
Latvia	94	78	100	87	50	100
Lebanon	100	98	100	99	95	100
Lithuania	90	69	100	82	58	100
Macedonia, Rep. of	99	96	100	97	92	100
Malaysia	99	98	100	99	97	100
Moldova, Rep. of	100	99	100	100	99	100
Morocco	94	86	100	86	69	95
Netherlands	90	70	100	84	61	100
New Zealand	98	92	100	93	84	100
Norway	95	83	100	91	80	100
Palestinian Nat'l Auth.	95	82	100	87	69	99
Philippines	98	89	100	94	83	99
Romania	99	96	100	98	94	100
Russian Federation	99	92	100	98	91	100

Exhibit 10.7 **TIMSS 2003 Within-Country Constructed-Response Scoring Reliability Science Items – Eighth Grade** *(...Continued)*

Countries	Correctness Score Agreement			Diagnostic Score Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement	
		Min	Max		Min	Max
Saudi Arabia	97	87	100	91	68	99
Scotland	97	89	100	94	85	100
Serbia	99	94	100	98	92	100
Singapore	100	99	100	99	98	100
Slovak Republic	99	95	100	97	89	100
Slovenia	90	70	100	81	61	100
South Africa	99	94	100	96	88	99
Sweden	92	76	100	85	68	99
Tunisia	98	90	100	94	73	100
United States	92	72	100	83	68	99
International Avg.	**97**	**88**	**100**	**92**	**80**	**99**
Benchmarking Participants						
Basque Country, Spain	96	87	100	92	79	100
Indiana State, US	94	82	100	87	67	100
Ontario Province, Can.	91	77	100	83	62	98
Quebec Province, Can.	92	80	100	84	66	100

Exhibit 10.7 **TIMSS 2003 Within-Country Constructed-Response Scoring Reliability Science Items – Fourth Grade**

Countries	Correctness Score Agreement			Diagnostic Score Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement	
		Min	Max		Min	Max
Armenia	99	97	100	97	91	100
Australia	99	94	100	98	91	100
Belgium (Flemish)	99	89	100	95	86	100
Chinese Taipei	98	89	100	96	89	100
Cyprus	94	76	100	89	75	99
England	98	87	100	96	86	100
Hong Kong, SAR	99	97	100	97	89	100
Hungary	95	80	100	91	78	100
Iran, Islamic Rep. of	96	85	100	93	83	99
Italy	94	77	100	90	77	100
Japan	97	86	100	94	83	100
Latvia	96	82	100	92	71	99
Lithuania	93	81	100	86	50	99
Moldova, Rep. of	100	100	100	100	100	100
Morocco	97	93	100	92	78	99
Netherlands	91	71	99	84	70	99
New Zealand	97	86	100	92	83	99
Norway	97	85	100	93	84	100
Philippines	97	89	100	91	77	99
Russian Federation	99	98	100	99	96	100
Scotland	98	90	100	96	85	100
Singapore	100	99	100	99	97	100
Slovenia	91	74	100	85	69	100
Tunisia	93	79	100	82	68	96
United States	93	70	100	86	68	99
International Avg.	**96**	**85**	**100**	**92**	**80**	**99**
Benchmarking Participants						
Indiana State, US	95	76	100	92	62	100
Ontario Province, Can.	95	80	100	90	75	100
Quebec Province, Can.	95	81	100	89	72	99

Exhibit 10.8 **TIMSS 2003 Trend Item Scoring Reliability Mathematics Items – Eighth Grade**

Countries	Correctness Score Agreement			Diagnostic Score Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement	
		Min	Max		Min	Max
Australia	98	88	100	94	73	100
Belgium (Flemish)	98	92	100	94	78	100
Bulgaria	99	82	100	94	71	100
Chile	99	97	100	92	73	100
Chinese Taipei	98	95	100	94	79	100
Cyprus	98	91	100	94	79	100
Hong Kong, SAR	98	91	100	96	84	100
Hungary	98	89	100	95	86	100
Indonesia	98	90	100	93	60	100
Iran, Islamic Rep.	98	83	100	89	24	99
Israel	98	91	100	92	74	100
Italy	99	91	100	97	86	100
Japan	98	87	100	96	76	100
Jordan	99	96	100	96	87	100
Korea, Rep. of	98	88	100	94	67	100
Latvia	90	34	100	78	32	100
Lithuania	98	93	100	94	74	100
Macedonia, Rep. of	99	85	100	96	70	100
Malaysia	99	91	100	95	84	100
New Zealand	99	96	100	94	85	100
Philippines	99	86	100	95	75	100
Romania	99	97	100	97	90	100
Russian Federation	98	94	100	92	62	100
Singapore	99	96	100	98	89	100
Slovak Republic	93	54	100	87	50	99
Slovenia	99	95	100	95	81	100
South Africa	99	92	100	93	47	100
United States	98	91	100	94	76	100
International Avg.	**98**	**88**	**100**	**93**	**72**	**100**
Benchmarking Participants						
Ontario Province, Can.	98	85	100	93	65	100
Quebec Province, Can.	98	85	100	93	65	100

Exhibit 10.9 **TIMSS 2003 Trend Item Scoring Reliability Science Items – Eighth Grade**

Countries	Correctness Score Agreement			Diagnostic Score Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement		Average of Exact Percent Agreement Across Items	Range of Exact Percent Agreement	
		Min	Max		Min	Max
Australia	93	75	100	81	56	100
Belgium (Flemish)	92	79	100	83	68	100
Bulgaria	96	87	100	83	45	100
Chile	91	80	100	77	47	100
Chinese Taipei	92	70	100	80	38	100
Cyprus	90	70	99	79	50	99
Hong Kong, SAR	89	74	100	80	58	100
Hungary	92	74	100	84	64	100
Indonesia	90	63	100	75	41	97
Iran, Islamic Rep.	92	68	100	82	55	99
Israel	93	80	100	81	46	100
Italy	94	86	100	88	73	100
Japan	92	72	100	84	62	100
Jordan	96	90	100	87	76	99
Korea, Rep. of	93	77	100	85	56	100
Latvia	79	36	100	65	21	98
Lithuania	86	66	100	74	40	100
Macedonia, Rep. of	99	89	100	98	80	100
Malaysia	92	80	100	74	35	100
New Zealand	94	87	99	79	52	98
Philippines	90	44	100	76	32	100
Romania	96	91	100	90	73	100
Russian Federation	93	80	100	79	55	99
Singapore	97	93	100	88	61	100
Slovak Republic	89	73	100	76	56	100
Slovenia	94	71	100	90	72	100
South Africa	93	71	100	79	19	100
United States	94	83	100	84	70	100
International Avg.	**92**	**75**	**100**	**82**	**54**	**100**
Benchmarking Participants						
Ontario Province, Can.	91	76	100	81	60	100
Quebec Province, Can.	91	76	100	81	60	100

Exhibit 10.10 Cross-Country Constructed-Response Scoring Reliability Data for Mathematics Items

Item Label	Total Valid Comparisons	Exact Percent Agreement	
		Correctness Score Agreement	Diagnostic Score Agreement
M022202	99900	99	98
M022156	99900	99	91
M022012	99900	94	86
M022261A	99900	99	98
M022261B	99900	99	98
M022261C	99900	90	84
M022227A	99900	99	99
M022227B	99900	97	90
M022227C	99900	94	86
M022234A	99900	95	88
M022234B	99900	91	87
M022110	99900	98	93
M032691	99900	98	94
M032640	99900	93	93
M032683	99900	92	85
M032681A	99900	99	99
M032681B	99900	99	98
M032681C	99900	97	97
M032233	99900	93	91
M032692	99900	95	95
Average		**96**	**92**

Exhibit 10.11 **Cross-Country Constructed-Response Scoring Reliability Data for Science Items**

Item Label	Total Valid Comparisons	Exact Percent Agreement	
		Correctness Score Agreement	Diagnostic Score Agreement
S032202	99900	83	73
S022283	99900	93	86
S022154	99900	83	70
S022191	99900	94	83
S022088A	99900	83	72
S022088B	99900	76	61
S022286	99900	91	77
S032625A	99900	97	94
S032625B	99900	92	72
S032120A	99900	78	61
S032120B	99900	87	69
S032063	99900	81	73
S032306	99900	88	83
S032640	99900	89	79
S032272	99900	95	88
S032650A	99900	90	84
S032650B	99900	87	80
S032056	99900	88	74
S032369	99900	80	71
S032565	99900	90	78
S032516	99900	84	74

10.5 Item Position in Booklet

As described in Chapter 2, TIMSS has a complicated student booklet design. Although each student completes just one booklet, there are 12 different student booklets at each grade level, with six blocks of mathematics and science items in each booklet. As illustrated in Exhibit 10.12, blocks of items appear in different positions in different booklets. For example, the items in block M1 appear as the first block in Booklet 1, as the second block in Booklet 6, and as the third block in Booklet 12. This allows the booklets to be linked together efficiently, but also to monitor and counterbalance any position effect.

An important step in the item review process, made possible by the counterbalanced booklet design, was to compare the characteristics of item blocks appearing in different booklet positions to detect any position effect. As the item statistics for each country were reviewed during this step, it became apparent that there was indeed an unexpectedly strong position effect in the data. As may be seen from Exhibit 10.13, this position effect occurred because some students in all countries did not reach all the items in the third block position, which was the end of the first half of each booklet before the break. The same effect was evident for the sixth block position, which was the last block in the booklets.

As described in Chapter 11, TIMSS addressed this problem using IRT scaling by treating items in the third and sixth block positions as if they were unique, even though they also appeared in other positions. For example, the mathematics items in block M1 from Booklet 1 (the first position) and from Booklet 6 (second position) were considered to be the same items for scaling and reporting purposes, but those in Booklet 12 (the third position) were scaled as items that were different and unique. This approach allowed all student responses to all items to be included in the calibration of the IRT scale and in estimating student achievement scores, while taking into account the booklet position effect. However, because items in blocks appearing in the third and sixth booklet positions were judged to have different properties to those same items when appearing in positions one, two, four, and five, student responses to items in positions three and six were not included when computing percent correct for individual example items, item statistics for use in scale anchoring, or average percent correct for measuring trends in mathematics or science content areas (see Chapter 12).

Exhibit 10.12 **TIMSS 2003 Booklet Design (Adapted from Exhibit 2.16)**

	Part 1			Part 2		
Booklet	**Position 1**	**Position 2**	**Position 3**	**Position 4**	**Position 5**	**Position 6**
1	M01	M02	S06	S07	M05	M07
2	M02	M03	S05	S08	M06	M08
3	M03	M04	S04	S09	M13	M11
4	M04	M05	S03	S10	M14	M12
5	M05	M06	S02	S11	M09	M13
6	M06	M01	S01	S12	M10	M14
7	S01	S02	M06	M07	S05	S07
8	S02	S03	M05	M08	S06	S08
9	S03	S04	M04	M09	S13	S11
10	S04	S05	M03	M10	S14	S12
11	S05	S06	M02	M11	S09	S13
12	S06	S01	M01	M12	S10	S14

Exhibit 10.13 **Average Percent Not Reached, by Booklet Position**

	Average Percent Not Reached					
	Position 1	**Position 2**	**Position 3**	**Position 4**	**Position 5**	**Position 6**
Grade 8						
Mathematics	0.4	2.5	8.4	0.3	0.9	7.7
Science	0.5	1.2	13.2	0.4	1.0	8.3
Grade 4						
Mathematics	1.1	5.4	10.9	0.7	3.1	13.2
Science	0.7	2.3	17.0	0.7	3.1	13.3

References

TIMSS (2002), *TIMSS 2003 Survey Operations Manual*, prepared by the International Study Center, Chestnut Hill, MA: Boston College.

11

Chapter 11

Scaling Methods and Procedures for the TIMSS 2003 Mathematics and Science Scales

Eugenio J. Gonzalez, Joseph Galia, and Isaac Li

11.1 Overview

As described in Chapter 1, the TIMSS 2003 goals of broad coverage of the mathematics and science curriculum and of measuring trends across assessments necessitated a complex matrix-sampling booklet design,[1] with individual students responding to just a subset of the mathematics and science items in the assessment, and not the entire assessment item pool. Given the complexities of the data collection and the need to have student scores on the entire assessment for analysis and reporting purposes, TIMSS 2003 relied on Item Response Theory (IRT) scaling to describe student achievement on the assessment and to provide accurate measures of trends from previous assessments. The TIMSS IRT scaling approach used multiple imputation or "plausible values" methodology to obtain proficiency scores in mathematics and science for all students, even though each student responded to only a part of the assessment item pool. To enhance the reliability of the student scores, the TIMSS scaling combined student responses to the items they were administered with information about students' backgrounds, a process known as "conditioning."

This chapter first reviews the psychometric models and the conditioning and multiple imputation or "plausible values" methodology used in scaling the TIMSS 2003 data, and then describes how this approach was applied to the TIMSS 2003 data and to the data from the previous TIMSS 1999 and TIMSS 1995 studies, in order to measure trends in achievement. The TIMSS

1 The TIMSS 2003 achievement test design is described in Chapter 2.

scaling was conducted at the TIMSS & PIRLS International Study Center at Boston College, using software from Educational Testing Service.[2]

11.2 TIMSS 2003 Scaling Methodology[3]

The IRT scaling approach used by TIMSS was developed originally by Educational Testing Service for use in the U.S. National Assessment of Educational Progress. It is based on psychometric models that were first used in the field of educational measurement in the 1950s and have become popular since the 1970s for use in large-scale surveys, test construction, and computer adaptive testing.[4] This approach also has been used to scale IEA's PIRLS data to measure progress in reading literacy.

Three distinct scaling models, depending on item type and scoring procedure, were used in the analysis of the TIMSS 2003 assessment data. Each is a "latent variable" model that describes the probability that a student will respond in a specific way to an item in terms of the respondent's proficiency, which is an unobserved or "latent" trait, and various characteristics (or "parameters") of the item. A three-parameter model was used with multiple-choice items, which were scored as correct or incorrect, and a two-parameter model for constructed-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 constructed-response items, i.e., those with more than two score points.

11.2.1 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:

$$P(x_i = 1 \mid \theta_k, a_i, b_i, c_i) = c_i + \frac{1 - c_i}{1 + \exp(-1.7 a_i (\theta_k - b_i))} \equiv P_{i1}(\theta_k) \tag{1}$$

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);

2 TIMSS is indebted to Matthias Von Davier, Ed Kulick, and John Barone of Educational Testing Service for their advice and support.

3 This section describing the TIMSS scaling methodology has been adapted with permission from the TIMSS 1999 Technical Report (Yamamoto and Kulick, 2000).

4 For a description of IRT scaling see Birnbaum (1968); Lord and Novick (1968); Lord (1980); Van Der Linden and Hambleton (1996). The theoretical underpinning of the imputed value methodology was developed by Rubin (1987), applied to large-scale assessment by Mislevy (1991), and studied further by Mislevy, Johnson and Muraki (1992) and Beaton and Johnson (1992). The procedures used in TIMSS have been used in several other large-scale surveys, including Progress in Reading Literacy Study (PIRLS), the U.S. National Assessment of Educational Progress (NAEP), the U.S. National Adult Literacy Survey (NALS), the International Adult Literacy Survey (IALS), and the International Adult Literacy and Life Skills Survey (IALLS).

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 chances of respondents of very low proficiency selecting the correct answer.

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

$$P_{i0} \equiv P(x_i = 0 \mid \theta_k, a_i, b_i, c_i) = 1 - P_{i1}(\theta_k) \tag{2}$$

The two-parameter (2PL) model was used for the short constructed-response items that were scored as correct or incorrect. The form of the 2PL model is the same as Equations (1) and (2) with the c_i parameter fixed at zero.

11.2.2 The IRT Model for Polytomous Items

In TIMSS 2003, as in TIMSS 1995 and TIMSS 1999, constructed-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:

$$P(x_i = 1 \mid \theta_k, a_i, b_i, d_{i,1}, \ldots, d_{i,m_i-1}) = \frac{\exp\left(\sum_{\nu=0}^{l} 1.7a_i(\theta_k - b_i + d_{i,\nu})\right)}{\sum_{g=0}^{m_i-1} \exp\left(\sum_{\nu=0}^{g} 1.7a_i(\theta_k - b_i + d_{i,\nu})\right)} \equiv P_{il}(\theta_k)$$

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 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.

Indeterminacy of model parameters of the polytomous model are resolved by setting $d_{i,0}$ =0 and setting $\sum_{i=1}^{m_i-1} d_{i,j} = 0$.

For all of the IRT models there is a linear indeterminacy between the values of item parameters and proficiency parameters, i.e., mathematically equivalent but different values of item parameters can be estimated on an arbitrarily linearly transformed proficiency scale. This linear indeterminacy can be resolved by setting the origin and unit size of the proficiency scale to arbitrary constants, such as a mean of 500 with a standard deviation of 100, as was done for TIMSS in 1995. The indeterminacy is most apparent when the scale is set for the first time.

IRT modeling relies on a number of assumptions, the most important being conditional independence. Under this assumption, item response probabilities depend only on θ_k (a measure of person proficiency) and the specified parameters of the item, and are unaffected by the demographic characteristics or unique experiences of the respondents, the data collection conditions, or the other items presented in the test. Under this assumption, the joint probability of a particular response pattern x across a set of n items is given by:

$$P(x \mid \theta_k, item\ parameters) = \prod_{i=1}^{n} \prod_{l=0}^{m_i-1} P_{il}(\theta_k)^{u_{il}}$$

where $P_{il}(\theta_k)$ is of the form appropriate to the type of item (dichotomous or polytomous), m_i is equal to 2 for the dichotomously scored items and is equal to 3 for the polytomous items, and u_{il} is an indicator variable defined by

$$u_{il} = \begin{cases} 1 \ \ if\ \ response\ \ x_i\ \ is\ in\ category\ l \\ 0\ otherwise \end{cases}$$

Replacing the hypothetical response pattern with the real scored data, the above function can be viewed as a likelihood function to be maximized by a given set of item parameters. In TIMSS 2003 analyses, estimates of both dichotomous and polytomous item parameters were obtained using the commercially available Parscale software (Muracki & Bock, 1991; version 4.1). The item parameters for 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 θ_s for each respondent.

11.2.3 Proficiency Estimation Using Plausible Values

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, whether 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 to see achievement tests designed to provide information on individual students that 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 that used in TIMSS. This design solicits relatively few responses from each sampled respondent while maintaining a wide range of content representation when responses are aggregated 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 generating multiple imputed scores, called 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).

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 θ represent the proficiency of interest. If θ were known for all sampled students, it would be possible to compute a statistic $t(\theta,y)$, 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 θ as "missing data" and approximate $t(\theta,y)$ by its expectation given (x,y), the data that actually were observed, as follows:

$$t^*(x,y) = E\left\lfloor t(\underline{\theta}, \underline{y}) \mid \underline{x}, \underline{y} \right\rfloor$$

$$= \int t(\underline{\theta}, \underline{y})\, p(\underline{\theta} \mid \underline{x}, \underline{y})\, d\underline{\theta}$$

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 TIMSS, NAEP, NALS, and 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 uncertainly associated with imputation can be quantified by "multiple imputation". For example, the average of multiple estimates of t, each computed from a different set of plausible values, is a numerical approximation of t^* of the above equation; the variance among them 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. That variability is estimated separately by jackknife variance estimation procedures, which are discussed in Chapter 12.

Note that plausible values are not test scores for individuals in the usual sense, but rather are imputed values that may be used to estimate population characteristics correctly. When the underlying model is correctly specified, plausible values will provide consistent estimates of population characteristics, even though they are not generally unbiased estimates of the proficiencies of the individuals with whom they are associated.[5]

Plausible values for each respondent j are drawn from the conditional distribution $P(\theta_j|x_j,y_j,\Gamma,\Sigma)$, 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

$$P(\theta_j|\, x_j, y_j, \Gamma, \Sigma) \propto P(x_j|\,\theta_j, y_j, \Gamma, \Sigma) P(\theta_j|\, y_j, \Gamma, \Sigma) = P(x_j|\,\theta_j) P(\theta_j|\, y_j, \Gamma, \Sigma) \tag{3}$$

where θ_j is a vector of scale values, $P(x_j,\theta_j)$ is the product over the scales of the independent likelihoods induced by responses to items within each scale, and $P(\theta_j|\, y_j, \Gamma, \Sigma)$ is the multivariate joint density of proficiencies for 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.

5 For further discussion, see Mislevy, Beaton, Kaplan, and Sheehan (1992).

11.2.4 Conditioning

A multivariate normal distribution was assumed for $P(\theta_j | y_j, \Gamma, \Sigma)$, 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 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 percent 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.

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

where ε is normally distributed with mean zero and variance Σ. As in a regression analysis, Γ is a matrix each of whose columns is the effects for each scale and Σ is the matrix of residual variance between scales.

Note that in order to be strictly correct for all functions Γ of θ, it is necessary that $P(\theta | y)$ be correctly specified for all background variables in the survey. Estimates of functions Γ involving background variables not conditioned on in this manner are subject to estimation error due to misspecification. The nature of these errors was discussed in detail in Mislevy (1991). In TIMSS 2003, however, principal component scores based on nearly all background variables were used. Those selected variables were chosen to reflect high relevance to policy and to education practices. The computation of marginal means and percentile points of θ for these variables is nearly optimal.

The basic method for estimating Γ and Σ with the Expectation and Maximization (EM) procedure is described in Mislevy (1985) for a single scale case. The EM algorithm requires the computation of the mean, θ, and variance, Σ, of the posterior distribution in equation (3).

11.2.5 Generating Proficiency Scores

After completing the EM algorithm, plausible values for all sampled students are drawn from the joint distribution of the values of Γ in a three-step process. First, a value of Γ is drawn from a normal approximation to $P(\Gamma, \Sigma | x_j, y_j)$ that fixes 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_j, and variance Σ_j^p of the posterior distribution in equation (3), where p is the number of scales, are computed using the methods applied in the EM algorithm. In the third step, the proficiency values are drawn independently from a multivariate normal distribution with mean θ_j and variance Σ_j^p. These three steps are repeated five times, producing five imputations of θ_j for each sampled respondent.

For respondents with an insufficient number of responses, the Γ and Σs described in the previous paragraph are fixed. Hence, all respondents - regardless of the number of items attempted - are assigned a set of plausible values.

The plausible values could then be employed to evaluate equation (1) for an arbitrary function *T* as follows:

1. Using the first vector of plausible values for each respondent, evaluate *T* as if the plausible values were the true values of θ. Denote the result T_1.
2. Evaluate the sampling variance of *T*, or Var(T_1), with respect to respondents' first vectors of plausible values.
3. Carry out steps 1 and 2 for the second through fifth vectors of plausible values, thus obtaining T_u and Var_u for u = 2, . . ., 5.
4. The best estimate of *T* obtainable from the plausible values is the average of the five values obtained from the different sets of plausible values:
$$\hat{T} = \frac{\sum_u T_u}{5}$$
5. An estimate of the variance of $\hat{T}$ is the sum of two components: an estimate of Var(T_u) obtained by averaging as in step 4, and the variance among the T_us. Let $\overline{U} = \frac{\sum_u Var_u}{M}$, and let $B_M = \frac{\sum_u \left(T_u - \hat{T}\right)^2}{M-1}$ be the variance among the M plausible values. Then the final estimate of the variance of $\hat{T}$ is:
$$Var\left(\hat{T}\right) = \overline{U} + \left(1 + M^{-1}\right) B_M$$

The first component in $Var\left(\hat{T}\right)$ reflects uncertainty due to sampling respondents from the population; the second reflects uncertainty due to the fact that sampled respondents' θs are not known precisely, but only indirectly through *x* and *y*.

11.2.6 Working with Plausible Values

Plausible values methodology was used in TIMSS 2003 to ensure the accuracy of estimates of the proficiency distributions for the TIMSS population as a whole and particularly for comparisons between subpopulations. A further advantage of this method is that the variation between the five plausible values generated for each respondent reflects the uncertainty associated with proficiency estimates for individual respondents. However, retaining this component of uncertainty requires that additional analytical procedures be used to estimate respondents' proficiencies, as follows.

If θ values were observed for all sampled respondents, the statistic $(t-T)/U^{1/2}$ would follow a t-distribution with d degrees of freedom. Then the incomplete-data statistic $(T-\hat{T})/\left[Var(\hat{T})\right]^{1/2}$ is approximately t-distributed, with degrees of freedom (Johnson & Rust, 1993) given by

$$\nu = \frac{1}{\frac{f_M{}^2}{M-1}+\frac{(1-f_M)^2}{d}}$$

where d is the degrees of freedom for the complete-data statistic, and f_M is the proportion of total variance due to not observing θ values:

$$f_M = \frac{\left(1+M^{-1}\right)B_M}{Var\left(\hat{T}\right)}$$

When B_M is small relative to $\overline{U}$, the reference distribution for the incomplete-data statistic differs little from the reference distribution for the corresponding complete-data statistics. If, in addition, d is large, the normal approximation can be used instead of the t-distribution.

For k-dimensional t, such as the k coefficients in a multiple regression analysis, each U and $\overline{U}$ is a covariance matrix, and B_M is an average of squares and cross-products rather than simply an average of squares. In this case, the quantity $\left(\underline{T}-\underline{\hat{T}}\right)Var^{-1}\left(\underline{\hat{T}}\right)\left(\underline{T}-\underline{\hat{T}}\right)'$ is approximately F-distributed with degrees of freedom equal to k and ν, with ν defined as above but with a matrix generalization of f_M:

$$f_M = \left(1+M^{-1}\right)Trace\left[B_M Var^{-1}\left(\hat{T}\right)\right]/k\,.$$

For the same reason that the normal distribution can approximate the t distribution, a chi-square distribution with k degrees of freedom can be used in place of the F-distribution for evaluating the significance of the above quantity $\left(\underline{T}-\underline{\hat{T}}\right)Var^{-1}\left(\underline{\hat{T}}\right)\left(\underline{T}-\underline{\hat{T}}\right)'$.

Statistics $\hat{T}$, the estimates of ability conditional on responses to cognitive items and background variables, are consistent estimates of the corresponding population values T, as long as background variables are included in the conditioning variables. The consequences of violating this restriction are described by Beaton & Johnson (1990), Mislevy (1991), and Mislevy & Sheehan (1987). To avoid such biases, the TIMSS 2003 analyses included nearly all background variables.

11.3 Implementing the Scaling Procedures for the TIMSS 2003 Assessment Data

The application of IRT scaling and plausible value methodology to the TIMSS 2003 assessment data involved four major tasks: calibrating the achievement test items (estimating model parameters for each item), creating principal components from the questionnaire data for use in conditioning; generating IRT scale scores (proficiency scores) for mathematics and science and for each of the mathematics and science content domains; and placing the proficiency scale scores on the metric used to report the results from previous assessments. The TIMSS eighth-grade reporting metric was established by setting the average of the mean scores of the countries that participated in TIMSS 1995 at the eighth grade to 500 and the standard deviation to 100. To enable comparisons between 1999 and 1995, the TIMSS 1999 eighth-grade data also were placed on this metric. Placing the 2003 eighth-grade results on this metric permitted trend results from three points in time: 1995, 1999, and 2003. Since TIMSS did not collect data at the fourth grade in 1999, the TIMSS 2003 fourth-grade data were placed directly on the 1995 fourth-grade scale, providing comparisons between results from 1995 and 2003. Scale metrics were aligned for trend reporting only for mathematics and science overall; there were insufficient trend items from 1995 and 1999 to measure trends in content areas reliably.

11.3.1 Calibrating the TIMSS 2003 Test Items

As described in Chapter 2, the TIMSS 2003 achievement test design consisted of a total of 14 mathematics blocks and 14 science blocks at each grade, distributed across 12 student booklets. Each block contained either mathematics or science items, drawn from a range of content and cognitive domains. The 14 mathematics blocks were designated M01 through M14, and the 14 science blocks S01 through S14. Each student booklet contained six blocks, which were chosen according to a matrix-sampling scheme that kept the number of booklets as few as possible while maximizing the number of times blocks were paired together in a booklet. Half of the booklets contained four mathematics blocks and two science blocks, and half four science blocks and two mathematics blocks. Each sampled student completed one of the twelve student booklets. During the testing sessions, each student responded to three blocks of items, took a short break, and then responded to the other three blocks. The booklets were distributed among the students in each sampled class according to a scheme that ensured comparable random samples of students responded to each booklet.

In line with the TIMSS assessment framework, IRT scales were constructed for reporting overall student achievement in mathematics and science, as well as for reporting separately for each of the mathematics and science content domains.

The first step in constructing these scales was to estimate the IRT model item parameters for each item on each of the scales. This item calibration was conducted using the commercially-available Parscale software (Muraki & Bock, 1991; version 4.1). Item calibration for the overall mathematics and science scales, which were used to measure trends from 1995 and 1999, included data from 1995 for fourth grade and from 1999 for eighth grade. The calibration was conducted using a self-weighting random sample of 1000 students from each country's TIMSS student sample from each assessment year. This ensured that the data from each country and each assessment year contributed equally to the item calibration, while keeping the amount of data to be analyzed to a reasonable size.

Several calibrations were conducted. At the eighth grade, to construct separate overall mathematics and science scales for reporting trends, as well as performance generally in 2003, item calibrations were conducted using data from the 29 countries that participated in both 1999 and 2003 assessments. These calibrations each included 29,000 student records from the 1999 assessment and 29,000 records from the 2003 assessment, for a total of 58,000 student records. The item parameters established in these calibrations were used subsequently for estimating student scores for all 49 countries and 4 benchmarking entities that participated in 2003.

At the fourth grade, item calibrations for the overall mathematics and science scales for reporting trends, as well as performance generally in 2003, were conducted using data from the 15 countries that participated in both 1995 and 2003 assessments. These calibrations each included 15,000 student records from the 1999 assessment and 15,000 records from the 2003 assessment, for a total of 30,000 student records. As for the eighth grade, the item parameters established in these calibrations were used subsequently for estimating student scores for all 26 countries and 3 benchmarking entities that participated in 2003.

Because there were insufficient items to construct reliable scales for measuring trends in each of the content domains, scales for these domains were constructed using 2003 data only. At the eighth grade, separate calibrations were conducted for each of the five mathematics and five science content domains. These calibrations were based on 46,000 student records, 1,000 from each of the 46 countries that participated in the 2003 assessment.[6] Similarly at the fourth grade, separate calibrations were conducted for each of the five mathematics and three science content domains. These calibrations were based on 26,000 student records, 1,000 from each of the 26 countries that participated in the 2003 assessment at the fourth grade. Although, because of the matrix-sampling design, not all students responded to every item, there were at least 2,000 student responses to each item in all calibrations.

All items in the TIMSS 2003 assessment were included in the item calibrations. However, a non-trivial position effect was detected during routine quality control checks on the data. As described in Chapter 2, TIMSS has a complicated booklet design, with blocks of items appearing in different positions in different booklets. For example, the items in block M1 appear as the first block in Booklet 1, as the second block in Booklet 6, and as the third block in Booklet 12. This allows the booklets to be linked together efficiently, but also to monitor and counterbalance any position effect. The counterbalanced booklet design made it possible to detect an unexpectedly strong position effect in the data as the item statistics for each country were reviewed. More specifically, this position effect occurred because some students in all countries did not reach all the items in the third block position, which was the end of the first half of each booklet before the break. The same effect was evident for the sixth block position, which was the last block in the booklets. The IRT scaling addressed this problem by treating items in the third and sixth block positions as if they were unique, even though they also appeared in other positions. For example, the mathematics items in block M1 from Booklet 1 (the first position) and from Booklet 6 (second position) were considered to be the same items for scaling and reporting purposes, but those in Booklet 12 (the third position) were scaled as items that were different and unique. This technique is also known as “splitting” the items, or “freeing” the item parameters.

Exhibits D.1 through D.22 in Appendix D present the item parameters generated from the calibrations. Items where the parameters have been freed have an "F" in the second character position of the item label. As a by-product of the calibrations, interim scores in mathematics, science, and the content domains for use in constructing conditioning variables were produced.

11.3.2 Omitted and Not-Reached Responses

Apart from missing data on items that by design were not administered to a student, missing data could also occur because a student did not answer an item – whether because the student did not know the answer, omitted it by mistake, or did not have time to attempt the item. An item was considered

6 Data from the four Benchmarking participants were not included in the item calibration.

not reached when (within part 1 or part 2 of the booklet) the item itself and the item immediately preceding were not answered, and there were no other items completed in the remainder of the booklet.

In TIMSS 2003, 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, and that were located in positions 1, 2, 4, and 5 of the test booklet, were treated as if they had not been administered. Items that were considered not to have been reached by the students, and that were located in positions 3 and 6 of the test booklet were treated as incorrect. This approach was considered optimal for parameter estimation. However, not-reached items were always considered as incorrect responses when student proficiency scores were generated.

11.3.3 Evaluating Fit of IRT Models to the TIMSS 2003 Data

After the calibrations were completed, checks were performed to verify that the item parameters obtained from Parscale adequately reproduced the observed distribution of responses across the proficiency continuum. The fit of the IRT models to the TIMSS 2003 data was examined by comparing the theoretical item response function curves generated using the item parameters estimated from the data with the empirical item response functions calculated from the posterior distributions of the θs for each respondent that received the item.

Exhibit 11.1 shows a plot of the empirical and theoretical item response functions for a dichotomous item. In the plot, the horizontal axis represents the proficiency scale, and the vertical axis represents the probability of a correct response. Values from the theoretical curve based on the estimated item parameters are shown as crosses. Empirical results are represented by circles. The centers of the circles represent the empirical proportions correct. The plotted values are the sums of these individual posteriors at each point on the proficiency scale for those students that responded correctly to the item, plus a fraction of the omitted responses, divided by the sum of the posteriors of all that were administered the item. The size of the circles is proportional to the sum of the posteriors at each point on the proficiency scale for all of those who received the item; this is related to the number of respondents contributing to the estimation of that empirical proportion correct.

Exhibit 11.2 contains a plot of the empirical and theoretical item response functions for a polytomous item. As for the dichotomous item plot

Exhibit 11.1 **TIMSS 2003 Mathematics Assessment Example Response Function for a Dichotomous Item**

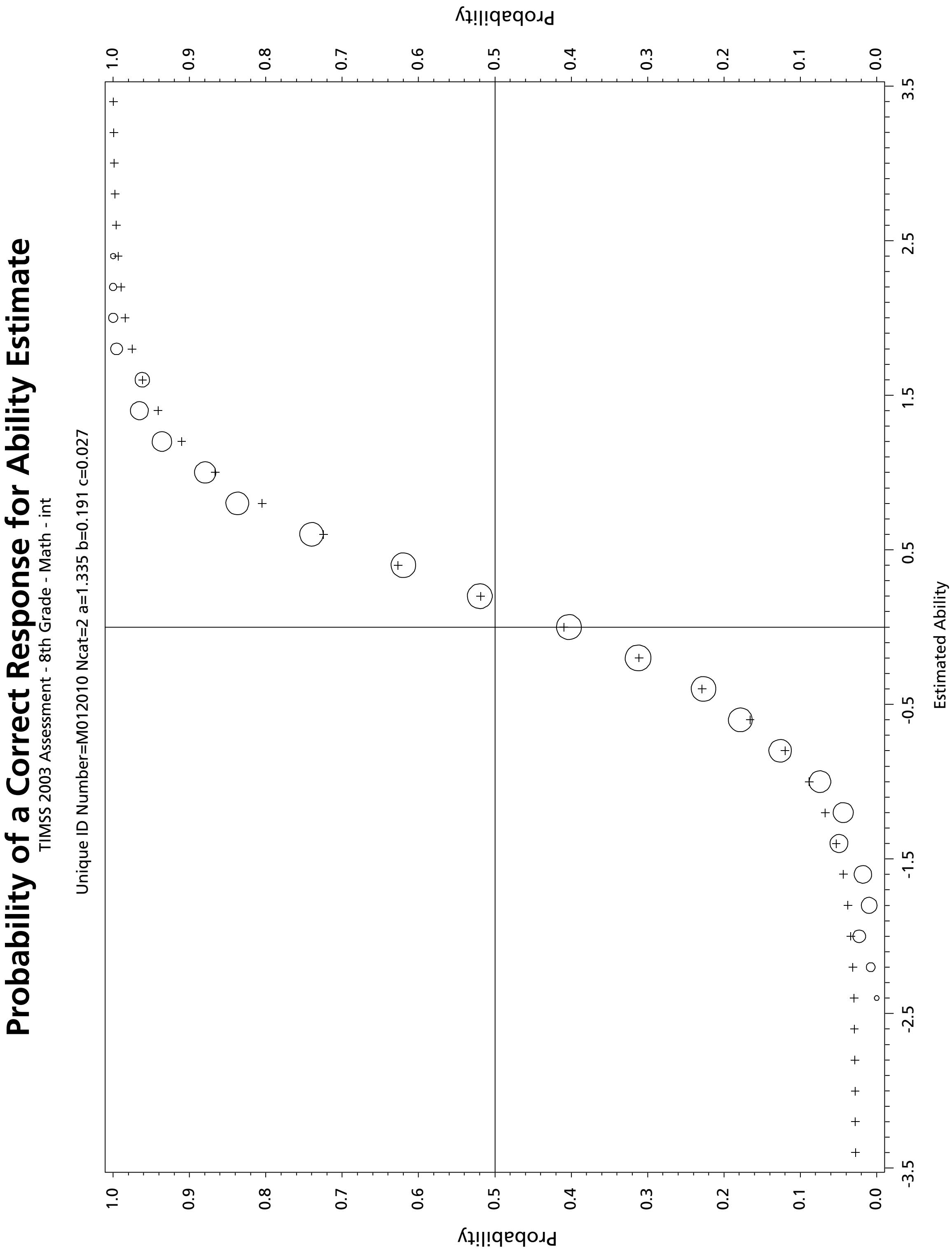
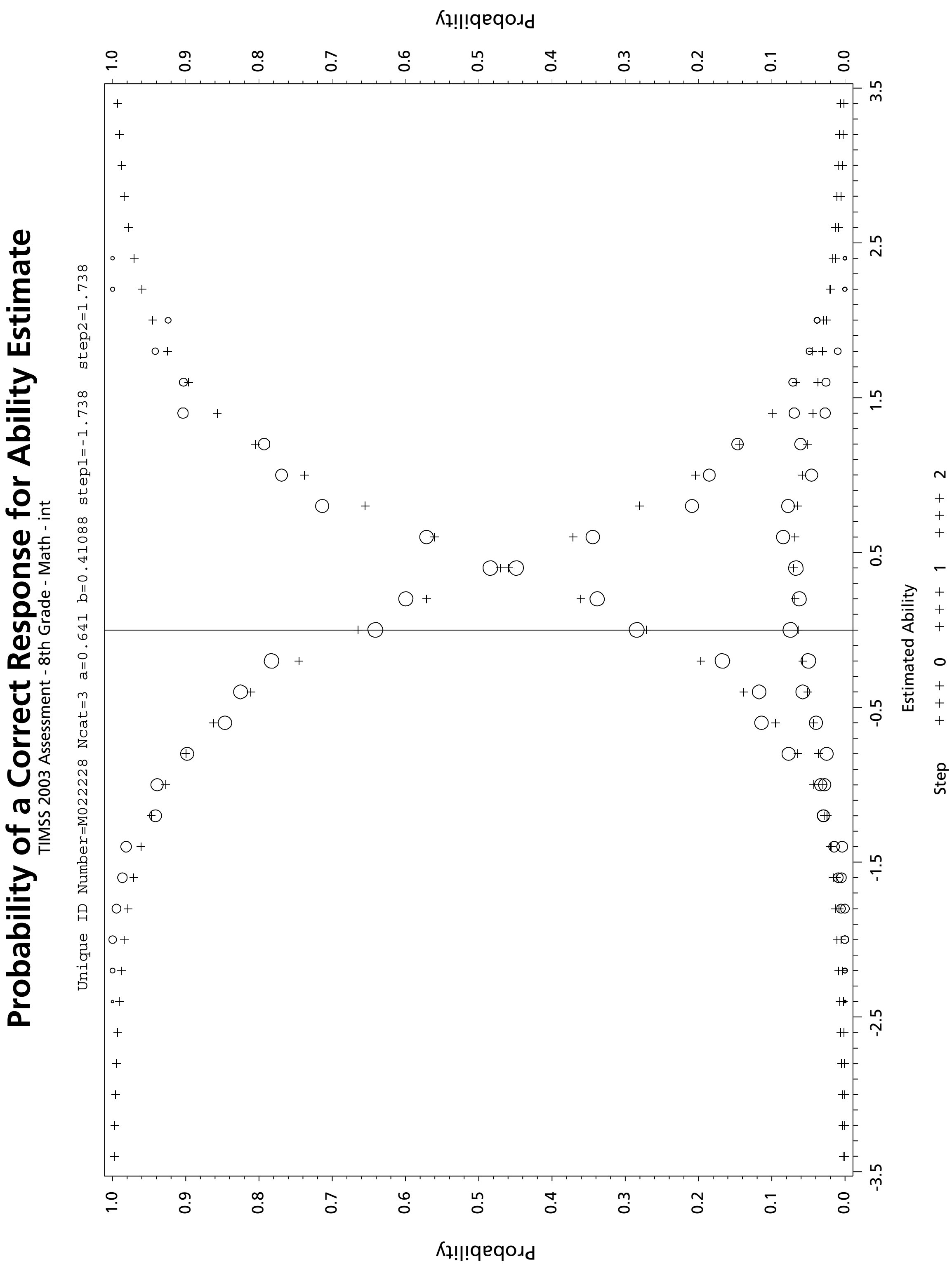


Exhibit 11.2 **TIMSS 2003 Mathematics Assessment Example Response Function for a Polytomous Item**



above, the horizontal axis represents the proficiency scale, but the vertical axis represents the probability of having a response fall in a given score category. For polytomous items, the sums for those who scored in the category of interest is divided by the sum for all those that were administered the item. The interpretation of the circles is the same as in Exhibit 11.2.

11.3.4 Variables for Conditioning the TIMSS 2003 Data

Because there were so many background variables that could be used in conditioning, TIMSS followed the practice established in other large-scale studies of using principal components analysis to reduce the number of variables while explaining most of their common variance. Principal components for the TIMSS 2003 background data were constructed as follows:

1. For categorical variables (questions with a small number of fixed response options), a "dummy coded" variable was created for each response option, with a value of one if the option was chosen and zero otherwise. If a student omitted or was not administered a particular question, all dummy coded variables associated with that question were assigned the value zero.
2. Background variables with numerous response options (such as year of birth, or number of people who live in the home) were recoded using criterion scaling.[7] This was done by replacing each response option with an interim achievement score. For the overall mathematics and science scales, the interim achievement scores were the average across the interim mathematics and science scores produced from the item calibration. For the content domain scales, the interim achievement scores from the calibration in each subject were averaged to form a composite mathematics and a composite science score, and the average of these composite scores was used as the interim achievement score.
3. Separately for each TIMSS country, all the dummy-coded and criterion-scaled variables were included in a principal components analysis. Those principal components accounting for 90 percent of the variance of the background variables were retained for use as conditioning variables. Because the principal components analysis was performed separately for each country, different numbers of principal components were required to account for 90% of the common variance in each country's background variables. Exhibit 11.3 and Exhibit 11.4 show the total number of variables that were used in the principal component analysis and the number of principal components selected to account for 90% of the background variance within each country.

7 The process of generating criterion scaled variables is described in Beaton (1969).

In addition to the principal components, student gender (dummy coded), the language of the test (dummy coded), an indicator of the classroom in the school to which the student belonged (criterion scaled), and an optional, country-specific variable (dummy coded) were included as conditioning variables.

Exhibit 11.3 **Number of Variables and Principal Components for Conditioning TIMSS 2003 Fourth Grade Data**

Country	Sample Size	Total number of conditioning variables	Total number of principal components only
ARM	5674	291	283
AUS	4321	301	216
BFL	4712	305	235
COT	4362	291	218
CQU	4350	291	217
CYP	4328	291	216
ENG	3585	295	179
HKG	4608	313	230
HUN	3319	307	165
IRN	4352	305	217
ITA	4282	311	214
JPN	4535	313	226
LTU	4422	290	221
LVA	3687	313	184
MAR	4263	297	213
MDA	3981	307	199
NLD	2937	289	146
NOR	4342	313	217
NZL	4308	311	215
PHL	4572	303	228
RUS	3963	305	198
SCO	3936	295	196
SGP	6668	301	333
SVN	3126	313	156
TUN	4334	311	216
TWN	4661	313	233
USA	9829	287	491
YEM	4205	313	210

Exhibit 11.4 **Number of Variables and Principal Components for Conditioning TIMSS 2003 Eighth Grade Data**

Country	Sample Size	Total Number of Conditioning Variables	Total Number of Principal Components Only
ARM	5726	893	286
AUS	4791	417	225
BFL	4970	762	248
BGR	4117	913	205
BHR	4199	432	209
BSQ	2514	431	125
BWA	5150	424	248
CHL	6377	416	240
COT	4217	410	210
CQU	4411	410	220
CYP	4002	897	200
EGY	7095	418	249
ENG	2830	410	141
EST	4040	903	202
GHA	5100	410	245
HKG	4972	432	233
HUN	3302	907	165
IDN	5762	897	288
IRN	4942	424	244
ISR	4318	432	215
ITA	4278	430	213
JOR	4489	432	224
JPN	4856	426	231
KOR	5309	432	234
LBN	3814	745	190
LTU	4964	811	248
LVA	3630	679	181
MAR	3160	408	158
MDA	4033	913	201
MKD	3893	919	194
MYS	5314	412	231
NLD	3065	735	153
NOR	4133	429	206
NZL	3801	430	190
PHL	6917	422	243
PSE	5357	432	251
ROM	4104	919	205
RUS	4667	912	233
SAU	4295	426	214
SCG	4296	919	214
SCO	3516	410	175

Exhibit 11.4 **Number of Variables and Principal Components for Conditioning TIMSS 2003 Eighth Grade Data** *(...Continued)*

Country	Sample Size	Total Number of Conditioning Variables	Total Number of Principal Components Only
SGP	6018	420	233
SVK	4215	912	210
SVN	3578	766	178
SWE	4256	916	212
SYR	4895	418	240
TUN	4931	410	242
TWN	5379	432	231
USA	8912	404	229
ZAF	8952	432	255

11.3.5 Generating IRT Proficiency Scores for the TIMSS 2003 Data

Educational Testing Service's MGROUP program (ETS, 1998; version 3.1)[8] was used to generate the IRT proficiency scores. This program takes as input the students' responses to the items they were given, the item parameters estimated at the calibration stage, and the conditioning variables, and generates as output the plausible values that represent student proficiency. Four MGROUP runs were conducted at each grade level using the 2003 assessment data: one unidimensional run for the overall mathematics scale, one unidimensional run for the overall science scale, one multidimensional run for the mathematics content domain scales, and one multidimensional run for the science content domain scales.

In addition to generating plausible values for the TIMSS 2003 data, the parameters estimated at the calibration stage also were used to generate plausible values on the overall mathematics and science scales using the 1999 eighth-grade data for the 29 trend countries that participated in the TIMSS 1999 eighth-grade assessment and the 1995 fourth-grade data for the 15 countries that participated in the 1995 fourth-grade assessment. These plausible values for the trend countries were called "bridge scores."

Plausible values generated by the conditioning program are initially on the same scale as the item parameters used to estimate them. This scale metric is generally not useful for reporting purposes since it is somewhat arbitrary, ranges between approximately –3 and +3, and has a mean of zero across all countries.

8 The MGROUP program was provided by ETS under contract to the TIMSS and PIRLS International Study Center at Boston College.

11.3.6 Transforming the Mathematics and Science Scores to Measure Trends from 1995 and 1999

To provide results for TIMSS 2003 that would be comparable to results from previous TIMSS' assessments, the 2003 proficiency scores (plausible values) for overall mathematics and science had to be transformed to the metric used in 1995 and 1999. To accomplish this, the means and standard deviations of the mathematics and science "bridge scores" were made to match the means and standard deviations of the scores reported in the earlier assessments by applying the appropriate linear transformations. Once the linear transformation constants had been established, all of the mathematics and science scores from the 2003 assessment were transformed by applying the same linear transformations. This provided mathematics and science student achievement scores for the TIMS 2003 assessment that were directly comparable to the scores from the 1995 and 1999 assessments.

11.3.7 Setting the Metric for the Mathematics and Science Content Domain Scales

As described earlier, the IRT scales for the mathematics and science content domains had no provision for measuring trends, and so there was no need to establish links to previous assessment metrics. Instead, the plausible values for each content domain scale were transformed to the same metric as the overall subject scale in 2003. For example, in eighth-grade mathematics, the mean and standard deviation for the number, algebra, measurement, geometry, and data scales were set to have the same mean and standard deviation as the 2003 eighth-grade mathematics scale.

References

Beaton, A.E., & Johnson, E.G. (1990). The average response method of scaling. *Journal of Educational Statistics*, 15, 9-38.

Beaton, A.E., & Johnson, E.G. (1992). Overview of the scaling methodology used in the National Assessment. *Journal of Educational Measurement*, 26(2), 163-175.

Birnbaum, A. (1968). "Some latent trait models and their use in inferring an examinee's ability" in F. M. Lord & M. R. Novick, *Statistical theories of mental test scores*. Reading, MA: Addison- Wesley Publishing.

Johnson, E.G., & Rust, K.F. (1992). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics.*

Lord, F.M.,& Novick, M.R. (1968). *Statistical theories of mental test scores*. Redding, MA: Addison-Wesley.

Lord, F.M. (1980). *Applications of items response theory to practical testing problems*. Hillsdales, NJ: Lawrence Erlbaum Associates.

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.

Mislevy, R.J., Beaton, A., Kaplan, B.A., & Sheehan, K. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29(2), 133-161.

Mislevy, R.J., Johnson, E.G. & Muraki, E. (1992). Scaling procedures in NAEP. *Journal of Educational Statistics*, 17(2), 131- 154.

Mislevy, R.J. & Sheehan, K. (1987). "Marginal estimation procedures" in A. E. Beaton (Ed.), Implementing the new design: *The* NAEP *1983-84 technical report* (pp. 293-360). (no. 15-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16(2), 159-176.

Muraki, E., & Bock, R.D. (1991). PARSCALE: *Parameter scaling of rating data*. Chicago, IL: Scientific Software, Inc.

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons.

Thomas, N. (1993). Asymptotic corrections for multivariate posterior moments with factored likelihood functions. *Journal of Computational and Graphical Statistics*, 2, 309-22.

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.

Van Der Linden, W.J. & Hambleton, R. (1996). *Handbook of Modern Item Response Theory*. New York. Springer-Verlag.

12

Chapter 12

Reporting Student Achievement in Mathematics and Science

Eugenio J. Gonzalez, Joseph Galia, Alka Arora, Ebru Erberber, and Dana Diaconu

12.1 Overview

The *TIMSS 2003 International Mathematics Report* (Mullis, Martin, Gonzalez, and Chrostowski, 2004) and the *TIMSS 2003 International Science Report* (Martin, Mullis, Gonzalez, and Chrostowski, 2004) summarize eighth- and fourth-grade students' mathematics and science achievement in each participating country. This chapter provides information about the international benchmarks established to help users of the achievement results understand the meaning of the achievement scales, and describes the scale anchoring procedure applied to describe student performance at these benchmarks. The chapter also describes the jackknifing technique employed by TIMSS to capture the sampling and imputation variances that follow from TIMSS' complex student sampling and booklet design, and describes how important statistics used to compare student achievement across the participating countries were calculated.

12.2 Describing International Benchmarks of Student Achievement on the TIMSS 2003 Mathematics and Science Scales[1]

It is important for users of TIMSS achievement results to understand what the scores on the TIMSS mathematics and science achievement scales mean. That is, what does it mean to have a scale score of 513 or 426? To describe student performance at various points along the TIMSS mathematics and science achievement scales, TIMSS used scale anchoring to summarize and describe student achievement at four points on the mathematics and science scales – Advanced International Benchmark (625), High International Bench-

1 The description of the scale anchoring procedure was adapted from Kelly (1999), Gregory and Mullis (2000), and Gonzalez and Kennedy (2003).

mark (550), Intermediate International Benchmark (475), and Low International Benchmark (400).

In brief, scale anchoring involves selecting Benchmarks (scale points) on the TIMSS achievement scales to be described in terms of student performance and then identifying items that students scoring at the anchor points (the international benchmarks) can answer correctly. The items, so identified, are grouped by content area within benchmarks for review by mathematics and science experts. For TIMSS, the Science and Mathematics Item Replacement Committee (SMIRC) conducted the review. They examined the content of each item and described the kind of mathematics or science knowledge demonstrated by students answering the item correctly. The panelists then summarized the detailed list in a brief description of performance at each anchor point. This procedure resulted in a content referenced interpretation of the achievement results that can be considered in light of the TIMSS 2003 Mathematics and Science Frameworks.

12.2.1 Identifying the Benchmarks

Identifying the scale points to serve as benchmarks has been a challenge in the context of measuring trends. For the TIMSS 1995 and 1999 assessments, the scales were anchored using percentiles. That is, the analysis was conducted using the Top 10 percent (90th percentile), the Top Quarter (75th percentile), the Top Half (50th percentile), and the Bottom Quarter (25th percentile). However, with different participating countries in each TIMSS cycle and different achievement for countries participating in previous cycles, it was pointed out by the National Research Coordinators (NRCs) that the percentile points were changing with each cycle and that stability was required.

It was clear that TIMSS needed a set of points to serve as benchmarks, that would not change in the future, that would look sensible, and that were similar to points used in 1999. After much consideration of points used in other international (IALS and PISA) and national assessments (e.g., NAEP in the United States), it was decided to use specific scale points with equal intervals as the international benchmarks. At the TIMSS Project Management Meeting in March 2004, a set of four points on the mathematics and science achievement scales was identified to be used as the international benchmarks, namely 400, 475, 550, and 625. These points were selected to be as close as possible to the percentile points anchored in 1999 at the eighth grade (i.e., Top 10% was 616 for mathematics and science, Top Quarter was 555 for mathematics and 558 for science, Top Half was 479 for mathematics and 488 for science, and Bottom Quarter was 396 for mathematics and 410 for science). The newly defined benchmark scale points were used as the basis for the scale anchoring descriptions. Exhibit 12.1 shows the scale scores representing each international benchmark for both grades in mathematics and science.

Exhibit 12.1 **TIMSS 2003 International Benchmarks for Eighth and Fourth Grade Mathematics and Science**

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

12.2.2 Identifying the Anchor Items

After selecting the benchmark points to be described on the TIMSS 2003 mathematics and science achievement scales, the first step in the scale-anchoring procedure was to establish criteria for identifying those students scoring at the international benchmarks. Following the procedure used in previous IEA studies, a student scoring within plus and minus five scale score points of a benchmark was identified for the benchmark analysis. The score ranges around each international benchmark and the number of students scoring in each range for mathematics and science are shown in Exhibit 12.2 for the eighth grade and in Exhibit 12.3 for the fourth grade. The range of plus and minus five points around a benchmark is intended to provide an adequate sample in each group, yet be small enough so that performance at each benchmark anchor point is still distinguishable from the next. The data analysis for the scale anchoring was based on these students scoring at each benchmark range.

Exhibit 12.2 **Range around Each Anchor Point and Number of Observations within Ranges – Eighth Grade**

	Low Benchmark	Intermediate Benchmark	High Benchmark	Advanced Benchmark
Range of Scale Scores	**395 - 405**	**470 - 480**	**545 - 555**	**620 – 630**
Mathematics Students	6372	8294	6955	3320
Science Students	5633	8731	8373	3477

Exhibit 12.3 **Range around Each Anchor Point and Number of Observations within Ranges – Fourth Grade**

	Low Benchmark	Intermediate Benchmark	High Benchmark	Advanced Benchmark
Range of Scale Scores	395 - 405	470 - 480	545 - 555	620 – 630
Mathematics Students	2352	4173	5169	2481
Science Students	2408	4559	4892	2085

12.2.3 Anchoring Criteria

Having identified the number of students scoring at each benchmark anchor point, the next step was establishing criteria for determining whether particular items anchor at each of the anchor points. An important feature of the scale anchoring method is that it yields descriptions of the performance demonstrated by students reaching the benchmarks on the TIMSS mathematics and science achievement scales, and that these descriptions reflect demonstrably different accomplishments of students reaching each successively higher benchmark. The process entails the delineation of sets of items that students at each benchmark anchor point are very likely to answer correctly and that discriminate between performance at the various benchmarks. Criteria were applied to identify the items that are answered correctly by most of the students at the anchor point, but by fewer students at the next lower point.

In scale anchoring, the anchor items for each point are intended to be those that differentiate between adjacent anchor points, e.g., between the Advanced and the High international benchmarks. To meet this goal, the criteria for identifying the items must take into consideration performance at more than one anchor point. Therefore, in addition to a criterion for the percentage of students at a particular benchmark correctly answering an item, it was necessary to use a criterion for the percentage of students scoring at the next lower benchmark who correctly answer an item. For multiple choice items, the criterion of 65% was used for the anchor point, since students would be likely (about two-thirds of the time) to answer the item correctly. The criterion of less than 50% was used for the next lower point, because with this response probability, students were more likely to have answered the item incorrectly than correctly. Because there is no possibility of guessing, for constructed response items the criterion of 50% was used for the anchor point and no criterion was used for the lower points.

The criteria used to identify multiple-choice items that "anchored" are outlined below:

For the Low International Benchmark (400), a multiple-choice item anchored if

- At least 65% of students scoring in the range answered the item correctly
- Because the Low International Benchmark was the lowest one described, items were not identified in terms of performance at a lower point

For the Intermediate International Benchmark (475), a multiple-choice item anchored if

- At least 65% of students scoring in the range answered the item correctly and
- Less than 50% of students at the Low International Benchmark answered the item correctly

For the High International Benchmark (550), a multiple-choice item anchored if

- At least 65% of students scoring in the range answered the item correctly and
- Less than 50% of students at the Intermediate International Benchmark answered the item correctly

For the Advanced International Benchmark (625), a multiple-choice item anchored if

- At least 65% of students scoring in the range answered the item correctly and
- Less than 50% of students at the High International Benchmark answered the item correctly

To include all of the items in the anchoring process and provide information about content areas and cognitive processes that might not have had many items anchor exactly, items that met a slightly less stringent set of criteria were also identified. The criteria to identify multiple-choice items that "almost anchored" were the following:

For the Low International Benchmark (400), a multiple-choice item almost anchored if

- At least 60% of students scoring in the range answered the item correctly
- Because Low International Benchmark was the lowest point, items were not identified in terms of performance at a lower point

For the Intermediate International Benchmark (475), a multiple-choice item almost anchored if

- At least 60% of students scoring in the range answered the item correctly and

- Less than 50% of students at the Low International Benchmark answered the item correctly

For the High International Benchmark (550), a multiple-choice item almost anchored if

- At least 60% of students scoring in the range answered the item correctly and
- Less than 50% of students at the Intermediate International Benchmark answered the item correctly

For the Advanced International Benchmark (625), a multiple-choice item almost anchored if

- At least 60% of students scoring in the range answered the item correctly and
- Less than 50% of students at the High International Benchmark answered the item correctly

To be completely inclusive for all items, items that met only the criterion that at least 60% of the students answered correctly (regardless of the performance of students at the next lower point) were also identified. The three categories of items were mutually exclusive, and ensured that all of the items were available to inform the descriptions of student achievement at the anchor levels. A multiple-choice item was considered to be "too difficult" to anchor if less than 60% of students at the Advanced Benchmark answered the item correctly.

Different criteria were used to identify constructed-response items that "anchored." A constructed-response item anchored at one of the international benchmarks if at least 50% of students at that benchmark answer the item correctly. A constructed-response item was considered to be "too difficult" to anchor if less than 50% of students at the Advanced Benchmark answered the item correctly.

12.2.4 Computing the Item Percent Correct At Each Anchor Level

The percentage of students scoring in the range around each anchor point that answered the item correctly was computed. To compute these percentages, students in each country were weighted to contribute proportional to the size of the student population in a country. Most of the TIMSS 2003 items are scored dichotomously. For these items, the percent of students at each anchor point who answered each item correctly was computed. For constructed-response items, percentages were computed for the students receiving full credit, even if the item was scored for partial as well as full credit.

12.2.5 Identifying Anchor Items

For the TIMSS 2003 mathematics and science scales, the criteria described above were applied to identify the items that anchored, almost anchored, and met only the 60 to 65 percent criterion. Exhibit 12.4 and Exhibit 12.5 present the number of these items, at the eighth grade, anchoring at each anchor point on the mathematics and science scales, respectively. Exhibit 12.6 and Exhibit 12.7 present the numbers at the fourth grade. All together, at the eighth grade, four mathematics items met the anchoring criteria at the Low International Benchmark, 40 did so for the Intermediate International Benchmark, 75 for the High International Benchmark, and 63 for the Advanced International Benchmark. Twelve items were too difficult for the Advanced International Benchmark. In science, 10 items met one of the criteria for anchoring at the Low International Benchmark, 23 for the Intermediate International Benchmark, 61 for the High International Benchmark, and 68 for the Advanced International Benchmark. Twenty-seven items were too difficult to anchor at the Advanced International Benchmark at the eighth grade.

At the fourth grade level, 17 mathematics items met the anchoring criteria at the Low International Benchmark, 43 did so for the Intermediate International Benchmark, 56 for the High International Benchmark, and 33 for the Advanced International Benchmark. Ten items were too difficult for the Advanced International Benchmark. In science, 32 items met one of the criteria for anchoring at the Low International Benchmark, 37 for the Intermediate International Benchmark, 28 for the High International Benchmark, and 37 for the Advanced International Benchmark. Sixteen items were too difficult to anchor at the Advanced International Benchmark at the fourth grade.

Including items meeting the less stringent anchoring criteria substantially increased the number of items that could be used to characterize performance at each benchmark, beyond what would have been available if only the items that met the 65 percent criteria were included. Even though these items did not meet the 65 percent anchoring criteria, they were still items that students scoring at the benchmarks had a high degree of probability of answering correctly.

Exhibit 12.4 **Number of Items Anchoring at Each Anchor Level Eighth Grade Mathematics**

	Anchored	Almost Anchored	Met 60-65% Criterion	Total
Low (400)	3*	1	-	4
Intermediate (475)	25	5*	10	40
High (550)	46	10	19*	75
Advanced (625)	41	5	17	63
Too Difficult to Anchor	12	-	-	12
Total	**127**	**21**	**46**	**194**

* These numbers where obtained based on the anchor points where the calculator-sensitive items anchor if considered without calculator (see Appendix A of the International Mathematics Report for more details on calculator use in TIMSS 2003 assessment))

Exhibit 12.5 **Number of Items Anchoring at Each Anchor Level Eighth Grade Science**

	Anchored	Almost Anchored	Met 60-65% Criterion	Total
Low (400)	6	4	-	10
Intermediate (475)	10	4	9	23
High (550)	35	5	21	61
Advanced (625)	40	5	23	68
Too Difficult to Anchor	27	-	-	27
Total	**118**	**18**	**53**	**189**

Exhibit 12.6 **Number of Items Anchoring at Each Anchor Level Fourth Grade Mathematics**

	Anchored	Almost Anchored	Met 60-65% Criterion	Total
Low (400)	15	2	-	17
Intermediate (475)	21	11	11	43
High (550)	36	7	13	56
Advanced (625)	23	1	9	33
Too Difficult to Anchor	10	-	-	10
Total	**105**	**21**	**33**	**159**[2]

2 Following the item review, two items were deleted out of 161 items in the Mathematics Grade 4 test, resulting in 159 items (see chapter 10 for more details on item review process).

Exhibit 12.7 **Number of Items Anchoring at Each Anchor Level Fourth Grade Science**

	Anchored	Almost Anchored	Met 60-65% Criterion	Total
Low (400)	26	6	-	32
Intermediate (475)	20	5	12	37
High (550)	18	2	8	28
Advanced (625)	25	3	9	37
Too Difficult to Anchor	16	-	-	16
Total	**105**	**16**	**29**	**150**[3]

12.2.6 Expert Review of Anchor Items by Content Area

Having identified the items that anchored at each of the international benchmarks, the next step was to have the items reviewed by the TIMSS 2003 Science and Mathematics Item Review Committee (SMIRC) to develop descriptions of student performance. In preparation for the review by the SMIRC, the mathematics and science items, respectively, were organized in binders grouped by benchmark anchor point and within anchor point, the items were sorted by content area and then by the anchoring criteria they met – items that anchored, followed by items that almost anchored, followed by items that met only the 60 to 65% criteria. The following information was included for each item: content area, main topic, cognitive domain, answer key, percent correct at each anchor point, and overall international percent correct. For open-ended items, the scoring guides were included.

The TIMSS & PIRLS International Study Center convened the SMIRC for a four-day meeting. The assignment consisted of three tasks: (1) work through each item in each binder and arrive at a short description of the knowledge, understanding, and/or skills demonstrated by students answering the item correctly; (2) based on the items that anchored, almost anchored, and met only the 60-65% criterion, draft a description of the level of comprehension demonstrated by students at each of the four benchmark anchor points; and (3) select example items to support and illustrate the anchor point descriptions. Following the meeting, these drafts were edited and revised as necessary for use in the TIMSS 2003 International Reports.

Exhibits 12.8 and 12.9 present, for each scale, the number of items per content area that met one of the anchoring criteria discussed above, at each International Benchmark, and the number of items that were too difficult for the Advanced International Benchmark, at the eighth grade level. Exhibits 12.10 and 12.11 present the same information for the fourth grade. The descriptions for each item developed by SMIRC and the summaries are

3 Following the item review, two items were deleted out of 152 items in the Science Grade 4 test, resulting in 150 items (see chapter 10 for more details on item review process).

presented in the TIMSS 2003 International Reports.

Exhibit 12.8 **Number of Items Anchoring at Each Anchor Level, by Content Area Eighth Grade Mathematics**

	Low (400)	Intermediate (475)	High (550)	Advanced (625)	Too Difficult to Anchor	Total
Number	2*	11*	22*	20	2	57
Algebra	0	11	16	16	4	47
Measurement	1	4	14	10	2	31
Geometry	0	8	12	10	1	31
Data	1	6	11	7	3	28
Total	**4**	**40**	**75**	**63**	**12**	**194**

* These numbers where obtained based on the anchor points where the calculator-sensitive items anchor if considered without calculator (see Appendix A of the International Mathematics Report for more details on calculator use in TIMSS 2003 assessment)

Exhibit 12.9 **Number of Items Anchoring at Each Anchor Level, by Content Area Eighth Grade Science**

	Low (400)	Intermediate (475)	High (550)	Advanced (625)	Too Difficult to Anchor	Total
Life Science	4	4	19	19	8	54
Chemistry	1	1	8	16	5	31
Physics	3	7	17	13	6	46
Earth Science	1	7	9	10	4	31
Environmental Science	1	4	8	10	4	27
Total	**10**	**23**	**62**	**68**	**27**	**189**

Exhibit 12.10 **Number of Items Anchoring at Each Anchor Level, by Content Area Fourth Grade Mathematics**

	Low (400)	Intermediate (475)	High (550)	Advanced (625)	Too Difficult to Anchor	Total
Number	7	18	22	12	4	63
Patterns and Relationships	1	6	8	4	4	23
Measurement	2	5	11	13	1	32
Geometry	5	6	10	2	1	24
Data	2	8	5	2	0	17
Total	**17**	**43**	**56**	**33**	**10**	**159**[4]

4 Following the item review, two items were deleted out of 161 items in the Mathematics Grade 4 test, resulting in 159 items (see chapter 10 for more details on item review process).

Exhibit 12.11 **Number of Items Anchoring at Each Anchor Level, by Content Area Fourth Grade Science**

	Low (400)	Intermediate (475)	High (550)	Advanced (625)	Too Difficult to Anchor	Total
Life Science	17	14	11	15	7	64
Physical Science	9	13	12	13	6	53
Earth Science	6	10	5	9	3	33
Total	**32**	**37**	**28**	**37**	**16**	**150**[5]

12.3 Capturing the Uncertainty in the TIMSS Student Achievement Measures

To obtain estimates of students' proficiency in mathematics and science that were both accurate and cost-effective, TIMSS 2003 made extensive use of probability sampling techniques to sample students from national eighth- and fourth-grade student populations, and applied matrix sampling methods to target individual students with a subset of the entire set of assessment materials. Statistics computed from these student samples were used to estimate population parameters. This approach made an efficient use of resources, in particular keeping student response burden to a minimum, but at a cost of some variance or uncertainty in the statistics. To quantify this uncertainty, each statistic in the TIMSS 2003 international reports (Mullis et al., 2004; Martin et al., 2004) is accompanied by an estimate of its standard error. These standard errors incorporate components reflecting the uncertainty due to generalizing from student samples to the entire eighth- or fourth-grade student population (sampling variance), and to inferring students' performance on the entire assessment from their performance on the subset of items that they took (imputation variance).

12.3.1 Estimating Sampling Variance

The TIMSS 2003 sampling design applied a stratified multistage cluster-sampling technique to the problem of selecting efficient and accurate samples of students while working with schools and classes. This design capitalized on the structure of the student population (i.e., students grouped in classes within schools) to derive student samples that permitted efficient and economical data collection. Unfortunately, however, such a complex sampling design complicates the task of computing standard errors to quantify sampling variability.

When, as in TIMSS, the sampling design involves multistage cluster sampling, there are several options for estimating sampling errors that avoid the assumption of simple random sampling (Wolter, 1985). The jackknife

5 Following the item review, two items were deleted out of 152 items in the Science Grade 4 test, resulting in 150 items (see chapter 10 for more details on item review process).

repeated replication technique (JRR) was chosen by TIMSS because it is computationally straightforward and provides approximately unbiased estimates of the sampling errors of means, totals, and percentages.

The variation on the JRR technique used in TIMSS 2003 is described in Johnson and Rust (1992). It assumes that the primary sampling units (PSUs) can be paired in a manner consistent with the sample design, with each pair regarded as members of a pseudo-stratum for variance estimation purposes. When used in this way, the JRR technique appropriately accounts for the combined effect of the between- and within-PSU contributions to the sampling variance. The general use of JRR entails systematically assigning pairs of schools to sampling zones, and randomly selecting one of these schools to have its contribution doubled and the other to have its contribution zeroed, so as to construct a number of "pseudo-replicates" of the original sample. The statistic of interest is computed once for all of the original sample, and once again for each pseudo-replicate sample. The variation between the estimates for each of the replicate samples and the original sample estimate is the jackknife estimate of the sampling error of the statistic.

12.3.1.1 Constructing Sampling Zones for Sampling Variance Estimation

To apply the JRR technique used in TIMSS 2003, the sampled schools had to be paired and assigned to a series of groups known as sampling zones. This was done at Statistics Canada by working through the list of sampled schools in the order in which they were selected and assigning the first and second schools to the first sampling zone, the third and fourth schools to the second zone, and so on. In total 75 zones were used, allowing for 150 schools per country. When more than 75 zones were constructed, they were collapsed to keep the total number to 75.

Sampling zones were constructed within design domains, or explicit strata. Where there was an odd number of schools in an explicit stratum, either by design or because of school nonresponse, the students in the remaining school were randomly divided to make up two "quasi" schools for the purposes of calculating the jackknife standard error. Each zone then consisted of a pair of schools or "quasi" schools. Exhibit 12.12 shows the range of sampling zones used in each country.

Exhibit 12.12 **Number of Sampling Zones Used in Each Country**

Country	TIMSS 2003 Sampling Zones	TIMSS 1999 Sampling Zones	TIMSS 1995 Sampling Zones
Armenia	75	-	-
Australia	75	-	74
Bahrain	75	-	-
Belgium (Flemish)	75	74	71
Botswana	73	-	-
Bulgaria	75	75	58
Chile	75	75	-
Chinese Taipei	75	75	-
Cyprus	75	61	55
Egypt	75	-	-
England	44	64	64
Estonia	75	-	-
Ghana	75	-	-
Hong Kong, SAR	63	69	43
Hungary	75	74	75
Indonesia	75	75	-
Iran, Islamic Rep. of	75	75	75
Israel	74	70	-
Italy	75	75	-
Japan	74	71	75
Jordan	70	74	-
Korea, Rep. of	75	75	75
Latvia	70	73	64
Lebanon	75	-	-
Lithuania	72	75	73
Macedonia, Rep. of	75	75	-
Malaysia	75	75	-
Moldova, Rep. of	75	75	-
Morocco	67	75	-
Netherlands	65	63	48
New Zealand	75	75	75
Norway	69	-	74
Palestinian Nat'l Auth.	73	-	-
Philippines	69	75	-
Romania	74	74	72
Russian Federation	69	56	41
Saudi Arabia	75	-	-

Exhibit 12.12 **Number of Sampling Zones Used in Each Country** *(...Continued)*

Country	TIMSS 2003 Sampling Zones	TIMSS 1999 Sampling Zones	TIMSS 1995 Sampling Zones
Scotland	65	-	64
Serbia	75	-	-
Singapore	75	73	69
Slovak Republic	75	73	73
Slovenia	75	-	61
South Africa	75	75	-
Sweden	75	-	60
Tunisia	75	75	-
United States	75	53	55

12.3.1.2 Computing Sampling Variance Using the JRR Method

The JRR algorithm used in TIMSS 2003 assumes that there are H sampling zones within each country, each containing two sampled schools selected independently. To compute a statistic t from the sample for a country, the formula for the JRR variance estimate of the statistic t is then given by the following equation:

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

where H is the number of pairs in the sample for the country. The term $t(S)$ corresponds to the statistic for the whole sample (computed with any specific weights that may have been used to compensate for the unequal probability of selection of the different elements in the sample or any other post-stratification weight). The element $t(J_h)$ denotes the same statistic using the h^{th} jackknife replicate. This is computed using all cases except those in the h^{th} zone of the sample; for those in the h^{th} zone, all cases associated with one of the randomly selected units of the pair are removed, and the elements associated with the other unit in the zone are included twice. In practice, this is 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.

The computation of the JRR variance estimate for any statistic in TIMSS 2003 required the computation of the statistic up to 76 times for any given country: once to obtain the statistic for the full sample, and up to 75 times to obtain the statistics for each of the jackknife replicates (J_h). The number of times a statistic needed to be computed for a given country depended on the number of implicit strata or sampling zones defined for that country.

Doubling and zeroing the weights of the selected units within the sampling zones was accomplished by creating replicate weights that were then used in the calculations. In this approach, a set of temporary replicate weights are created for each pseudo-replicate sample. Each replicate weight is equal to k times the overall sampling weight, where k can take values of 0, 1, or 2 depending on whether the case is to be removed from the computation, left as it is, or have its weight doubled. The value of k for an individual student record for a given replicate depends on the assignment of the record to the specific PSU and zone.

Within each zone the members of the pair of schools are assigned an indicator (u_i), coded randomly to 1 or 0 so that one of them has a value of 1 on the variable u_i, and the other a value of 0. This indicator determines whether the weights for the elements in the school in this zone are to be doubled or zeroed. The replicate weight $W_h^{g,i,j}$ for the elements in a school assigned to zone h is computed as the product of k_h times their overall sampling weight, where k_h can take values of 0, 1, or 2 depending on whether the school is to be omitted, be included with its usual weight, or have its weight doubled for the computation of the statistic of interest. In TIMSS 2003, the replicate weights were not permanent variables, but were created temporarily by the sampling variance estimation program as a useful computing device.

To create replicate weights, each sampled student was first assigned a vector of 75 weights, $W_h^{g,i,j}$, where h takes values from 1 to 75. The value of $W_0^{g,i,j}$ is the overall sampling weight, which is simply the product of the final school weight, classroom weight, and student weight, as described in Chapter 9.

The replicate weights for a single case were then computed as

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

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

In the TIMSS 2003 analysis, 75 replicate weights were computed for each country regardless of the number of actual zones within the country. If a country had fewer than 75 zones, then the replicate weights W_h, where h was greater than the number of zones within the country, were each the same as the overall sampling weight. Although this involved some redundant computation, having 75 replicate weights for each country had no effect on the size of the error variance computed using the jackknife formula, but it facilitated the computation of standard errors for a number of countries at a time.

Standard errors presented in the international reports were computed using SAS programs developed at the TIMSS & PIRLS International Study Center. As a quality control check, results were verified using the WesVarPC software (Westat, 1997).

12.3.2 Estimating Imputation Variance

The TIMSS 2003 item pool was far too extensive to be administered in its entirety to any one student, and so a matrix-sampling test design was developed whereby each student was given a single test booklet containing only a part of the entire assessment.[6] The results for all of the booklets were then aggregated using item response theory to provide results for the entire assessment. Since each student responded to just a subset of the assessment items, multiple imputation (the generation of "plausible values") was used to derive reliable estimates of student performance on the assessment as a whole. Since every student proficiency estimate incorporates some uncertainty, TIMSS followed the customary procedure of generating five estimates for each student and using the variability among them as a measure of this imputation uncertainty, or error. In the TIMSS 2003 international report the imputation error for each variable has been combined with the sampling error for that variable to provide a standard error incorporating both.

The general procedure for estimating the imputation variance using plausible values is the following (Mislevy, R.J., Beaton, A.E., Kaplan, B., and Sheenan, K.M., 1992). First compute the statistic t, for each set of M plausible values. The statistics t_m, where m = 1, 2, …, 5, can be anything estimable from the data, such as a mean, the difference between means, percentiles, and so forth.

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

$$Var_{imp} = (1 + 1/M)\, Var(t_1, \ldots, t_M)$$

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

12.3.3 Combining Sampling and Imputation Variance

The standard errors of the mathematics and science proficiency statistics reported by TIMSS include both sampling and imputation variance components. The standard errors were computed using the following formula:[7]

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

6 Details of the TIMSS test design may be found in Chapter 2.

7 Under ideal circumstances and with unlimited computing resources, the imputation variance for the plausible values and the JRR sampling variance for each of the plausible values would be computed. This would be equivalent to computing the same statistic up to 380 times (once overall for each of the five plausible values using the overall sampling weights, and then 75 times more for each plausible value using the complete set of replicate weights). An acceptable shortcut, however, is to compute the JRR variance component using one plausible value, and then the imputation variance using the five plausible values. Using this approach, a statistic needs to be computed only 80 times.

where $Var_{jrr}(t_1)$ is the sampling variance for the first plausible value and Var_{imp} is the imputation variance. The User Guide for the TIMSS 2003 International Database contains programs in SAS and SPSS that compute each of these variance components for the TIMSS 2003 data.

Exhibits 12.13 through 12.16 show basic summary statistics for mathematics and science achievement in the TIMSS 2003 assessment for the eighth and fourth grades. Each exhibit presents the student sample size, the mean and standard deviation, averaged across the five plausible values, the jackknife standard error for the mean, and the overall standard errors for the mean including imputation error. Appendix E contains tables showing the same summary statistics for the mathematics and science content areas for the eighth and fourth grades.

12.4 Calculating National and International Statistics for Student Achievement

As described in earlier chapters, TIMSS 2003 made extensive use of imputed proficiency scores to report student achievement, both in the major content domains (number, algebra, measurement, geometry, and data for mathematics and life science, chemistry, physics, earth science, and environmental science for science) and mathematics and science as overall subjects. This section describes the procedures followed in computing the principal statistics used to summarize achievement in the International Reports (Mullis, et al., 2004; Martin et al., 2004), including means based on plausible values, gender differences, performance in content domains, and performance on example items.

For each of the TIMSS 2003 mathematics and science scales, the item response theory (IRT) scaling procedure described in Chapter 11 yields five imputed scores or plausible values for each student. The difference between the five values reflects the degree of uncertainty in the imputation process. When the process yields consistent results, the differences between the five values are very small. To obtain the best estimate for each of the TIMSS statistics, each one was computed five times, using each of the five plausible values in turn, and the results averaged to derive the reported value. The standard errors that accompany each reported statistic include two components as described in the previous section: one quantifying sampling variation and the other quantifying imputation variation.

Exhibit 12.13 **Summary Statistics and Standard Errors for Proficiency in Mathematics - Eighth Grade**

Country	Sample Size	Mean Proficiency	Standard Deviation	Jackknife Sampling Error	Overall Standard Error
Armenia	5726	478.127	83.522	2.952	2.997
Australia	4791	504.703	81.538	4.613	4.638
Bahrain	4199	401.196	76.317	1.571	1.727
Belgium (Flemish)	4970	536.710	73.494	2.696	2.772
Botswana	5150	366.345	71.554	2.189	2.581
Bulgaria	4117	476.169	84.077	4.222	4.315
Chile	6377	386.880	83.233	3.060	3.269
Chinese Taipei	5379	585.252	99.969	4.507	4.607
Cyprus	4002	459.366	81.377	1.474	1.653
Egypt	7095	406.168	92.754	3.423	3.505
England	2830	498.464	77.231	4.653	4.674
Estonia	4040	530.915	69.334	2.931	2.997
Ghana	5100	275.704	90.996	4.339	4.657
Hong Kong, SAR	4972	586.051	71.924	3.245	3.324
Hungary	3302	529.275	79.506	3.212	3.221
Indonesia	5762	410.702	88.789	4.796	4.844
Iran, Islamic Rep. of	4942	411.447	74.303	2.316	2.351
Israel	4318	495.648	84.682	3.360	3.422
Italy	4278	483.599	76.675	3.145	3.192
Japan	4856	569.921	79.874	1.985	2.074
Jordan	4489	424.352	89.007	4.068	4.086
Korea, Rep. of	5309	589.092	83.855	1.853	2.191
Latvia	3630	508.327	73.094	3.131	3.174
Lebanon	3814	433.045	66.747	3.040	3.091
Lithuania	4964	501.615	78.291	2.442	2.458
Macedonia, Rep. of	3893	434.983	88.380	3.500	3.542
Malaysia	5314	508.336	74.263	4.035	4.079
Moldova, Rep. of	4033	459.895	80.563	4.006	4.050
Morocco	2943	386.539	68.126	2.134	2.483
Netherlands	3065	536.273	69.391	3.788	3.820
New Zealand	3801	494.040	78.318	5.264	5.275
Norway	4133	461.470	70.859	2.427	2.499
Palestinian Nat’l Auth.	5357	390.486	91.839	3.037	3.104
Philippines	6917	377.690	87.339	5.164	5.208
Romania	4104	475.282	90.230	4.786	4.822
Russian Federation	4667	508.041	76.619	3.532	3.709
Saudi Arabia	4295	331.682	78.324	4.466	4.574
Scotland	3516	497.654	74.820	3.585	3.711

Exhibit 12.13 **Summary Statistics and Standard Errors for Proficiency in Mathematics - Eighth Grade** *(...Continued)*

Country	Sample Size	Mean Proficiency	Standard Deviation	Jackknife Sampling Error	Overall Standard Error
Serbia	4296	476.637	88.850	2.477	2.595
Singapore	6018	605.450	80.090	3.508	3.583
Slovak Republic	4215	507.740	82.382	3.250	3.308
Slovenia	3578	492.956	71.101	2.089	2.193
South Africa	8952	263.614	107.151	5.330	5.490
Sweden	4256	499.058	71.182	2.550	2.622
Tunisia	4931	410.329	60.340	2.121	2.186
United States	8912	504.366	79.993	3.270	3.309

Exhibit 12.14 **Summary Statistics and Standard Errors for Proficiency in Mathematics - Fourth Grade**

Country	Sample Size	Mean Proficiency	Standard Deviation	Jackknife Sampling Error	Overall Standard Error
Armenia	5674	455.925	86.681	3.473	3.489
Australia	4321	498.663	80.862	3.821	3.882
Belgium (Flemish)	4712	550.601	58.948	1.773	1.783
Chinese Taipei	4661	563.949	63.029	1.696	1.752
Cyprus	4328	509.810	85.391	2.399	2.424
England	3585	531.182	87.407	3.701	3.736
Hong Kong, SAR	4608	574.782	63.389	3.080	3.161
Hungary	3319	528.502	77.251	3.045	3.130
Iran, Islamic Rep. of	4352	389.052	85.697	4.012	4.153
Italy	4282	502.762	82.050	3.662	3.679
Japan	4535	564.556	73.749	1.515	1.598
Latvia	3687	535.855	72.517	2.789	2.835
Lithuania	4422	534.017	73.806	2.797	2.804
Moldova, Rep. of	3981	504.149	87.334	4.818	4.879
Morocco	4264	346.807	90.250	4.940	5.081
Netherlands	2937	540.373	54.625	2.013	2.109
New Zealand	4308	493.464	84.230	2.139	2.151
Norway	4342	451.342	80.240	2.260	2.298
Philippines	4572	358.195	109.709	7.861	7.911
Russian Federation	3963	531.682	78.249	4.734	4.746
Scotland	3936	490.321	77.541	3.166	3.252
Singapore	6668	594.427	84.222	5.558	5.597
Slovenia	3126	478.795	77.946	2.575	2.619
Tunisia	4334	339.300	99.591	4.567	4.730
United States	9829	518.284	76.272	2.429	2.436

Exhibit 12.15 **Summary Statistics and Standard Errors for Science Proficiency - Eighth Grade**

Country	Sample Size	Mean Proficiency	Standard Deviation	Jackknife Sampling Error	Overall Standard Error
Armenia	5726	461.267	81.041	3.413	3.465
Australia	4791	527.014	75.307	3.763	3.800
Bahrain	4199	438.255	74.470	1.625	1.793
Belgium (Flemish)	4970	515.506	66.954	2.457	2.487
Botswana	5150	364.569	86.472	2.771	2.840
Bulgaria	4117	478.843	92.987	5.072	5.151
Chile	6377	412.851	84.096	2.827	2.890
Chinese Taipei	5379	571.092	79.064	3.381	3.457
Cyprus	4002	441.474	79.496	1.589	2.049
Egypt	7095	421.117	103.720	3.825	3.898
England	2830	543.896	76.832	4.070	4.140
Estonia	4040	552.258	65.049	2.382	2.456
Ghana	5100	255.324	120.145	5.726	5.882
Hong Kong, SAR	4972	556.089	65.545	2.965	3.039
Hungary	3302	542.761	75.903	2.800	2.837
Indonesia	5762	420.221	78.769	3.981	4.055
Iran, Islamic Rep. of	4942	453.428	72.593	2.176	2.329
Israel	4318	488.200	84.965	3.028	3.082
Italy	4278	490.891	78.125	2.996	3.062
Japan	4856	552.178	71.011	1.691	1.739
Jordan	4489	474.845	89.396	3.755	3.848
Korea, Rep. of	5309	558.399	69.575	1.581	1.641
Latvia	3630	512.363	67.343	2.532	2.551
Lebanon	3814	393.399	92.556	4.271	4.315
Lithuania	4964	519.380	69.632	2.126	2.143
Macedonia, Rep. of	3893	449.373	91.641	3.575	3.596
Malaysia	5314	510.452	65.855	3.643	3.651
Moldova, Rep. of	4033	472.423	73.553	3.258	3.365
Morocco	2943	396.474	69.138	2.141	2.501
Netherlands	3065	535.765	61.278	3.046	3.077
New Zealand	3801	519.730	73.716	5.010	5.044
Norway	4133	493.863	69.755	2.107	2.170
Palestinian Nat'l Auth.	5357	435.387	92.463	3.215	3.240

Exhibit 12.15 **Summary Statistics and Standard Errors for Science Proficiency - Eighth Grade** *(...Continued)*

Country	Sample Size	Mean Proficiency	Standard Deviation	Jackknife Sampling Error	Overall Standard Error
Philippines	6917	377.373	102.264	5.659	5.803
Romania	4104	469.604	91.090	4.865	4.936
Russian Federation	4667	513.621	75.184	3.561	3.679
Saudi Arabia	4295	397.741	72.491	3.618	3.985
Scotland	3516	511.546	75.689	3.319	3.351
Serbia	4296	467.686	83.688	2.412	2.467
Singapore	6018	577.849	91.817	4.249	4.262
Slovak Republic	4215	516.785	75.587	3.159	3.215
Slovenia	3578	520.498	66.696	1.725	1.786
South Africa	8952	243.664	131.640	6.357	6.683
Sweden	4256	524.258	73.901	2.587	2.688
Tunisia	4931	403.547	60.483	1.914	2.082
United States	8912	527.298	80.681	3.095	3.143

Exhibit 12.16 **Summary Statistics and Standard Errors for Science Proficiency - Fourth Grade**

Country	Sample Size	Mean Proficiency	Standard Deviation	Jackknife Sampling Error	Overall Standard Error
Armenia	5674	436.528	95.954	4.219	4.299
Australia	4321	520.691	82.093	4.137	4.206
Belgium (Flemish)	4712	518.342	54.858	1.542	1.769
Chinese Taipei	4661	551.355	68.622	1.589	1.727
Cyprus	4328	480.485	74.171	2.214	2.379
England	3585	540.240	83.167	3.383	3.608
Hong Kong, SAR	4608	542.483	59.804	2.907	3.059
Hungary	3319	529.727	79.351	2.887	2.979
Iran, Islamic Rep. of	4352	413.923	96.600	4.070	4.104
Italy	4282	515.640	84.861	3.749	3.766
Japan	4535	543.469	73.117	1.343	1.509
Latvia	3687	531.521	68.794	2.464	2.489
Lithuania	4422	512.106	66.362	2.171	2.551
Moldova, Rep. of	3981	496.420	84.966	4.576	4.599
Morocco	4264	304.392	124.834	6.582	6.705
Netherlands	2937	525.125	53.351	1.816	2.001
New Zealand	4308	519.671	85.050	2.375	2.460
Norway	4342	466.346	83.994	2.154	2.619
Philippines	4572	331.620	145.326	9.293	9.433
Russian Federation	3963	526.187	82.019	5.115	5.167
Scotland	3936	501.975	77.719	2.808	2.887
Singapore	6668	565.148	86.786	5.517	5.548
Slovenia	3126	490.365	77.195	2.462	2.530
Tunisia	4334	313.989	125.686	5.583	5.655
United States	9829	535.631	81.247	2.408	2.526

National averages were computed as the average of the weighted means for each of the five plausible values. The weighted mean for each plausible value was computed as follows:

$$\overline{X}_{pvl} = \frac{\sum_{j=1}^{N} W^{i,j} . pv_{lj}}{\sum_{j=1}^{N} W^{i,j}}$$

where

$\overline{X}_{pvl}$ is the country mean for plausible value l

pv_{lj} is the l-th plausible value for the j-th student

$W^{i,j}$ is the weight associated with the j-th student in class i, described in Chapter 9

N is the number of students in the country's sample.

These five weighted means were then averaged to obtain the national average for each country. To provide a reference point for comparison purposes, TIMSS presented the international average of many of the national statistics (means and percentages). International averages were calculated by first computing the national average for each plausible value for each country and then averaging across countries. These five estimates of the international average were then themselves averaged to derive the international average presented in the TIMSS reports, as shown below:

$$\overline{X}_{\bullet pvl} = \frac{\sum_{k=1}^{K} \overline{X}_{pvl,k}}{K}$$

where

$\overline{X}_{\bullet pvl}$ is the international mean for plausible value l

$\overline{X}_{pvl,k}$ is the k-th country mean for plausible value l

and K is the number of countries.

12.4.1 Comparing Achievement Differences Across Countries

A basic aim of the TIMSS 2003 International Reports is to provide fair and accurate comparisons of student achievement across the participating countries. Most of the exhibits in the TIMSS reports summarize student achievement by means of a statistic such as a mean or percentage, and each statistic is accompanied by its standard error, which is a measure of the uncertainty due to student sampling and the imputation process. In comparisons of performance across countries, standard errors can be used to assess the statistical significance of the difference between the summary statistics.

The exhibits presented in the TIMSS 2003 international reports allow comparisons of average performance of a country with that of other participating countries. If repeated samples were taken from two populations with the same mean and variance and in each one the hypothesis that the means from the two samples are significantly different at the $\alpha = .05$ level (i.e. with 95% confidence) was tested, then in about five percent of the comparisons it would be expected to find significant differences between the sample means even though no difference exists in the population. In such a test of the difference between two means, the probability of finding significant differences in the samples when none exist in the populations (the so-called type I error) is given by $\alpha = .05$. Conversely, the probability of not making such an error is $1 - \alpha$, which in the case of a single test is .95.

Mean proficiencies are considered significantly different if the absolute difference between them, divided by the standard error of the difference, is greater than the critical value. For differences between countries, which can be considered as independent samples, the standard error of the difference between means is computed as the square root of the sum of the squared standard errors of each mean:

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

where se_1 and se_2 are the standard errors of the means. Exhibits 12.17 and 12.18 show the means and standard errors used in the calculation of statistical significance for mathematics and science achievement in the eighth and fourth grades.

In contrast to the practice in previous TIMSS reports, the significance tests presented in the TIMSS 2003 International Reports have NOT been adjusted for multiple comparisons among countries. Although adjustments such as the Bonferroni procedure guard against misinterpreting the outcome of multiple simultaneous significance tests, and have been used in previous TIMSS studies, the results vary depending on the number of countries included in the adjustment, leading to apparently conflicting results from comparisons using different numbers of countries.

12.4.2 Comparing National Achievement Against the International Mean

Many of the data exhibits in the TIMSS 2003 international reports show countries' mean achievement compared with the international mean, together with a test of the statistical significance between the two. These significance tests were based on the standard errors of the national and international means.

Exhibit 12.17 **Means and Standard Errors for Country Comparisons of Mathematics and Science Achievement in the Eighth Grade**

	Mathematics		Science	
Country	**Mean**	**S.E.**	**Mean**	**S.E.**
Armenia	478.127	2.997	461.267	3.465
Australia	504.703	4.638	527.014	3.800
Bahrain	401.196	1.727	438.255	1.793
Basque Country, Spain	487.061	2.732	488.754	2.678
Belgium (Flemish)	536.710	2.772	515.506	2.487
Botswana	366.345	2.581	364.569	2.840
Bulgaria	476.169	4.315	478.843	5.151
Chile	386.880	3.269	412.851	2.890
Chinese Taipei	585.252	4.607	571.092	3.457
Cyprus	459.366	1.653	441.474	2.049
Egypt	406.168	3.505	421.117	3.898
England	498.464	4.674	543.896	4.140
Estonia	530.915	2.997	552.258	2.456
Ghana	275.704	4.657	255.324	5.882
Hong Kong, SAR	586.051	3.324	556.089	3.039
Hungary	529.275	3.221	542.761	2.837
Indiana State, US	508.257	5.215	530.609	4.769
Indonesia	410.702	4.844	420.221	4.055
Iran, Islamic Rep. of	411.447	2.351	453.428	2.329
Israel	495.648	3.422	488.200	3.082
Italy	483.599	3.192	490.891	3.062
Japan	569.921	2.074	552.178	1.739
Jordan	424.352	4.086	474.845	3.848
Korea, Rep. of	589.092	2.191	558.399	1.641
Latvia	508.327	3.174	512.363	2.551
Lebanon	433.045	3.091	393.399	4.315
Lithuania	501.615	2.458	519.380	2.143
Macedonia, Rep. of	434.983	3.542	449.373	3.596
Malaysia	508.336	4.079	510.452	3.651
Moldova, Rep. of	459.895	4.050	472.423	3.365
Morocco	386.539	2.483	396.474	2.501
Netherlands	536.273	3.820	535.765	3.077
New Zealand	494.040	5.275	519.730	5.044
Norway	461.470	2.499	493.863	2.170
Ontario Province, Can.	520.932	3.105	532.920	2.656
Palestinian Nat'l Auth.	390.486	3.104	435.387	3.240

Exhibit 12.17 **Means and Standard Errors for Country Comparisons of Mathematics and Science Achievement in the Eighth Grade** *(...Continued)*

	Mathematics		Science	
Country	**Mean**	**S.E.**	**Mean**	**S.E.**
Philippines	377.690	5.208	377.373	5.803
Quebec Province, Can.	543.075	3.031	531.013	3.044
Romania	475.282	4.822	469.604	4.936
Russian Federation	508.041	3.709	513.621	3.679
Saudi Arabia	331.682	4.574	397.741	3.985
Scotland	497.654	3.711	511.546	3.351
Serbia	476.637	2.595	467.686	2.467
Singapore	605.450	3.583	577.849	4.262
Slovak Republic	507.740	3.308	516.785	3.215
Slovenia	492.956	2.193	520.498	1.786
South Africa	263.614	5.490	243.664	6.683
Sweden	499.058	2.622	524.258	2.688
Tunisia	410.329	2.186	403.547	2.082
United States	504.366	3.309	527.298	3.143

Exhibit 12.18 **Means and Standard Errors for Country Comparisons of Mathematics and Science Achievement in the Fourth Grade**

	Mathematics		Science	
Country	**Mean**	**S.E.**	**Mean**	**S.E.**
Armenia	455.925	3.489	436.528	4.299
Australia	498.663	3.882	520.691	4.206
Belgium (Flemish)	550.601	1.783	518.342	1.769
Chinese Taipei	563.949	1.752	551.355	1.727
Cyprus	509.810	2.424	480.485	2.379
England	531.182	3.736	540.240	3.608
Hong Kong, SAR	574.782	3.161	542.483	3.059
Hungary	528.502	3.130	529.727	2.979
Indiana State, US	532.874	2.806	553.287	3.710
Iran, Islamic Rep. of	389.052	4.153	413.923	4.104
Italy	502.762	3.679	515.640	3.766
Japan	564.556	1.598	543.469	1.509
Latvia	535.855	2.835	531.521	2.489
Lithuania	534.017	2.804	512.106	2.551
Moldova, Rep. of	504.149	4.879	496.420	4.599
Morocco	346.807	5.081	304.392	6.705
Netherlands	540.373	2.109	525.125	2.001
New Zealand	493.464	2.151	519.671	2.460
Norway	451.342	2.298	466.346	2.619
Ontario Province, Can.	511.184	3.830	540.205	3.746
Philippines	358.195	7.911	331.620	9.433
Quebec Province, Can.	505.848	2.409	500.392	2.484
Russian Federation	531.682	4.746	526.187	5.167
Scotland	490.321	3.252	501.975	2.887
Singapore	594.427	5.597	565.148	5.548
Slovenia	478.795	2.619	490.365	2.530
Tunisia	339.300	4.730	313.989	5.655
United States	518.284	2.436	535.631	2.526

When comparing each 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, TIMSS adjusted the standard error of the difference. The sampling component of the standard error of the difference for country *j* is

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

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,

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 for country *j* was computed by taking the square root of the imputation variance calculated as follows

$$se_{i\_dif\_j} = \sqrt{\frac{6}{5} Var(d_1,...,d_l,...,d_5)},$$

where d_l is the difference between the international mean and the country mean for plausible value *l*.

Finally, the standard error of the difference was calculated as

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

12.4.3 Reporting Gender Differences Within Countries

TIMSS reported gender differences in overall student achievement in mathematics and science overall, as well as in mathematics and science content areas. Gender differences were presented in an exhibit showing mean achievement for males and females and the differences between them, with an accompanying graph indicating whether the difference was statistically significant.

Because in most countries males and females attend the same schools, the samples of males and females cannot be treated as independent samples for the purpose of statistical tests. Accordingly, TIMSS used a jackknife procedure applicable to correlated samples for estimating the standard errors of the male-female differences. This involved computing the average difference between boys and girls in each country once for every one of the 75 replicate samples, and five more times, once for each plausible value, as described above.

12.4.4 Examining Profiles of Relative Performance by Content Areas

In addition to performance on mathematics and science overall, it was of interest to see how countries performed in the content areas or domains within each subject relative to their performance on the subject overall. There were five content areas in mathematics and five content areas for science that were used in this analysis.[8] The relative performance of the countries in the content areas was examined separately for each subject. TIMSS 2003 computed the average across content area scores for each country, and then displayed country performance in each content area as the difference between the content area average and the overall average. Confidence intervals were estimated for each difference.

In order to do this, TIMSS computed the vector of average proficiencies for each of the content areas on the test, and joined each of these column vectors to form a matrix R_{ks}, where a row contains the average proficiency score for country k on scale s for a specific subject. This R_{ks} matrix also had a "zeroth" row and column. The elements in r_{k0} contained the average of the elements on the k^{th} row of the R_{ks} matrix. These were the country averages across the content areas. The elements in r_{0s} contained the average of the elements of the *sth* column of the R_{ks} matrix. These are the content area averages across all countries. The element r_{00} contains the overall average for the elements in vector r_{0s} or r_{k0}. Based on this information the matrix I_{ks} was constructed in which the elements are computed as

$$i_{ks} = r_{ks} + r_{00} - r_{0s} - r_{k0}$$

Each of these elements can be considered as the interaction between the performance of country k on content area s. A value of zero for an element i_{ks} indicates a level of performance for country k on content area s that would be expected given its performance on other content areas and its performance relative to other countries on that content area. A negative value for an element i_{ks} indicates a performance for country k on content area s lower than would be expected on the basis of the country's overall performance. A positive value for an element i_{ks} indicates a performance for country k on content area s better than expected. This procedure was applied to each of the five plausible values and the results averaged.

To construct confidence intervals it was necessary first to estimate the standard error for each content area in each country. These were then combined with an adjustment for multiple comparisons, based on the number of content areas.[9] The imputation portion of the error was obtained from combining the results from the five calculations, one with each separate plausible value.

8 Science at fourth grade had just three content areas.

9 Note that the adjustment was for multiple comparisons between content areas, and not across countries.

To compute the JRR portion of the standard error, the vector of average proficiency was computed for each of the country replicates for each of the content areas on the test. For each country and each content area 75 replicates were created.[10] Each replicate was randomly reassigned to one of 75 sampling zones or replicates. These column vectors were then joined to form a new set of matrices each called R_{ks}^{h} where a row contains the average proficiency for country *k* on content area *s* for a specific subject, for the h^{th} international set of replicates. Each of these R_{ks}^{h} matrices had also a "zeroth" row and column. The elements in r_{k0}^{h} contained the average of the elements on the *kth* row of the R_{ks}^{h} matrix. These are the country averages across the content areas. The elements in r_{0s}^{h} contained the average of the elements of the *sth* column of the R_{ks}^{h} matrix. These were the content area averages across all countries. The element r_{0}^{h} contains the overall average for the elements in vector r_{0s}^{h} or r_{k0}^{h}. Based on this information the set of matrices R_{ks}^{h} were constructed, in which the elements were computed as

$$i_{ks}^{h} = r_{ks}^{h} + r_{00}^{h} - r_{0s}^{h} - r_{k0}^{h}$$

The JRR standard error is then given by the formula

$$jse_{r_{ks}} = \sqrt{\sum_{h} \left(i_{ks} - i_{ks}^{h}\right)^{2}}$$

The overall standard error was computed by combining the JRR and imputation variances. A relative performance was considered significantly different from the expected if the 95% confidence interval built around it did not include zero. The confidence interval for each of the i_{ks} elements was computed by adding and subtracting to the i_{ks} element its corresponding standard error multiplied by the critical value for the number of comparisons.

The critical values were determined by adjusting the critical value for a two-tailed test, at the alpha 0.05 level of significance for multiple comparisons. The critical value for mathematics and science with five content scales was 2.5758. For the three content scales in fourth grade science, the critical value was 2.3939.

12.4.5 Reporting Student Performance on Individual Items

To portray student achievement as fully as possible, the TIMSS 2003 international reports present many examples of the items used in the TIMSS 2003 tests, together with the percentages of students in each country responding correctly to or earning full credit on the items. The base of these percentages was the total number of students that were administered the item. For multiple-choice items, the weighted percentage of students that answered the item correctly was reported. For constructed-response items with more than one

10 In countries where the were less than 75 jackknife zones, 75 replicates were also created by assigning the overall mean to the as many replicates as were necessary to have 75.

score level, it was the weighted percentage of students that achieved full credit on the item. Omitted and not-reached items were treated as incorrect.

When the percent correct for example items was computed, student responses were classified in the following way. For multiple-choice items, the responses to item *j* were classified as correct (C_j) when the correct option for an item was selected, incorrect (W_j) when the incorrect option or no option at all was selected, invalid (I_j) when two or more choices were made on the same question, not reached (R_j) when it was assumed that the student stopped working on the test before reaching the question, and not administered (A_j) when the question was not included in the student's booklet or had been mistranslated or misprinted. For constructed-response items, student responses to item *j* were classified as correct (C_j) when the maximum number of points was obtained on the question, incorrect (W_j) when the wrong answer or an answer not worth all the points in the question was given, invalid (N_j) when the student's response was not legible or interpretable, or simply left blank, not reached (R_j) when it was determined that the student stopped working on the test before reaching the question, and not administered (A_j) when the question was not included in the student's booklet or had been mistranslated or misprinted. The percent correct for an item (P_j) was computed as

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

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

As described in Chapters 10 and 11, student responses to items in block positions 3 and 6 of the student booklets were found to have different properties to student responses than the same items located in other positions in the booklets. Although these student responses were included in the IRT scaling, albeit with different item parameters, they were not included in the calculation of percent correct on individual example items.

12.5 Examining the TIMSS 2003 Test in the Light of National Curricula

TIMSS 2003 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 2003 Science and Mathematics Item Review 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 of the participating countries. Although every effort was made in TIMSS 2003 to ensure the widest possible subject matter coverage, no test can measure all that is taught or learned in every participating country. Given that no test can cover the curriculum in every country completely, the question arises as to how well the items on the tests match the curricula of each of the participating countries. To address this issue, TIMSS 2003 asked each country to indicate which items on the tests, if any, were inappropriate to its curriculum. For each country, in turn, TIMSS 2003 took the list of remaining items, and computed the average percentage correct on these items for that country and all other countries. This allowed each country to select only those items on the tests that they would like included, and to compare the performance of their students on those items with the performance of the students in each of the other participating countries on that set of items. 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).

The TCMA results show that the TIMSS 2003 tests provide a reasonable basis for comparing achievement across the participating countries. The analysis shows that omitting items considered by one country to be difficult for their students tends to improve the results for that country, but also tends to improve the results for all other countries as well, so that the overall pattern of relative performance is largely unaffected.

References

Beaton, A.E. & Gonzalez, E. J. (1997). TIMSS Test-Curriculum Matching Analysis. In M. O. Martin & D.L. Kelly (Eds.), *TIMSS technical report, volume II: Implementation and analysis.* Chestnut Hill, MA: Boston College.

Gregory, K. D. & Mullis, I. V. S. (2000). Describing international benchmarks of student achievement. In M. O. Martin, K. D. Gregory, & S. E. Stemler (Eds.), *TIMSS 1999 technical report*. Chestnut Hill, MA: Boston College.

Gonzalez, E. J. & Kennedy, A.M. (2003). Statistical analysis and reporting of the PIRLS data. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2001 technical report*. Chestnut Hill, MA: Boston College.

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

Kelly, D. L. (1999). *Interpreting the Third International Mathematics and Science Study (TIMSS) achievement scales using scale anchoring*. Unpublished doctoral dissertation, Boston College.

Mislevy, R.J., Beaton, A.E., Kaplan, B., and Sheenan, K.M. (1992). Estimating Population Characteristics from Sparse Matrix Samples of Item Responses, *Journal of Educational Measurement*, 29, 133-161.

Martin, M.O., Mullis, I.V.S., Gonzalez, E. J., Chrostowski, S.J. (2004). *TIMSS 2003 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: Boston College.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Chrostowski, S. J. (2004). *TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: Boston College.

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

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

13

Chapter 13

Reporting TIMSS 2003 Questionnaire Data

María José Ramírez and Alka Arora

13.1 Overview

The purpose of TIMSS is to provide information that policymakers, curriculum specialists, and researchers can use to understand better the performance of their educational systems. With this aim, TIMSS collects data on hundreds of contextual variables from nationally representative samples of students, their science and mathematics teachers, and their schools. Once the data are collected, one of the major challenges for TIMSS is reporting this vast array of information in a useful and meaningful way. The challenge is to focus on the most important educational contexts, inputs, and processes without overburdening the audiences with unmanageable amounts of information. TIMSS strives to report educational indicators that are easy to understand and interpret by policymakers and school personnel.

This chapter documents the analysis and reporting procedures used for the background questionnaire data in producing the TIMSS 2003 International Reports in mathematics and science. It provides an overview of the consensus process used to develop the report outlines and prototype exhibits; explains how single- and multiple-item indicators from the student, teacher, and school data were developed and computed; describes methods used by TIMSS to compute these indicators; and details the analysis and reporting of curriculum data. The final section explains how the data are displayed in the exhibits, and addresses issues regarding the unit of analysis, trend data, and response rates.

13.2 General Procedures

As described in Chapter 3, TIMSS 2003 used four types of questionnaires at both the fourth and eighth grades to gather information at various levels of the educational system:

- Student Questionnaire (separate versions for general/integrated science countries and separate science countries at eighth grade)
- Teacher Questionnaire (separate versions for mathematics and science at eighth grade)
- School Questionnaire
- Curriculum Questionnaire (separate versions for mathematics and science at both eighth and fourth grades)

The TIMSS & PIRLS International Study Center (ISC) at Boston College produced data almanacs summarizing the basic data from the student, teacher, and school questionnaires. For each participating country, these almanacs presented descriptive statistics for each question (variable) in the survey instruments. The statistics included the percentages of students checking each response option for categorical and ordinal data, as well as means, standards deviations, and percentile scores for continuous data. The almanacs were distributed periodically to the National Research Coordinators (NRCs) for review. Each time, a new data version was provided with more cases and updated cleaning rules and corrections implemented.

The ISC began working on the analysis of background data in May 2003. The main steps involved in this process were as follows. First, the TIMSS 2003 questionnaires were reviewed in the light of the contextual framework (see Chapter 3) to identify major conceptual categories or constructs that would enable a better understanding of the participating countries' educational systems and a fuller interpretation of their students' achievement in mathematics and science. Second, an outline describing the chapters and exhibits to be included in the TIMSS 2003 International Reports was prepared. Third, questions that could be used to measure the constructs of interest were identified, and extensive exploratory data analysis was conducted to decide what information to show and how to display it in each of the exhibits of the International Reports.

At the time the ISC started working on the reporting of data from the background questionnaires, data from the countries that operated with the southern hemisphere schedule were available for preliminary analyses.[1] These countries – Australia, Botswana, Chile, Malaysia, New Zealand, Singapore, and South Africa – provided data from some 40,000 students covering the entire spectrum of achievement on the TIMSS 2003 assess-

1 Countries that used the southern hemisphere schedule collected their data during September-November 2002, approximately six months earlier than countries using the northern hemisphere schedule.

ment and representing great cultural diversity. The preliminary analyses used background data from the Student Questionnaire (general version), Mathematics Teacher Questionnaire, and School Questionnaire from the TIMSS 2003 eighth-grade population.

As a first step, staff at the ISC reviewed the data thoroughly to ensure its quality. Descriptive analyses were run for each country separately, as well as for all the countries together. Statistics showing total number of cases, response rates, mean scores, standard deviations, and minimum and maximum scores were computed. For open-ended questions, ranges of valid responses were defined. When there were questions about the data, the national versions of the questionnaires were reviewed, and in some cases the NRC was contacted for further clarifications. As a result of this data review, the IEA Data Processing Center (DPC) in Hamburg implemented a number of revisions to the data cleaning rules.

Several preliminary versions of the indicators were developed and reviewed at the ISC. As explained in the following section, TIMSS 2003 used three methods for reporting background data: the direct reporting method (for single-item indicators), the scale method, and the combination of responses method (for multiple-item indicators). At this exploratory stage, all the analyses were run on unweighted data, using the first plausible value for mathematics as a criterion.[2] All the programming at this stage was done using SPSS version 11.5 (SPSS Inc., Chicago IL).

Once there was a clearer idea about how to combine the data into multiple-item indicators, the analyses were extended and adapted to the TIMSS 2003 fourth-grade population as well as to the science-specific instruments – Student Questionnaire (integrated science), Student Questionnaire (separate science subjects) and Science Teacher Questionnaire. All the indicators were reviewed for their effectiveness in providing information about educational contexts in the participating countries. Starting in October 2003, data from the northern hemisphere countries became available and was included in the analyses. The suitability of the preliminary indicators was checked again for these additional countries, and changes in the measures were made as necessary.

For each exhibit (table or figure) in the International Reports, analysis notes were created to document how the data were to be analyzed. These notes identified the source questions used to gather the data, explained how the data were processed before reporting, and described how the data would be displayed in the exhibits. The analysis notes also served as directions for programming the analyses in SAS version 9.0 (SAS Institute Inc., Cary NC),

2 See Chapters 11 and 12 for more information on plausible values.

the software used by TIMSS in implementing the data analysis. The exhibits in the International Reports were produced in SAS using all five plausible values in the TIMSS 2003 dataset, and standard errors were computed using the jackknife procedure (see Chapter 12). Based on the analysis notes also, the graphic production staff at the ISC designed and prepared prototype exhibits to display the background information.

Representatives from the participating countries reviewed the outlines for the International Reports, the proposed exhibits and indicators, and the analysis notes at the seventh NRC meeting held in Cape Town, South Africa, in November 2003. At that time, although data were available for just a few countries, they were useful in providing a sense of how the complex exhibits would look. NRCs approved the report outlines and almost all the proposed indicators; revisions were required in some exhibits based on suggestions for improvements from NRCs.

In January 2004, the ISC posted to its website revised Chapter 4 (Mathematics/Science Student Background) exhibits for the NRCs to review. Weighted data from 45 countries at the eighth-grade and 22 countries at the fourth-grade were available at that time. In March 2004, a revised version of the exhibits in Chapter 5 (Mathematics/Science Curriculum), Chapter 6 (Teachers of Mathematics/Science), Chapter 7 (Instruction in Mathematics/Science), and Chapter 8 (Mathematics/Science School Context) were posted to the ISC website, together with updated analyses notes. NRCs reviewed their national data and informed the ISC about any problems or anomalies that required further attention. In the meantime, staff at the ISC continued checking the data. All analyses were conducted in SAS, and repeated independently in SPSS to ensure that the same results were obtained.

The penultimate version of the TIMSS background exhibits was presented at the eighth NRC meeting held in Santiago, Chile, in June 2004. Country representatives reviewed their data and approved the exhibits for the International Reports. In a few cases, changes in the exhibits' format and type of information displayed were requested. NRCs informed the ISC about any questionable results that required further examination. After the meeting, staff at the ISC made final revisions to the exhibits.

Once the final exhibits of the background chapters were available, the companion text for those chapters was written. The background chapters with final exhibits and draft text were posted to the ISC website from August 16-30, 2004. NRCs reviewed the text and shared their comments with the ISC.

13.3 Methods for Reporting Background Data

This section describes the specific methods used to report TIMSS 2003 questionnaire data: the direct reporting method (for single-item indicators); scale method and combination of responses method (for multiple-item indicators).

13.3.1 Direct Reporting Method

Direct reporting was the simplest method used by TIMSS to report background data. The direct reporting method simply used the response categories in the questionnaires as reporting categories in the exhibits in the International Reports. In some cases, slight modifications were introduced: some response categories were collapsed, or were presented in a different order. Although the direct reporting method had the advantage of simplicity, it would have been impossible to report the vast amount of information collected by TIMSS in this way. Some data reduction was required, necessitating the use of more sophisticated approaches, as described below.

13.3.2 Methods for Computing Multiple-Item Indicators

Around one-fourth of the exhibits in the TIMSS 2003 International Reports were multiple-item indicators (derived variables) that combined data from several questions in the TIMSS 2003 questionnaires. Multiple-item indicators were used with complex constructs, such as the teacher's emphasis on mathematics homework, or school climate. Because the source items making up a multiple-item indicator target different facets of the construct, these measures can provide a more global and thorough picture of the phenomenon being studied than can single variables. Multiple-item indicators also have the advantage of providing more reliable measures of the construct, since random errors tend to cancel out when data are combined from different sources (see DeVellis, 1991; Spector, 1992).

Multiple-item indicators maximize the information that can be preserved in the presence of missing data. TIMSS required that at least two-thirds of the component questions have valid responses before computing an index. For instance, if an index was based on five questions, this rule allowed for one missing response only.

The starting point for creating a multiple-item indicator was to identify the questions in the TIMSS 2003 questionnaires that were related to the construct of interest. In some cases, these source questions were all sub-items of a more general question, and all had the same format. In other cases, the source questions came from different parts of the questionnaires, and did not share the same format. Depending upon the construct of interest and the item formats, TIMSS used two different methods to create derived variables: the scale method and the combination of responses method.

13.3.2.1 Scale Method

The "scale method" was used when the construct of interest had an underlying quantitative continuum. For example, schools can have a better or a worse climate for learning, or students can have higher or lower self-confidence in learning science. The scale method also required that all the questions (items) have the same number of response categories. These conditions allowed data to be combined from several items into one underlying scale while retaining the original metric of the items.

Before combining data from different questions, TIMSS gathered evidence that the source questions had the expected relationship with the achievement scores. For instance, it was expected that students who agreed with a statement such as "I usually do well in mathematics" would have higher mathematics scores than students who disagreed with the statement. Descriptive statistics, analysis of variance (ANOVA), and eta-squared (η^2) were useful in assessing whether the expected relationships held true (see Hinkle, Wiersma & Jurs, 1998, pp. 565-569; Pedhazur, 1997, pp. 355, 505-507).

Questions addressing a construct were expected to be correlated in the data. Chi-square (χ^2) and Spearman's rank order correlation coefficient were used to measure the association between pairs of categorical or ordinal items. Principal component analysis (PCA) was used to identify questions related to a common construct. Building on these analyses, new variables (components) were created that accounted for most of the variance in the source items.

Once there was enough evidence that a set of questions or items was measuring the construct of interest, TIMSS examined the reliability of a scale made up from these items. Cronbach's alpha (α) was used to measure the internal consistency of these scales; item-total correlations (or point-biserial correlations) were used to identify questions that did not cluster together with the others.

Using the scale method, TIMSS computed index scores by averaging the numerical values associated with each response option. This procedure had the advantage of preserving the original scale categories, thus allowing for a straightforward interpretation of the index scores. The TIMSS 2003 questionnaires made extensive use of the 4-point Likert scale format, with "*strongly agree*" coded 4, "*agree*" coded 3, "*disagree*" coded 2, and "*strongly disagree*" coded 1. Before averaging the scores associated with the responses,

responses were recoded as necessary, with items coded so that high scores were associated with the response category indicating higher levels of the attribute being measured.

Whenever the scale method was used to create an index, TIMSS classified the students into three levels: high, medium, and low. In the International Reports, these derived variables are referred to as indices. To classify the cases into three groups, two cutoff points were established. Three main criteria were used in setting the cutoff points. First, the high level of the index should correspond to conditions or activities generally associated with good educational practice or high academic achievement. Second, there should be a reasonably even distribution of students across the three index levels. Third, the scale categories should be about the same size.

Once the cutoff points were defined, a critical step was to check the overall quality of the indices. Indices were intended to discriminate among students with high and low achievement. The extent of the association with achievement was measured using eta-squared (η^2). This was computed for each country separately and for all the countries together. Only indices that discriminated reasonably well in most of the participating countries were included in the International Reports.

Line graphs plotting mean achievement by index level also were useful in checking the hypothesized positive association between index levels and achievement scores. The slope of the line joining the means served as an indicator of how well the index discriminated among students with different achievement levels. The steeper the line the greater were the differences between the average achievement scores of one index level and the next.

13.3.2.2 Combination of Responses Method

TIMSS also made extensive use of the “combination of responses method” to construct indices. Cases were classified into the high, medium, or low level of an index depending upon the combination of responses provided to the source items. For example, in the index of *Good School and Class Attendance*, cases were classified into the high index level if the three source items (arriving late at school, absenteeism, and skipping classes) were reported to be *not a problem*. Cases went to the low index level when two or more behaviors were reported to be a *serious problem* or two behaviors were reported to be a *minor problem* and the third a *serious problem*. The medium level included all other combinations of responses.

In addition to constructing indices, the combination of responses method also was used to construct some specific derived variables. An example is students' *Use of Computer*. Students were asked if they use a computer "at home," "at school," "at a library," "at a friend's house," "at an Internet cafe," or "elsewhere." The reporting categories for this derived variable were "use computer both home and at school," "use computer at home but not at school," "use computer at school but not at home," "use computer only at places other than home and school," and "do not use computer at all."

13.3.2.3 Summary of Derived Variables in the TIMSS 2003 International Reports

The TIMSS 2003 International Reports in mathematics and science each present some 60 exhibits with background information, providing data on some 250 indicators. The mathematics report presents data on 17 derived variables and the science report on 16; each report includes 11 indices. Exhibits 13.1 and 13.2 list the indices computed for the TIMSS 2003 International Reports in mathematics and science, respectively. Exhibit 13.3 lists the other derived variables presented in the mathematics and science reports. The name of the indicators, the label used to identify them in the International Reports and database, the mathematics or science exhibit where the data are reported, and the analysis method used to compute the data are provided.

13.4 Analysis of Curriculum Data

The Mathematics and Science Curriculum Questionnaires were used to collect information about the intended curriculum in each participating country. The NRC for each country, with the help of curriculum specialists, completed curriculum questionnaires for the grade assessed (fourth grade and/or eighth grade). Chapter 5 in the TIMSS 2003 International Reports combined data from the Curriculum Questionnaires and the Teacher Questionnaire to inform about both the intended and implemented Mathematics and Science curricula in the participating countries. The following information was presented:

- Existence of a national curriculum, the year it was introduced, and whether it was under revision
- Methods used to support and monitor curriculum implementation
- Use of public examinations and grades tested
- Instructional time intended for mathematics and science
- Differentiation of curriculum for students with different levels of ability
- Emphasis on different approaches and processes in the intended curriculum (e.g., knowing facts, understanding concepts)
- Coverage of the TIMSS 2003 topics in the intended and implemented curriculum
- Science subjects offered through the eighth grade (science only)

Exhibit 13.1 **Summary Indices in the TIMSS 2003 International Mathematics Report**

Index	Analysis Method
Exhibit 4.7 Index of Time Students Spend Doing Mathematics Homework (TMH)	Index based on students' reports on the frequency and amount of mathematics homework they are given. High level indicates more than 30 minutes of mathematics homework assigned 3-4 times a week. Low level indicates no more than 30 minutes of mathematics homework no more than twice a week. Medium level includes all other possible combinations of responses.
Exhibit 4.9 Index of Students' Self-Confidence in Learning Mathematics (SCM)	Index based on students' responses to four statements about mathematics: 1) I usually do well in mathematics; 2) Mathematics is more difficult for me than for many of my classmates (Reversed); 3) Mathematics is not one of my strengths (Reversed); 4) I learn things quickly in mathematics. Average is computed across the four items based on a 4-point scale: 1. Agree a lot; 2. Agree a little; 3. Disagree a little; 4. Disagree a lot. Students agreeing a little or a lot on average across the four statements are assigned to the high level. Students disagreeing a little or a lot on average are assigned to the low level. All other students are assigned to the middle level.
Exhibit 4.10 Index of Students' Valuing Mathematics (SVM) (Grade 8 only)	Index based on students' responses to seven statements about mathematics: 1) I would like to take more mathematics in school; 2) I enjoy learning mathematics; 3) I think learning mathematics will help me in my daily life; 4) I need mathematics to learn other school subjects; 5) I need to do well in mathematics to get into the university of my choice; 6) I would like a job that involved using mathematics; 7) I need to do well in mathematics to get the job I want. Average is computed across the seven items based on a 4-point scale: 1. Agree a lot; 2. Agree a little; 3. Disagree a little; 4. Disagree a lot. Students agreeing a little or a lot on average across the seven statements are assigned to the high level. Students disagreeing a little or a lot on average are assigned to the low level. All other students are assigned to the middle level.
Exhibit 7.2 Index of Teachers' Reports on Teaching Mathematics Classes with Few or No Limitations on Instruction due to Student Factors (MCFL) (Grade 8 only)	Index based on teachers' responses to six statements about student factors limiting mathematics instruction: 1) Students with different academic abilities; 2) Students who come from a wide range of backgrounds; 3) Students with special needs; 4) Uninterested students; 5) Low morale among students; 6) Disruptive students. Average is computed across the six statements based on a 4-point scale: 1. Not at all/Not applicable; 2. A little; 3. Some; 4. A lot. High level indicates average is less than or equal to 2. Medium level indicates average is greater than 2 and less than 3. Low level indicates average is greater than or equal to 3.
Exhibit 7.13 Index of Teachers' Emphasis on Mathematics Homework (EMH)	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. High level indicates the assignment of more than 30 minutes of homework about half of the lessons or more. Low level indicates no assignment or the assignment of less than 30 minutes of homework about half of the lessons or less. Medium level includes all other possible combinations of responses.
Exhibit 8.3 Index of Availability of School Resources for Mathematics Instruction (ASRMI)	Index based on principals' average response to five questions about shortages that affect general capacity to provide instruction: instructional materials (e.g., textbook); budget for supplies (e.g., paper, pencils); school buildings and grounds; heating/cooling and lighting systems; and instructional space (e.g., classrooms); and the average response to five questions about shortages that affect mathematics instruction: computers for mathematics instruction; computer software for mathematics instruction; calculators for mathematics instruction; library materials relevant to mathematics instruction; and audio-visual resources for mathematics instruction. Average is computed based on a 4-point scale: 1. None; 2. A little; 3. Some; 4. A lot. High level indicates that both shortages are on average lower than 2. Low level indicates that both shortages are on average greater than or equal to 3. Medium level includes all other possible combinations of responses.

Exhibit 13.1 **Summary Indices in the TIMSS 2003 International Mathematics Report**
(...Continued)

Index	Analysis Method
Exhibit 8.4 Index of Principals' Perception of School Climate (PPSC)	Index based on principals' responses to eight questions about their schools: teachers' job satisfaction; teachers' understanding of the school's curricular goals; teachers' degree of success in implementing the school's curriculum; teachers' expectations for student achievement; parental support for student achievement; parental involvement in school activities; students' regard for school property; and students' desire to do well in school. Average is computed based on a 5-point scale: 1. Very high; 2. High; 3. Medium; 4. Low; 5. Very low. High level indicates average is less than or equal to 2. Medium level indicates that average is greater than 2 and less or equal to 3. Low level indicates average is greater than 3.
Exhibit 8.5 Index of Mathematics Teachers' Perception of School Climate (TPSC)	Index based on teachers' responses to eight questions about their schools: teachers' job satisfaction; teachers' understanding of the school's curricular goals; teachers' degree of success in implementing the school's curriculum; teachers' expectations for student achievement; parental support for student achievement; parental involvement in school activities; students' regard for school property; and students' desire to do well in school. Average is computed based on a 5-point scale: 1. Very high; 2. High; 3. Medium; 4. Low; 5. Very low. High level indicates average is less than or equal to 2. Medium level indicates that average is greater than 2 and less or equal to 3. Low level indicates average is greater than 3.
Exhibit 8.6 Index of Good School and Class Attendance (GSCA)	Index based on principals' responses to three questions about the seriousness of attendance problems in the school: arriving late at school; absenteeism (i.e., unjustified absences); and skipping class. High level indicates that all three behaviors either never occur or are reported not to be 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.
Exhibit 8.7 Index of Mathematics Teachers' Perception of Safety in the Schools (TPSS)	Index based on teachers' responses to three statements about their schools: this school is located in a safe neighborhood; I feel safe at this school; this school's security policies and practices are sufficient. High level indicates that the teacher agrees a lot or agrees to all three statements. Low level indicates that teacher disagrees or disagrees a lot to all three statements. Medium level includes all other combinations of responses.
Exhibit 8.8 Index of Students' Perception of Being Safe in the Schools (SPBSS)	Index based on students' responses to five statements about things that happened in their schools in the last month (1 = yes, 2 = no): something of mine was stolen; I was hit or hurt by other student(s) (e.g., shoving, hitting, kicking); I was made to do things that I didn't want to do by other students; I was made fun of or called names; I was left out of activities by other students. High level indicates that the student answered NO to all five statements. Low level indicates that the student answered YES to three or more statements. Medium level includes all other possible combinations of responses.

Note: Detailed information about the computation of indices can be found in the TIMSS 2003 User Guide.

Exhibit 13.2 **Summary Indices in the TIMSS 2003 International Science Report**

Index	Analysis Method
Exhibit 4.7 Index of Time Students Spend Doing Science Homework (TSH)	Index based on students' reports on the frequency and amount of science homework they are given. High level indicates more than 30 minutes of science homework assigned 3-4 times a week. Low level indicates no more than 30 minutes of science homework no more than twice a week. Medium level includes all other possible combinations of responses.
Exhibit 4.9 Index of Students' Self-Confidence in Learning Science (SCS)	Index based on students' responses to four statements about science: 1) I usually do well in science; 2) Science is more difficult for me than for many of my classmates (Reversed); 3) Science is not one of my strengths (Reversed); 4) I learn things quickly in science. Average is computed across the four items based on a 4-point scale: 1. Agree a lot; 2. Agree a little; 3. Disagree a little; 4. Disagree a lot. Students agreeing a little or a lot on average across the four statements are assigned to the high level. Students disagreeing a little or a lot on average are assigned to the low level. All other students are assigned to the middle level.
Exhibit 4.10 Index of Students' Valuing Sciences (SVS) (Grade 8 only)	Index based on students' responses to seven statements about science: 1) I would like to take more science in school; 2) I enjoy learning science; 3) I think learning science will help me in my daily life; 4) I need science to learn other school subjects; 5) I need to do well in science to get into the university of my choice; 6) I would like a job that involved using science; 7) I need to do well in science to get the job I want. Average is computed across the seven items based on a 4-point scale: 1. Agree a lot; 2. Agree a little; 3. Disagree a little; 4. Disagree a lot. Students agreeing a little or a lot on average across the seven statements are assigned to the high level. Students disagreeing a little or a lot on average are assigned to the low level. All other students are assigned to the middle level.
Exhibit 7.2 Index of Teachers' Reports on Teaching Science Classes with Few or No Limitations on Instruction due to Student Factors (SCFL) (Grade 8 only)	Index based on teachers' responses to six statements about student factors limiting science instruction: 1) Students with different academic abilities; 2) Students who come from a wide range of backgrounds; 3) Students with special needs; 4) Uninterested students; 5) Low morale among students; 6) Disruptive students. Average is computed across the six statements based on a 4-point scale: 1. Not at all/Not applicable; 2. A little; 3. Some; 4. A lot. High level indicates average is less than or equal to 2. Medium level indicates average is greater than 2 and less than 3. Low level indicates average is greater than or equal to 3.
Exhibit 7.10 Index of Teachers' Emphasis on Science Homework (ESH)	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. High level indicates the assignment of more than 30 minutes of homework about half of the lessons or more. Low level indicates no assignment or the assignment of less than 30 minutes of homework about half of the lessons or less. Medium level includes all other possible combinations of responses.
Exhibit 8.3 Index of Availability of School Resources for Science Instruction (ASRSI)	Index based on principals' average response to five questions about shortages that affect general capacity to provide instruction: instructional materials (e.g., textbook); budget for supplies (e.g., paper, pencils); school buildings and grounds; heating/cooling and lighting systems; and instructional space (e.g., classrooms); and the average response to six questions about shortages that affect science instruction: science laboratory equipment and materials; computers for science instruction; computer software for science instruction; calculators for science instruction; library materials relevant to science instruction; and audio-visual resources for science instruction. Average is computed based on a 4-point scale: 1. None; 2. A little; 3. Some; 4. A lot. High level indicates that both shortages are on average lower than 2. Low level indicates that both shortages are on average greater than or equal to 3. Medium level includes all other possible combinations of responses.

Exhibit 13.2 **Summary Indices in the TIMSS 2003 International Science Report**
(…Continued)

Index	Analysis Method
Exhibit 8.4 Index of Principals' Perception of School Climate (PPSC)	Index based on principals' responses to eight questions about their schools: teachers' job satisfaction; teachers' understanding of the school's curricular goals; teachers' degree of success in implementing the school's curriculum; teachers' expectations for student achievement; parental support for student achievement; parental involvement in school activities; students' regard for school property; and students' desire to do well in school. Average is computed based on a 5-point scale: 1. Very high; 2. High; 3. Medium; 4. Low; 5. Very low. High level indicates average is less than or equal to 2. Medium level indicates that average is greater than 2 and less or equal to 3. Low level indicates average is greater than 3.
Exhibit 8.5 Index of Science Teachers' Perception of School Climate (TPSC)	Index based on teachers' responses to eight questions about their schools: teachers' job satisfaction; teachers' understanding of the school's curricular goals; teachers' degree of success in implementing the school's curriculum; teachers' expectations for student achievement; parental support for student achievement; parental involvement in school activities; students' regard for school property; and students' desire to do well in school. Average is computed based on a 5-point scale: 1. Very high; 2. High; 3. Medium; 4. Low; 5. Very low. High level indicates average is less than or equal to 2. Medium level indicates that average is greater than 2 and less or equal to 3. Low level indicates average is greater than 3.
Exhibit 8.6 Index of Good School and Class Attendance (GSCA)	Index based on principals' responses to three questions about the seriousness of attendance problems in the school: arriving late at school; absenteeism (i.e., unjustified absences); and skipping class. High level indicates that all three behaviors either never occur or are reported not to be 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.
Exhibit 8.7 Index of Science Teachers' Perception of Safety in the Schools (TPSS)	Index based on teachers' responses to three statements about their schools: this school is located in a safe neighborhood; I feel safe at this school; this school's security policies and practices are sufficient. High level indicates that the teacher agrees a lot or agrees to all three statements. Low level indicates that teacher disagrees or disagrees a lot to all three statements. Medium level includes all other combinations of responses.
Exhibit 8.8 Index of Students' Perception of Being Safe in the Schools (SPBSS)	Index based on students' responses to five statements about things that happened in their schools in the last month (1 = yes, 2 = no): something of mine was stolen; I was hit or hurt by other student(s) (e.g., shoving, hitting, kicking); I was made to do things that I didn't want to do by other students; I was made fun of or called names; I was left out of activities by other students. High level indicates that the student answered NO to all five statements. Low level indicates that the student answered YES to three or more statements. Medium level includes all other possible combinations of responses.

Note: Detailed information about the computation of indices can be found in the TIMSS 2003 User Guide

Exhibit 13.3 **Summary of Derived Variables Other than Indices in the TIMSS 2003 International Mathematics and Science Reports**

Derived Variable	Analysis Method
Exhibit 4.1 Highest Level of Education of Either Parent (Grade 8 only)	Derived variable based on students' responses to the highest level of education of mother and father. Cases classified in four categories: 1. Finished University or Equivalent or Higher 2. Finished Post-secondary Vocational/Technical Education but Not University 3. Finished Upper Secondary Schooling 4. Finished Lower Secondary Schooling 5. No More Than Primary Schooling
Exhibit 4.2 Students' Educational Aspirations Relative to Parents' Educational Level (Grade 8 only)	Derived variable based on students' responses to the highest level of education of mother and father, and students' expectations for further education. Cases were classified in four categories: 1. Finish University and Either Parent Went to University or Equivalent 2. Finish University but Neither Parent Went to University Equivalent 3. Not Finish University Regardless of Parents' Education 4. Do Not Know Regardless of Parents' Education
Exhibit 4.6 Use of Computer	Derived variable based on students' responses to where do they use a computer. Cases were classified in five categories: 1. Use Computer Both at Home and at School 2. Use Computer at Home but Not at School 3. Use Computer at School but Not at Home 4. Use Computer only at places other than home and school 5. Do Not Use Computer at All
Exhibit 6.5 Preparation to Teach Mathematics (Grade 4 only)*	Derived variable based on teachers' responses to main area of study during post-secondary education, and main area in specialization. Cases were classified in five categories: 1. Primary/Elementary Education with a Major or Specialization in Mathematics 2. Primary/Elementary Education with a Major or Specialization in Science but Not in Mathematics 3. Mathematics or Science Major or Specialization without a Major in Primary/Elementary Education 4. Primary/Elementary Education without a Major or Specialization in Mathematics or Science 5. Other
Exhibit 6.5 Preparation to Teach Science (Grade 4 only)*	Derived variable based on teachers' responses to main area of study during post-secondary education, and main area in specialization. Cases were classified in five categories: 1. Primary/Elementary Education with a Major or Specialization in Mathematics 2. Primary/Elementary Education with a Major or Specialization in Science but Not in Mathematics 3. Mathematics or Science Major or Specialization without a Major in Primary/Elementary Education 4. Primary/Elementary Education without a Major or Specialization in Mathematics or Science 5. Other

Note: Detailed information about the computation of indices can be found in the TIMSS 2003 User Guide
* At grade 8, "Preparation to teach" was reported using the direct reporting method.

In general, information from the curriculum questionnaires was directly reported in the exhibits. The information extracted from these questionnaires is mostly textual and qualitative in nature. In the case of quantitative information, descriptive statistics were provided. NRCs reviewed and approved the display of the curriculum information at the seventh NRC meeting. At that time, exhibits with data were available only for the mathematics curriculum at the eighth grade. After that meeting, ISC staff implemented the suggested changes to the curriculum exhibits, and completed them for both grades and subjects. Given the qualitative nature of the curriculum data, extensive follow-up and data cleaning were required. From January to June 2004, ISC staff carefully reviewed the curriculum data and asked NRCs to provide missing data, correct inconsistent data, and clarify questionable data. The final version of the curriculum exhibits was presented and approved at the eighth NRC meeting, when any lingering questions about the curriculum data were resolved.

13.5 Display of Background Data

TIMSS 2003 results were reported separately by subject area, with the mathematics and science results appearing in separate reports. Final exhibits with background data were organized into chapters 4 through 8 in the International Reports (the first three chapters reported achievement data). Chapter 4 reported data on students' characteristics, Chapter 5 on the curriculum, Chapter 6 on teachers' characteristics, Chapter 7 on instructional practices, and Chapter 8 on the schools.

It is important to note that in the data reported in the exhibits the student was always the unit of analysis, even when information from the teacher or school questionnaire was reported. In general, the exhibits presented the percentage of students having certain characteristics, or the percentage of students whose teachers or schools have various characteristics. For example, the International Reports give the percentage of students taught by teachers having a teaching certificate. This approach is consistent with the main goal of TIMSS, which is to inform about students' educational contexts and performance. The percentages in the exhibits were often accompanied by the students' mean achievement (mathematics or science). Information for each country was presented in individual rows, with the international average for all the participating countries (mean of countries' means) displayed separately. In general, where only one variable with several categories was reported in an exhibit, countries were displayed in rank order based on one of the categories, and where more than one variable was reported, countries were displayed in alphabetical order.

Whenever possible and relevant, the International Reports included trend data from 1995 (fourth and eighth grades) and 1999 (eighth grade only). Significant differences between the percentages of students having a given trait in each cycle were indicated. In other exhibits, data were displayed separately for boys and girls, and significant differences were also indicated.

In the science report, eighth grade background information was reported separately for the integrated science countries and for the separate science countries. The integrated science countries were reported in a “General/ Integrated Science” panel. The separate science countries were reported in four different panels: Biology, Earth Science, Chemistry, and Physics.

The exhibits in the International Reports contained special notations regarding response rates for the background variables. Although in general there were high response rates, some indicators and some countries had less than acceptable response rates. Since the student was the unit of analysis, the notation used in the International Reports always reflected the percentage of students for whom the responses from students, teachers, or schools were available. The following special notations were used to convey information about response rates in the exhibits in the International Reports:

- For a country where student, teacher, or school responses were available for 70 to 85 percent of the students, an “r” appeared next to the data for that country.
- Where student, teacher, or school responses were available for 50 to 69 percent of the students, an “s” appeared next to the data for that country.
- Where student, teacher, or school responses were available for less than 50 percent of the students, “x” replaced the data.
- Where the percentage of students in a particular category was less than two percent, achievement data were not reported in that category; the data were replaced by a tilde (~).
- Where data were not comparable for all respondents in a country, a dash (–) was used in place of data in all of the affected columns.[3]

3 A dash usually indicates that a background question was not administered in a country, but could also be due to translation problems or to the administration of a question that was determined to be not internationally comparable. In the exhibits based on the separate science subjects, the inclusion of dashes for specific countries is by design and reflects the specific science subjects not included in each country.

References

DeVellis, R. (1991). *Scale development, theory and applications*. Sage Publications on Applied Social Research Method Series Vol. 26. Newburry Park, California: Sage Publications, Inc.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1998). *Applied statistics for the behavioral sciences* (4th ed.). Boston: Houghton Mifflin Company.

Pedhazur, E. (1997). *Multiple regression in behavioral research* (3rd ed.). Boston: Harcourt Brace College Publishers.

SAS Institute (2002). *SAS system for Windows* (version 9). Cary, NC: SAS Institute.

Spector, P. (1992). *Summated rating scale construction, an introduction*. Sage University Papers Series on Quantitative Applications in the Social Sciences, Series No. 07-082. Beverly Hills, California: Sage University Press.

SPSS Inc. (2002). *SPSS for Windows* (version 11.5). CHicago, IL: SPSS Inc.

A

Appendix A
Acknowledgements

Developing and implementing TIMSS 2003 was an extremely ambitious and truly collaborative effort involving 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 TIMSS & PIRLS International Study Center (ISC) at Boston College, advisors, and funding agencies worked closely to develop and implement TIMSS 2003. The project would not have been possible without the tireless efforts of all involved. Below, the individuals and organizations are acknowledged for their contributions. Given that implementing TIMSS 2003 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 2003 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.

Funding Agencies

Funding for the international coordination of TIMSS 2003 was provided by the National Center for Education Statistics of the US Department of Education, the US National Science Foundation, the World Bank, the United Nations Development Programme (UNDP), Boston College, and participating countries. Valena Plisko, Patrick Gonzales, Elois Scott, and Eugene Owen of the National Center for Education Statistics; Janice Earle, Larry Suter, Finbarr Sloane, and Elizabeth VanderPutten of the National Science Foundation; Marlaine Lockheed of the World Bank; and Maen Nsour of the UNDP each played a crucial role in making TIMSS 2003 possible and for ensuring the quality of the study. Each participating country was responsible for funding national project costs and implementing TIMSS 2003 in accordance with the international procedures.

Management and Operations

TIMSS 2003 was conducted under the auspices of the IEA. The study was directed by Michael O. Martin and Ina V.S. Mullis, and managed centrally by the staff of the TIMSS & PIRLS 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 2003, 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 managed the ambitious translation verification effort conducted for the field test and main assessment and recruited international quality control monitors in each country. The IEA Data Processing Center was responsible for processing and verifying the data from the participating countries and for constructing the international database. Statistics Canada was responsible for collecting and evaluating the sampling documentation from each country and for calculating the sampling weights. Educational Testing Service in Princeton, New Jersey provided consultation on psychometric issues as well as technical support and software for scaling the achievement data. The Project Management Team, comprising the study directors and representatives from the International Study Center, IEA, Statistics Canada, and Educational Testing Service, met regularly throughout the study to discuss the study's progress, procedures, and schedule.

IEA Secretariat

Alejandro Tiana, IEA Chair
Hans Wagemaker, Executive Director
Barbara Malak-Minkiewicz, Manager Membership Relations
Juriaan Hartenberg, Financial Manager
Isabelle Braun-Gemin, Financial Manager Assistant
Katazyna Krohn, Management Assistant

TIMSS & PIRLS International Study Center at Boston College

Michael O. Martin, Co-Director
Ina V.S. Mullis, Co-Director
Eugenio J. Gonzalez, Director of Operations and Data Analysis
Teresa Smith Neidorf, TIMSS Science Coordinator
Robert Garden, TIMSS Mathematics Coordinator
Steven Chrostowski, TIMSS Project Coordinator

Ann Kennedy, PIRLS Project Coordinator
Joseph Galia, Senior Statistician/Programmer
Isaac Li, Statistician/Programmer
Dana Diaconu, Research Associate
Ebru Erberber, Research Associate
Cheryl Flaherty, Research Associate
Alka Arora, Graduate Assistant
María José Ramírez, Graduate Assistant
Keith Morgan, Publications Design & Production Manager (2003 to present)
José Nieto, Publications Production Manager (until 2003)
Mario Pita, Data Graphics Specialist
Betty Hugh, Data Graphics Specialist
Susan Messner, Graphics Specialist
Christine Hoage, Manager, Finance
Marcie Petras, Manager, Office Administration
Laura Brown, Administrative Coordinator
Rita Holmes, Administrative Coordinator

IEA Data Processing Center

Dirk Hastedt, Senior Researcher
Pierre Foy, Senior Researcher
Oliver Neuschmidt, Researcher
Juliane Barth, Researcher
Ieva Johansone, Junior Researcher
Milena Taneva, Junior Researcher
Christine Busch, Junior Researcher
Dirk Oehler, Programmer
Stefan Petzchen, Programmer
Harpreet Singh Choudry, Programmer
Olaf Zühlke, Research Assistant
Warnuna Ratnayake, Research Assistant
Jayprakash Kooraram, Research Assistant
Marta Kostek Drosin, Research Assistant

Statistics Canada

Marc Joncas, Senior Methodologist

Educational Testing Service

John Barone, Executive Director, Center for Data Analysis Research
Matthias Von Davier, Principal Research Scientist
Ed Kulick, Manager, Research Data Analysis

Project Management Team

Michael O. Martin, International Study Center
Ina V.S. Mullis, International Study Center
Eugenio J. Gonzalez, International Study Center
Hans Wagemaker, IEA Secretariat
Dirk Hastedt, IEA Data Processing Center
Pierre Foy, IEA Data Processing Center
Marc Joncas, Statistics Canada
Matthias Von Davier, Educational Testing Service

TIMSS 2003 Special Consultants

Sampling Referee

Keith Rust, Westat, Inc.

Psychometric Design

Eugene Johnson, American Institutes for Research

IEA Editorial Review

David Robitaille, University of British Columbia

IEA Technical Executive Group

TIMSS 2003 Advisory Committees and Task Forces

The TIMSS & PIRLS International Study Center at Boston College was supported in its work by a number of advisory committees. The International Expert Panel in Mathematics and Science played a crucial role in developing the TIMSS 2003 frameworks and specifications for the assessment. The Mathematics and Science Item Development Task Forces coordinated the work of the national research coordinators in developing and reviewing the mathematics and science achievement items. The Science and Mathematics Item Review Committee reviewed and revised successive drafts of the achievement items and was an integral part of the scale anchoring process. The Questionnaire

Item Review Committee revised the TIMSS context questionnaires for the 2003 assessment.

International Expert Panel

Mathematics

Khattab Abu-Libdeh, Jordan
Anica Aleksova, Republic of Macedonia
Kiril Bankov, Bulgaria
Aarnout Brombacher, South Africa
Anna Maria Caputo, Italy
Joan Ferrini-Mundy, United States
Jim Fey, United States
Derek Holton, New Zealand
Jeremy Kilpatrick, United States
Pekka Kupari, Finland
Mary Lindquist, United States
David Robitaille, Canada
Graham Ruddock, England
Hanako Senuma, Japan

Science

K.Th. (Kerst) Boersma, the Netherlands
Rodger Bybee, United States
Audrey Champagne, United States
Reinders Duit, Germany
Martin Hollins, England
Eric Jakobsson, United States
Galina Kovalyova, Russian Federation
Svein Lie, Norway
Jan Lokan, Australia
Francisco Mazzitelli, Argentina
Gabriela Noveanu, Romania
Margery Osborne, United Status
Jana Paleckova, Czech Republic
Hong Kim Tan, Singapore
Khadija Zaim-Idrissi, Morocco

Mathematics Item Development Task Force

Robert Garden, New Zealand (Mathematics Coordinator)
Chancey Jones, United States
Graham Ruddock, England

Science Item Development Task Force

Teresa Smith Neidorf, United States (Science Coordinator)
Svein Lie, Norway
Christine O'Sullivan, United States

Science and Mathematics Item Review Committee

Mathematics

Anica Aleksova, Republic of Macedonia
Kiril Bankov, Bulgaria
Aarnout Brombacher, South Africa
Francine Jaques, Canada
Jeremy Kilpatrick, United States
Mary Lindquist, United States
Graham Ruddock, England
Hanako Senuma, Japan

Science

Audrey Champagne, United States
Chang Chu-Nan, Chinese Taipei
Galina Kovalyova, Russian Federation
Svein Lie, Norway
Jan Lokan, Australia
Francisco Mazzitelli, Argentina
Gabriela Noveanu, Romania
Ahmed Muhammed Rafea, Bahrain
Vivien Talisayon, Philippines
Sandy Tan, Singapore

Questionnaire Item Review Committee

Khattab Abu Lebdeh, Jordan
Klaas Bos, the Netherlands
Megan Chamberlain, New Zealand
Chiu Mei-Hung, Chinese Taipei
Rich Coley, United States

Patrick Gonzales, United States
Mike Marshall, Canada

National Research Coordinators

The TIMSS 2003 National Research Coordinators and their staff had the enormous task of implementing the TIMSS 2003 design. This involved 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 2003 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 some cases, the national centers were operating with limited resources. All of the TIMSS 2003 National Research Coordinators and their staff members are to be commended for their professionalism and their dedication in conducting all aspects of TIMSS.

ARGENTINA

Margarita Poggi
Ministerio de Educación
Dirección Nacional de Información y Evaluación
Paraguay 1657, 2er Piso – Of 201
Buenos Aires C1062ACA

ARMENIA, REPUBLIC OF

Arsen Baghdasaryan
Yerevan State University
26 Halabyan Str Apt 31
Yerevan 375036

AUSTRALIA

Sue Thomson
Australian Council for Educational Research(ACER)
19 Prospect Hill Rd.
Camberwell, Victoria 3124

BAHRAIN

Ahmed Muhammad Rafea
Chief, Assessment and Curriculum Development Division
PO Box 43
Ministry Education
Manama

BELGIUM (Flemish)

Christiane Brusselmans-Dehairs
Vakgroep Onderwijskunde Universiteit Gent
Henri Dunantlaan 2
B 9000 Gent

Ann Van Den Broeck
LIVO
Dekenstraat 2
Leuven B-3000

BOTSWANA

Cyprian Ismael Cele
Examinations Research and Testing Division
Ministry of Education
Private Bag 0070
Gaborone

BULGARIA

Kiril Bankov
Faculty of Mathematics and Informatics
University of Sofia
5, bul James Boucher
Sofia 1164

CHILE

Leonor Cariola Huerta
Ministerio de Educación
Alameda 1146
Piso 8, Sector B
Santiago

CHINESE TAIPEI

Chu-Nan Chang
Dean of College of Science
National Taiwan Normal University
88 Sec 4, Ting-Chou Road
Taipei, 116

CYPRUS

Constantinos Christou
Research and Evaluation Department
Pedagogical Institue
University of Cyprus-Dept. of Education
Kallipoleous 75, P.O. Box 20537
Nicosia 1678

EGYPT

Solaiman El-Khodary El-Sheikh
National Center of Examinations and Educational Evaluation
Hadaba Olya
Al-Mokkatam
Cairo

ENGLAND

Graham Ruddock
National Foundation for Educational Research (NFER)
The Mere, Upton Park
Slough, Berkshire
SL1 2DQ

ESTONIA

Kristi Mere
Estonian Ministry of Education
Munga 18
Tartu EE 50088

GHANA
Aba Mansa Folson
Head of Inspectorate Division
Ministry of Education
Ghana Education Service
Arakan Printing Press Building
Kotobabi, Accra

HONG KONG, SAR
Frederick Leung
The University of Hong Kong – Hong Kong IEA Centre
Pokfulam Road
Hong Kong, SAR

HUNGARY
Peter Vari
National Institute of Public Education
Centre for Evaluation Studies
Dorottya u.8, PF 701/420
Budapest 1051

INDONESIA
Jahja Umar
Ministry of National Education
Examiniation Development Center
Jalan Gunung Sahari Raya – 4
Jakarta Pusat, 1000 Jakarta

IRAN, ISLAMIC REPUBLIC OF
Abbass Rahiminezhad
Institute for National Research
Hojjat Doost Alley Naderi St.
196 Keshavarz Blvd.
Tehran

ISRAEL

Ruth Zuzovsky
Tel Aviv University
School of Education
Center for Science and Technology Education
Ramat Aviv, Tel Aviv 69978

ITALY

Anna Maria Caputo
Istituto Nazionale per la
Valutazione del Sistema dell'Istruzione (CEDE)
Borromini 5 – Villa Falconieri
Frascati (Roma) 00044

JAPAN

Yuji Saruta
National Institute for Educational Research (NIER)
6-5-22 Shimomeguro
Meguro-ku, Tokyo
153-8681

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 11941

KOREA, REPUBLIC OF

Chung Park
Korea Institute of Curriculum & Evaluation(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

LEBANON

Leila Maliha Fayad
Ministry of Education
The Educational Center for Research and Development
Dekwanen, Beirut

LITHUANIA

Algirdas Zabulionis
Ministry of Education and Science
National Examination Center
M. Katkaus 44
Vilnius LT2051

MACEDONIA, REPUBLIC OF

Anica Aleksova
Ministry of Education and Science
Bureau for Development of Education
Ruder Boskovic St bb
1000 Skopje

MALAYSIA

Azmi Zakaria
Ministry of Education
Educational Planning and Research Division
Level 2, 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

MOROCCO
Mohamed Sassi
Direction de l'Évaluation du Système Educatif
Ministère de l'Éducation Nationale
Innovations Pédagogiques
32 Boulevard Ibn Toumert
Place Bob Rouah, Rabat

THE NETHERLANDS
Martina Meelissen
University of Twente
Centre for Applied Research in Education(OCTO)
PO Box 217
7500 AE Enschede

NEW ZEALAND
Megan Chamberlain
Ministry of Education
CER Unit-Research Division
45-47 Pipitea Street
Thorndon, Wellington

Fiona Sturrock
Ministry of Education
CER Unit-Research Division
45-47 Pipitea Street
Thorndon, Wellington

NORWAY
Liv Sissel Grønmo
University of Oslo ILS
Postboks 1099 Blindern
0316 Oslo 3

PALESTINIAN NATIONAL AUTHORITY
Ola Khalili
Ministry of Education
Assessment Center
Box 719
Ramallah – West Bank

PHILIPPINES
Vivien Talisayon
University of the Philippines
TIMSS 2003 Coordination Office
Vidal Tan Hall
Diliman, Quezon City

ROMANIA
Gabriela Noveanu
Institute for Educational Sciences
Evaluation and Forecasting Division
Revoltiei B1. C7/7
Bucharest Ro-70732

RUSSIAN FEDERATION
Galina Kovalyova
Center for Evaluating the Quality of Education
Institute of General Secondary Education
Russian Academy of Education
Ul. Pogodinskaya, 8 - Moscow 119905

SAUDI ARABIA
Ali Alhakami
Ministry of Education
Center for Educational Development
PO Box 102298
Riyadh, 11675

SCOTLAND
Jo MacDonald
Scottish Executive Education Dept Research
Economic and Corporate Strategy Unit
1B (South) Victoria Quay
Edinburgh EH6 6QQ

SERBIA
Slobodanka Milanovic-Nahod
Institute for Educational Research
Dobrinjska 11/lll
PF 546
11001 Belgrade

SINGAPORE
Kok Leong Boey
Research and Evaluation Section
#15-31 MOE Building
1 North Buona Vista Drive
138675 Singapore

SLOVAK REPUBLIC
Jozef Kuraj
National Institute for Education
P O Box 26, Pluhova 8
Brastislava SK – 830 00

SLOVENIA
Barbara Japelj
Educational Research Institute
Gerbiceva 62
PO Box 76
Ljubljana 1000

SOUTH AFRICA
Vijay Reddy
Human Sciences Research Council (HSRC)
134 Pretorius Street
Private Bag X07
Pretoria 0001

SWEDEN
Jan-Olof Lindstrom
Department of Educational
Measurement/ TIMSS
Umeå University
Enheten for pedagogiska matningar
Samhallsvetarhuset, 90187 Umeå

SYRIA
Aychoua Ishak
Ministry of Education
Damascus

TUNISIA
Mejib Ayed
CNIPRE
130 Boulevard du 9 Avril 1938
Tunis 1006

USA
Patrick Gonzales
National Center for Education Statistics
US Department of Education
1990 K St., NW Rm 9071
Washington, DC 20006

YEMEN, REPUBLIC OF
Omar Ba-Fadhel
Center for Educational Research and Development
Ministry of Education
General Market (Nokom),
Sana'a 23049

Benchmarking Participants

Basque Country, SPAIN
Josu Sierra
ISEI-IVEI
Asturias 9
Bilbao, Basque Country 48015

State of Indiana, USA
Carole Gallagher
Division of Assessment
Indiana Department of Education
Room 229, State House
Indianapolis, IN 46204

Ontario Province, CANADA

Francine Jaques
Education Quality and Accountability Office
2 Carlton Street
Suite 1200
Toronto, Ontario M5B 2M9

Québec Province, CANADA

Serge Baillargeon
MEQ Direction de la Sanction des Études
1035 rue de la Chevrotière
13e Étage
Québec, PQ G1R 5 A5

B

Appendix B

Characteristics of National Samples

Introduction

For each country participating in TIMSS 2003, this appendix describes the target population definition (where necessary), the extent of coverage and exclusions, the use of stratification variables, and any deviations from the general TIMSS sample design.

B.1 Armenia

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and very small schools (less than six eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by region, for a total of 11 implicit strata
- Same schools sampled in Fourth Grade and Eighth Grade

Exhibit B.1.1 **Allocation of School Sample in Armenia – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Armenia	150	0	148	0	0	2
Total	150	0	148	0	0	2

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and very small schools (less than six eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by region, for a total of 11 implicit strata
- Same schools sampled in Fourth Grade and Eighth Grade

Exhibit B.1.2 **Allocation of School Sample in Armenia – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Armenia	150	0	149	0	0	1
Total	**150**	**0**	**149**	**0**	**0**	**1**

B.2 Australia

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools, hospital schools, schools with radically different curricula, remote schools in the Northern Territory, and very small schools (less than five eligible students)

Sample Design

- Explicit stratification by States and Territories, for a total of eight explicit strata
- Implicit stratification by school type (Government, Catholic, Independent), for a total of 24 implicit strata

Exhibit B.2.1 **Allocation of School Sample in Australia – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
New South Wales	40	0	29	2	4	5
Victoria	35	0	26	6	0	3
Queensland	35	2	29	1	1	2
South Australia	30	0	25	1	1	3
Western Australia	30	1	20	6	1	2
Tasmania	30	0	25	0	0	5
Northern Territory	15	0	11	2	0	2
Australian Capital Territory	15	0	13	1	0	1
Total	**230**	**3**	**178**	**19**	**7**	**23**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools, hospital schools, schools with radically different curricula, remote schools in the Northern Territory, and very small schools (less than five eligible students)

Sample Design

- Explicit stratification by States and Territories, for a total of eight explicit strata
- Implicit stratification by school type (Government, Catholic, Independent), for a total of 24 implicit strata
- Schools were sampled with equal probabilities in the “Tasmania”, “Northern Territory”, and “Australian Capital Territory” strata

Exhibit B.2.2 **Allocation of School Sample in Australia – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
New South Wales	40	0	27	4	1	8
Victoria	35	0	31	2	1	1
Queensland	35	1	29	1	3	1
South Australia	30	0	25	2	0	3
Western Australia	30	1	23	2	1	3
Tasmania	30	1	25	1	0	3
Northern Territory	15	1	13	1	0	0
Australian Capital Territory	15	0	13	1	1	0
Total	**230**	**4**	**186**	**14**	**7**	19

B.3 Bahrain

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- There were no reported school-level exclusions

Sample Design

- No explicit stratification
- Implicit stratification by school type (girl schools, boy schools, private schools), for a total of three implicit strata
- All schools in the sample

Exhibit B.3.1 **Allocation of School Sample in Bahrain – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Bahrain	67	0	67	0	0	0
Total	**67**	**0**	**67**	**0**	**0**	**0**

B.4 Basque Country, Spain

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of other language schools and very small schools (less than ten eligible students)

Sample Design

- Explicit stratification by school type (public, private) and language (Basque, Castilian, mixed), for a total of six explicit strata
- No implicit stratification
- Small schools were sampled with probabilities proportional to size
- Four schools were sampled with certainty in the "Public – Type A (Castilian)" stratum

Exhibit B.4.1 **Allocation of School Sample in Basque Country, Spain – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Private – Type A (Castilian)	20	0	20	0	0	0
Private – Type B (Mixed)	20	0	20	0	0	0
Private – Type D (Basque)	20	0	20	0	0	0
Public – Type A (Castilian)	20	0	19	1	0	0
Public – Type B (Mixed)	20	0	20	0	0	0
Public – Type D (Basque)	20	0	20	0	0	0
Total	**120**	**0**	**119**	**1**	**0**	**0**

B.5 Belgium (Flemish)

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and very small schools (less than five eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by school type (catholic, communal, state), for a total of three implicit strata

Exhibit B.5.1 **Allocation of School Sample in Belgium (Flemish) – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Belgium (Flemish)	150	0	133	12	4	1
Total	**150**	**0**	**133**	**12**	**4**	**1**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and very small schools (less than five eligible students)

Sample Design

- Explicit stratification by school program (academic, professional) and school size (very large, large) in the “Academic” stratum, for a total of three explicit strata
- Implicit stratification by school type (catholic, communal, state), for a total of seven implicit strata
- Schools sampled with equal probabilities in the “Academic – Very Large Schools” stratum

Exhibit B.5.2 **Allocation of School Sample in Belgium (Flemish) – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Academic – Large Schools	114	0	96	13	4	1
Professional	30	0	21	6	2	1
Total	**150**	**0**	**122**	**20**	**6**	**2**

B.6 Botswana

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools

Sample Design

- Explicit stratification by school type (government, private), for a total of two explicit strata
- Implicit stratification by region (five regions) and urbanization (rural, semi-urban, urban), for a total of 21 implicit strata
- Schools sampled with equal probabilitie

Exhibit B.6.1 **Allocation of School Sample in Botswana – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Government	145	0	142	0	0	3
Private	5	0	4	0	0	1
Other	2	2	0	0	0	0
Total	**152**	**2**	**146**	**0**	**0**	**4**

B.7 Bulgaria

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of very small schools (less than five eligible students)

Sample Design

- Explicit stratification by school size (very large, large), for a total of two explicit strata
- Implicit stratification within large schools by entrance examination (with, without), for a total of three implicit strata
- The one "Very Large School" was in fact a cluster of smaller schools. One of them was sampled with PPS

Exhibit B.7.1 **Allocation of School Sample in Bulgaria – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Very Large Schools	1	0	1	0	0	0
Large Schools	169	1	162	1	0	5
Total	**170**	**1**	**163**	**1**	**0**	**5**

B.8 Chile

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of remote schools, schools on Easter Island, and very small schools (less than 11 eligible students)

Sample Design

- Explicit stratification by region (North & Region 8, all other regions) and school type (municipal, subsidized, private), for a total of six explicit strata

- Implicit stratification by urbanization (rural, urban), for a total of 12 implicit strata

Exhibit B.8.1 **Allocation of School Sample in Chile – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
North & Region 8 – Municipal	45	0	43	2	0	0
North & Region 8 – Subsidized	34	0	33	1	0	0
North & Region 8 – Private	31	0	31	0	0	0
All Other Regions – Municipal	50	0	50	0	0	0
All Other Regions - Subsidized	21	0	21	0	0	0
All Other Regions - Private	14	0	13	1	0	0
Total	**195**	**0**	**191**	**4**	**0**	**0**

B.9 Chinese Taipei

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and very small schools (less than eight eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by region (five regions), for a total of five implicit strata
- Small schools were sampled with probabilities proportional to size

Exhibit B.9.1 **Allocation of School Sample in Chinese Taipei – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Chinese Taipei	150	0	150	0	0	0
Total	**150**	**0**	**150**	**0**	**0**	**0**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and very small schools (less than eight eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by region (five regions) and gender (girls, boys, mixed), for a total of ten implicit strata
- Small schools were sampled with probabilities proportional to size

Exhibit B.9.2 **Allocation of School Sample in Chinese Taipei – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Chinese Taipei	150	0	150	0	0	0
Total	**150**	**0**	**150**	**0**	**0**	**0**

B.10 Cyprus

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of isolated schools and very small schools (less than seven eligible students)

Sample Design

- Explicit stratification by district, for a total of four explicit strata
- Implicit stratification by urbanization (rural, urban), for a total of eight implicit strata
- Schools sampled with equal probabilities

Exhibit B.10.1 **Allocation of School Sample in Cyprus – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Nicosia	54	0	54	0	0	0
Larnaka	38	0	38	0	0	0
Limassol	42	0	42	0	0	0
Pafos	16	0	16	0	0	0
Total	**150**	**0**	**150**	**0**	**0**	**0**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of isolated schools and very small schools (less than 15 eligible students)

Sample Design

- Explicit stratification by district, for a total of four explicit strata
- Implicit stratification by urbanization (rural, urban), for a total of eight implicit strata
- All schools in the sample

Exhibit B.10.2 **Allocation of School Sample in Cyprus – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Nicosia	22	0	22	0	0	0
Larnaka	13	0	13	0	0	0
Limassol	16	0	16	0	0	0
Pafos	8	0	8	0	0	0
Total	**59**	**0**	**59**	**0**	**0**	**0**

B.11 Egypt

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of schools for the blind, handicraft schools, sport schools, and very small schools (less than 12 eligible students)

Sample Design

- Explicit stratification by school type, for a total of six explicit strata
- Implicit stratification by gender (boys, girls, mixed), urbanization (rural, urban), school type (public, free private) in the "Afternoon 2nd Shift" stratum, schedule (full time, morning shift, noon shift) in the "Public" stratum, for a total of 42 implicit strata

Exhibit B.11.1 **Allocation of School Sample in Egypt – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Afternoon 2nd Shift	25	0	25	0	0	0
Public Schools	115	0	115	1	0	0
Experimental Language Schools	25	0	25	0	0	0
Free Private Schools	2	0	2	0	0	0
Private Schools	25	0	25	0	0	0
Private Language Schools	25	0	25	1	0	0
Total	**217**	**0**	**217**	**2**	**0**	**0**

B.12 England

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and very small schools (less than eight eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by school performance (six levels) and school type (primary, junior, middle, independent), for a total of 24 implicit strata

Exhibit B.12.1 **Allocation of School Sample in England – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
England	150	0	79	31	13	27
Total	**150**	**0**	**79**	**31**	**13**	**27**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and very small schools (less than seven eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by school performance (six levels) and school type (comprehensive to 16, comprehensive to 18, independent, grammar, other), for a total of 27 implicit strata
- Small schools were sampled with probabilities proportional to size

Exhibit B.12.2 **Allocation of School Sample in England – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
England	160	0	62	22	3	73
Total	**160**	**0**	**62**	**22**	**3**	**73**

B.13 Estonia

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and very small schools (less than seven eligible students)

Sample Design

- Explicit stratification by language (Estonian, Russian) and school size (very large, large), for a total of four explicit strata
- Implicit stratification by urbanization (five levels) and school type (years 1-12, years 1-9), for a total of 26 implicit strata
- All schools sampled in the two “Very Large Schools” strata

Exhibit B.13.1 **Allocation of School Sample in Estonia – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Very Large Estonian Schools	16	0	16	0	0	0
Large Estonian Schools	94	1	92	0	0	1
Very Large Russian Schools	7	0	7	0	0	0
Large Russian Schools	37	1	36	0	0	0
Total	**154**	**2**	**151**	**0**	**0**	**1**

B.14 Ghana

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and very small schools (less than 11 eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by region, for a total of ten implicit strata

Exhibit B.14.1 **Allocation of School Sample in Ghana – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Ghana	150	0	150	0	0	0
Total	**150**	**0**	**150**	**0**	**0**	**0**

B.15 Hong Kong, SAR

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools, international schools, and very small schools (less than nine eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by gender (single-sex, mixed), school type (aided, government & private), and schedule (morning, afternoon, whole day), for a total of 12 implicit strata

Exhibit B.15.1 **Allocation of School Sample in Hong Kong, SAR – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Hong Kong, SAR	150	0	116	14	2	18
Total	**150**	**0**	**116**	**14**	**2**	**18**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and international schools

Sample Design

- No explicit stratification
- Implicit stratification by gender (single-sex, mixed), school type (aided, government & private), and language (Chinese, English), for a total of eight implicit strata

Exhibit B.15.2 **Allocation of School Sample in Hong Kong, SAR – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Hong Kong, SAR	150	0	112	12	1	25
Total	**150**	**0**	**112**	**12**	**1**	**25**

B.16 Hungary

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and very small schools (less than 15 eligible students)

Sample Design

- Explicit stratification by grade (Fourth Grade only, Fourth Grade and Eighth Grade), for a total of two explicit strata
- Implicit stratification by province (20 provinces) and urbanization (village, town, county seat, Budapest), for a total of 109 implicit strata
- Small schools were sampled with probabilities proportional to size
- Maximum school sample overlap between Fourth Grade and Eighth Grade

Exhibit B.16.1 **Allocation of School Sample in Hungary – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Fourth Grade Only	14	0	14	0	0	0
Fourth Grade & Eighth Grade	146	1	142	1	0	2
Total	**160**	**1**	**156**	**1**	**0**	**2**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and very small schools (less than 15 eligible students)

Sample Design

- Explicit stratification by grade (Eighth Grade only, Fourth Grade and Eighth Grade), for a total of two explicit strata
- Implicit stratification by province (20 provinces) and urbanization (village, town, county seat, Budapest), for a total of 113 implicit strata
- Small schools were sampled with probabilities proportional to size
- Maximum school sample overlap between Fourth Grade and Eighth Grade

Exhibit B.16.2 **Allocation of School Sample in Hungary – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Eighth Grade Only	22	2	20	0	0	0
Fourth Grade & Eighth Grade	138	1	134	1	0	2
Total	**160**	**3**	**154**	**1**	**0**	**2**

B.17 Indiana State, U.S.

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- There were no reported school-level exclusions

Sample Design

- Explicit stratification by school type (public, private), for a total of two explicit strata
- Implicit stratification urbanization (eight levels) and minority status (above 15%, below 15%), for a total of 32 implicit strata
- Small schools were sampled with probabilities proportional to size

Exhibit B.17.1 **Allocation of School Sample in Indiana State, U.S. – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Public	50	0	50	0	0	0
Private	6	0	6	0	0	0
Total	**56**	**0**	**56**	**0**	**0**	**0**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- There were no reported school-level exclusions

Sample Design

- Explicit stratification by school type (public, private), for a total of two explicit strata
- Implicit stratification urbanization (eight levels) and minority status (above 15%, below 15%), for a total of 31 implicit strata
- Small schools were sampled with probabilities proportional to size

Exhibit B.17.2 **Allocation of School Sample in Indiana State, U.S. – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Public	50	0	49	0	0	1
Private	6	0	5	0	0	1
Total	**56**	**0**	**54**	**0**	**0**	**2**

B.18 Indonesia

EIGHTH GRADE

Coverage and Exclusions

- Coverage in Indonesia was restricted to students in non-Islamic schools (80% of International Desired Target Grade)
- School-level exclusions consisted of very small schools (less than ten eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by school type (public, private) and performance (high, average, low), for a total of six implicit strata

Exhibit B.18.1 **Allocation of School Sample in Indonesia – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Indonesia	150	0	148	2	0	0
Total	**150**	**0**	**148**	**2**	**0**	**0**

B.19 Iran, Islamic Republic of

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of remote schools and very small schools (less than seven eligible students)

Sample Design

- Explicit stratification by school size (small, large) and school type (public, private), for a total of four explicit strata
- No implicit stratification
- Small schools were sampled with probabilities proportional to size

Exhibit B.19.1 **Allocation of School Sample in Iran, Islamic Republic of – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Small – Public	24	1	23	0	0	0
Small – Private	8	0	8	0	0	0
Large – Public	108	4	104	0	0	0
Large – Private	36	0	36	0	0	0
Total	**176**	**5**	**171**	**0**	**0**	**0**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of adult schools and remote schools

Sample Design

- Explicit stratification by school size (small, large) and school type (public, private), for a total of four explicit strata
- No implicit stratification
- Small schools were sampled with probabilities proportional to size

Exhibit B.19.2 **Allocation of School Sample in Iran, Islamic Republic of – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Small – Public	20	0	20	0	0	0
Small – Private	5	1	4	0	0	0
Large – Public	148	5	143	0	0	0
Large – Private	15	1	14	0	0	0
Total	**188**	**7**	**181**	**0**	**0**	**0**

B.20 Israel

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools, Ultra Orthodox schools, Arab schools (East Jerusalem), and very small schools (less than nine eligible students)

Sample Design

- Explicit stratification by ethnicity (Hebrew secular, Hebrew religious, Arab), for a total of three explicit strata
- Implicit stratification by school type (five types) and socio-economic status (four levels), for a total of 40 implicit strata
- Small schools were sampled with probabilities proportional to size

Exhibit B.20.1 **Allocation of School Sample in Israel – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Hebrew Secular	70	1	68	0	1	0
Hebrew Religious	40	1	37	2	0	0
Arab	40	1	38	0	0	1
Total	**150**	**3**	**143**	**2**	**1**	**1**

B.21 Italy

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of very small schools (less than eight eligible students)

Sample Design

- No explicit stratification

- Implicit stratification by province (20 provinces) and urbanization (capital town, other towns), for a total of 40 implicit strata
- Small schools were sampled with probabilities proportional to size

Exhibit B.21.1 **Allocation of School Sample in Italy – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Italy	172	1	165	6	0	0
Total	**172**	**1**	**165**	**6**	**0**	**0**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of very small schools (less than eight eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by province (20 provinces) and urbanization (capital town, other towns), for a total of 40 implicit strata
- Small schools were sampled with probabilities proportional to size

Exhibit B.21.2 **Allocation of School Sample in Italy – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Italy	172	1	164	6	1	0
Total	**172**	**1**	**164**	**6**	**1**	**0**

B.22 Japan

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of schools for educable mentally disabled students and schools for functionally disabled students

Sample Design

- Explicit stratification by urbanization (big city area, city area, non-city area), for a total of three explicit strata
- No implicit stratification

Exhibit B.22.1 **Allocation of School Sample in Japan – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Big City	27	0	27	0	0	0
City	84	0	84	0	0	0
Non-City	39	0	39	0	0	0
Total	**150**	**0**	**150**	**0**	**0**	**0**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of schools for educable mentally disabled students and schools for functionally disabled students

Sample Design

- Explicit stratification by school type (public, private or national), urbanization (big city area, city area, non-city area) in the "Public" stratum, for a total of four explicit strata
- No implicit stratification

Exhibit B.22.2 **Allocation of School Sample in Japan – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Public – Big City	24	0	24	0	0	0
Public – City	79	0	79	0	0	0
Public – Non-City	37	0	36	0	0	1
Private Or National	10	0	7	0	0	3
Total	**150**	**0**	**146**	**0**	**0**	**4**

B.23 Jordan

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of very small schools (less than nine eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by school type (public, private, UNRWA), urbanization (rural, urban) in the "Public" and "Private" strata, and gender (boys, girls, mixed), for a total of 15 implicit strata

Exhibit B.23.1 **Allocation of School Sample in Jordan – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Jordan	150	10	140	0	0	0
Total	**150**	**10**	**140**	**0**	**0**	**0**

B.24 Korea, Republic of

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of remote schools, special education schools, sports schools, and very small schools (less than 11 eligible students)

Sample Design

- Explicit stratification by province (16 provinces), for a total of 16 explicit strata
- Implicit stratification by urbanization (large city, middle, rural) and gender (boys, girls, mixed), for a total of 83 implicit strata
- Small schools were sampled with probabilities proportional to size

Exhibit B.24.1 **Allocation of School Sample in Korea, Republic of – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Seoul	30	0	30	0	0	0
Pusan	12	0	12	0	0	0
Taegu	9	0	9	0	0	0
Inchon	9	0	9	0	0	0
Kwangju	5	0	5	0	0	0
Taejon	5	1	4	0	0	0
Ulsan	4	0	4	0	0	0
Kyunggi-do	30	0	30	0	0	0
Kangwon-do	4	0	4	0	0	0
Chungchongbuk-do	5	0	5	0	0	0
Chungchongnam-do	6	0	6	0	0	0
Chollabuk-do	6	0	6	0	0	0
Chollanam-do	6	0	6	0	0	0
Kyongsangbuk-do	8	0	7	0	0	1
Kongsangnam-do	10	0	10	0	0	0
Cheju-do	2	0	2	0	0	0
Total	**151**	**1**	**149**	**0**	**0**	**1**

B.25 Latvia

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools, other language schools, and very small schools (less than six eligible students in both Fourth Grade and Eighth Grade)

Sample Design

- Explicit stratification by grade (Fourth Grade only, Fourth Grade and Eighth Grade) and school size (very large, large) in the “Fourth Grade and Eighth Grade” stratum, for a total of three explicit strata
- Implicit stratification by language (Latvian, Russian, mixed) and urbanization (rural, urban), for a total of 15 implicit strata

- Schools sampled with equal probabilities in the "Fourth Grade and Eighth Grade – Very Large" stratum
- Maximum school sample overlap between Fourth Grade and Eighth Grade

Exhibit B.25.1 **Allocation of School Sample in Latvia – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Fourth Grade Only	15	0	15	0	0	0
Fourth & Eighth Grade – Very Large	27	1	25	0	0	1
Fourth & Eighth Grade – Large	108	0	97	1	2	8
Total	**150**	**1**	**137**	**1**	**2**	**9**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools, other language schools, and very small schools (less than six eligible students in both Fourth Grade and Eighth Grade)

Sample Design

- Explicit stratification by grade (Eighth Grade only, Fourth Grade and Eighth Grade) and school size (very large, large) in the "Fourth Grade and Eighth Grade" stratum, for a total of three explicit strata
- Implicit stratification by language (Latvian, Russian, mixed) and urbanization (rural, urban), for a total of 12 implicit strata
- Schools sampled with equal probabilities in the "Eighth Grade Only" and "Fourth Grade & Eighth Grade – Very Large" strata
- Maximum school sample overlap between Fourth Grade and Eighth Grade

Exhibit B.25.2 **Allocation of School Sample in Latvia – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Eighth Grade Only	15	1	14	0	0	0
Fourth & Eighth Grade – Very Large	27	0	26	0	0	1
Fourth & Eighth Grade – Large	108	0	97	1	2	8
Total	**150**	**1**	**137**	**1**	**2**	**9**

B.26 Lebanon

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of very small schools (less than nine eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by school type (public, private), urbanization (rural, urban), and gender (boys, girls, mixed), for a total of ten implicit strata

Exhibit B.26.1 **Allocation of School Sample in Lebanon – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Lebanon	160	0	148	4	0	8
Total	**160**	**0**	**148**	**4**	**0**	**8**

B.27 Lithuania

FOURTH GRADE

Coverage and Exclusions

- Coverage in Lithuania was restricted to students whose language of instruction is Lithuanian (92% of International Desired Target Grade).
- School-level exclusions consisted of special education schools and very small schools (less than five eligible students)

Sample Design

- Explicit stratification by grade (Fourth Grade only, Fourth Grade and Eighth Grade), for a total of two explicit strata
- Implicit stratification by school type (basic, secondary, primary), for a total of five implicit strata
- Maximum school sample overlap between Fourth Grade and Eighth Grade

Exhibit B.27.1 **Allocation of School Sample in Lithuania – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Fourth Grade Only	42	0	37	3	1	1
Fourth Grade & Eighth Grade	118	0	110	2	0	6
Total	**160**	**0**	**147**	**5**	**1**	**7**

EIGHTH GRADE

Coverage and Exclusions

- Coverage in Lithuania was restricted to students whose language of instruction is Lithuanian (89% of International Desired Target Grade).
- School-level exclusions consisted of special education schools and very small schools (less than six eligible students)

Sample Design

- Explicit stratification by grade (Eighth Grade only, Fourth Grade and Eighth Grade), for a total of two explicit strata

- Implicit stratification by school type (basic, secondary), for a total of four implicit strata
- Maximum school sample overlap between Fourth Grade and Eighth Grade

Exhibit B.27.2 **Allocation of School Sample in Lithuania – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Eighth Grade Only	23	0	20	2	1	0
Fourth Grade & Eighth Grade	127	0	117	3	0	7
Total	**150**	**0**	**137**	**5**	**1**	**7**

B.28 Macedonia, Republic of

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools, other language schools (Turkish and Serbian), schools in politically sensitive regions (near the border with Kosovo), and very small schools (less than seven eligible students)

Sample Design

- Explicit stratification by school size (large, very large), for a total of two explicit strata
- Implicit stratification by language (Macedonian, Albanian) and urbanization (rural, urban), for a total of seven implicit strata
- All schools sampled in the "Very Large Schools" stratum

Exhibit B.28.1 **Allocation of School Sample in the Macedonia, Republic of – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Very Large Schools	28	0	28	0	0	0
Large Schools	122	0	114	7	0	1
Total	**150**	**0**	**142**	**7**	**0**	**1**

B.29 Malaysia

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of private schools, international schools, and special education schools

Sample Design

- No explicit stratification
- Implicit stratification by state (14 states) and urbanization (rural, urban), for a total of 28 implicit strata
- Small schools were sampled with probabilities proportional to size

Exhibit B.29.1 **Allocation of School Sample in Malaysia – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Malaysia	150	0	150	0	0	0
Total	**150**	**0**	**150**	**0**	**0**	**0**

B.30 Moldova

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and very small schools (less than six eligible students in both Fourth Grade and Eighth Grade)

Sample Design

- Explicit stratification by grade (Fourth Grade only, Fourth Grade and Eighth Grade), for a total of two explicit strata
- Implicit stratification by urbanization (rural, urban), school type (Gymnasium, Lyceum, General School, other) in the “Fourth Grade and

Eighth Grade" stratum, and language (National, Russian, mixed) in the "Fourth Grade and Eighth Grade" stratum, for a total of 18 implicit strata

- Maximum school sample overlap between Fourth Grade and Eighth Grade

Exhibit B.30.1 **Allocation of School Sample in Moldova – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Fourth Grade Only	14	2	12	0	0	0
Fourth & Eighth Grade	139	0	135	4	0	0
Total	**153**	**2**	**147**	**4**	**0**	**0**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and very small schools (less than six eligible students in both Fourth Grade and Eighth Grade)

Sample Design

- Explicit stratification by grade (Eighth Grade only, Fourth Grade and Eighth Grade), for a total of two explicit strata
- Implicit stratification by urbanization (rural, urban), school type (Gymnasium, Lyceum, General School, other) in the "Fourth Grade and Eighth Grade" stratum, and language (National, Russian, mixed) in the "Fourth Grade & Eighth Grade" stratum, for a total of 18 implicit strata
- Maximum school sample overlap between Fourth Grade and Eighth Grade

Exhibit B.30.2 **Allocation of School Sample in Moldova – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Eighth Grade Only	11	1	10	0	0	0
Fourth Grade & Eighth Grade	139	0	137	2	0	0
Total	**150**	**1**	**147**	**2**	**0**	**0**

B.31 Morocco

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and very small schools (less than six eligible students in both Fourth Grade and Eighth Grade)

Sample Design

- Explicit stratification by grade (Fourth Grade only, Fourth Grade and Eighth Grade) and region strata (eight strata) in the “Fourth Grade Only” stratum, for a total of nine explicit strata
- The 16 regions of Morocco were combined into eight region strata
- Implicit stratification by school type (public, private), urbanization (rural, urban) in the “Public” stratum, and administration (four types) in the “Fourth Grade Only – Public” stratum, for a total of 66 implicit strata
- Maximum school sample overlap between Fourth Grade and Eighth Grade

Exhibit B.31.1 **Allocation of School Sample in Morocco – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Fourth Grade Only – Region Stratum 1	25	0	24	0	0	1
Fourth Grade Only – Region Stratum 2	25	0	8	0	0	17
Fourth Grade Only – Region Stratum 3	35	0	34	0	0	1
Fourth Grade Only – Region Stratum 4	25	0	24	0	0	1
Fourth Grade Only – Region Stratum 5	30	0	29	0	0	1
Fourth Grade Only – Region Stratum 6	30	0	28	0	0	2
Fourth Grade Only – Region Stratum 7	25	0	23	0	0	2
Fourth Grade Only – Region Stratum 8	30	0	27	0	0	3
Fourth & Eighth Grade	2	2	0	0	0	0
Total	**227**	**2**	**197**	**0**	**0**	**28**

EIGHTH GRADE

Coverage and Exclusions

- Coverage in Morocco was restricted to students outside the regions of Souss Massa Draa, Casablanca and Gharb-Chrardais (69% of the International Desired Target Grade).
- School-level exclusions consisted of special education schools and very small schools (less than six eligible students in both Fourth Grade and Eighth Grade)

Sample Design

- Explicit stratification by grade (Eighth Grade only, Fourth Grade and Eighth Grade), region strata (eight strata) in the "Fourth Grade Only" stratum, and school size (large, very large) in the "Region Stratum 1" stratum, for a total of ten explicit strata
- The 16 regions of Morocco were combined into eight region strata
- Implicit stratification by school type (public, private) and urbanization (rural, urban) in the "Public" stratum, for a total of 27 implicit strata

- All schools sampled in the "Eighth Grade Only – Region 1 – Very Large" stratum
- Maximum school sample overlap between Fourth Grade and Eighth Grade

Exhibit B.31.2 **Allocation of School Sample in Morocco – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Eighth Grade Only - Reg.1 - Very Large	4	0	4	0	0	0
Eighth Grade Only - Reg.1 - Large	21	0	16	0	0	5
Eighth Grade Only – Region Stratum 2	25	25	0	0	0	0
Eighth Grade Only – Region Stratum 3	35	35	0	0	0	0
Eighth Grade Only – Region Stratum 4	25	0	23	0	0	2
Eighth Grade Only – Region Stratum 5	30	0	26	0	0	4
Eighth Grade Only – Region Stratum 6	30	0	23	0	0	7
Eighth Grade Only – Region Stratum 7	25	0	20	0	0	5
Eighth Grade Only – Region Stratum 8	30	0	19	0	0	11
Fourth & Eighth Grade	2	2	0	0	0	0
Total	**227**	**62**	**131**	**0**	**0**	**34**

B.32 Netherlands

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and very small schools (less than seven eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by mean National Student Weight (low, medium, high), for a total of three implicit strata

Exhibit B.32.1 **Allocation of School Sample in the Netherlands – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Netherlands	150	1	77	36	17	19
Total	**150**	**1**	**77**	**36**	**17**	**19**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special secondary education, schools with recovery program (“vrije scholen”), and very small schools (less than seven eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by school program (VMBO, HAVO / VWO, mixed), for a total of three implicit strata
- Minimum school sample overlap between TIMSS and PISA

Exhibit B.32.2 **Allocation of School Sample in the Netherlands – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Netherlands	150	0	118	12	0	20
Total	**150**	**0**	**118**	**12**	**0**	**20**

B.33 New Zealand

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools, correspondence schools, Rudolf Steiner schools, and very small schools (less than four eligible students)

Sample Design

- Explicit stratification by language of instruction (Maori, English), for a total of two explicit strata
- Implicit stratification by school type (state, private) in the "Non-Maori" stratum, school decile indicator (low, medium, high) in the "Non-Maori - State" stratum, and urbanization (rural, urban) in the "Non-Maori - State" stratum, for a total of eight implicit strata

Exhibit B.33.1 **Allocation of School Sample in New Zealand – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Maori language instruction	10	0	5	1	0	4
English language instruction	218	0	189	23	2	4
Total	**228**	**0**	**194**	**24**	**2**	**8**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools, correspondence schools, Maori immersion schools, Rudolf Steiner schools, and very small schools (less than seven eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by school type (state, private), school decile indicator (low, medium, high) in the "State" stratum, urbanization

(rural, urban) in the "State" stratum, and gender (boys, girls, mixed) in the "State" stratum, for a total of ten implicit strata

- Schools were sampled with equal probabilities

Exhibit B.33.2 **Allocation of School Sample in New Zealand – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
New Zealand	175	1	149	20	0	5
Total	**175**	**1**	**149**	**20**	**0**	**5**

B.34 Norway

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools, Sami schools, and very small schools (less than five eligible students in both Fourth Grade and Eighth Grade)

Sample Design

- Explicit stratification by grade (Fourth Grade only, Fourth Grade and Eighth Grade) and language (Bokmål, other), for a total of four explicit strata
- No implicit stratification
- Small schools were sampled with probabilities proportional to size in the "Fourth Grade & Eighth Grade – Bokmål" stratum
- Maximum school sample overlap between Fourth Grade and Eighth Grade

Exhibit B.34.1 **Allocation of School Sample in Norway – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Fourth Grade Only – Bokmål	46	0	42	2	0	2
Fourth Grade Only – Other	79	0	68	3	0	8
Fourth & Eighth Grade – Other	20	0	20	0	0	0
Total	**150**	**0**	**134**	**5**	**0**	**11**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools, Sami schools, and very small schools (less than five eligible students in both Fourth Grade and Eighth Grade)

Sample Design

- Explicit stratification by grade (Eighth Grade only, Fourth Grade and Eighth Grade) and language (Bokmål, other), for a total of four explicit strata
- No implicit stratification
- Small schools were sampled with probabilities proportional to size in the “Eighth Grade Only – Bokmål”, “Fourth Grade & Eighth Grade – Bokmål”, and “Eighth Grade Only – Other” strata
- Maximum school sample overlap between Fourth Grade and Eighth Grade

Exhibit B.34.2 **Allocation of School Sample in Norway – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Eighth Grade Only – Bokmål	42	0	40	0	0	2
Fourth & Eighth Grade – Bokmål	8	0	6	0	0	2
Eighth Grade Only – Other	73	0	68	0	0	5
Fourth & Eighth Grade – Other	27	0	24	0	0	3
Total	**150**	**0**	**138**	**0**	**0**	**12**

B.35 Ontario Province, Canada

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools, remote schools (northern regions), and very small schools (less than ten eligible students in both Fourth Grade and Eighth Grade)

Sample Design

- Explicit stratification by grade (Fourth Grade only, Fourth Grade and Eighth Grade) and language (English, French), for a total of four explicit strata
- Implicit stratification by school type (public, private, separate), for a total of 12 implicit strata
- Maximum school sample overlap between Fourth Grade and Eighth Grade

Exhibit B.35.1 **Allocation of School Sample in Ontario Province, Canada – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Fourth Grade Only – English	32	2	27	1	1	1
Fourth Grade Only – French	25	0	24	1	0	0
Fourth & Eighth Grade – English	88	2	75	4	1	6
Fourth & Eighth Grade – French	55	0	53	2	0	0
Total	**200**	**4**	**179**	**8**	**2**	**7**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools, native schools, overseas schools, and very small schools (less than ten eligible students in both Fourth Grade and Eighth Grade)

Sample Design

- Explicit stratification by grade (Eighth Grade only, Fourth Grade and Eighth Grade) and language (English, French), for a total of four explicit strata
- Implicit stratification by school type (public, private, separate), for a total of 11 implicit strata
- All schools sampled in the “Eighth Grade Only – French” stratum
- Maximum school sample overlap between Fourth Grade and Eighth Grade

Exhibit B.35.2 **Allocation of School Sample in Ontario Province, Canada – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Eighth Grade Only – English	32	1	22	4	3	2
Eighth Grade Only – French	25	1	23	0	0	1
Fourth & Eighth Grade – English	88	1	75	5	1	6
Fourth & Eighth Grade – French	55	1	51	2	0	1
Total	**200**	**4**	**171**	**11**	**4**	**10**

B.36 Palestinian Nat'l Authority

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of very small schools (less than 11 eligible students)

Sample Design

- Explicit stratification by school size (very large, large), for a total of two explicit strata
- Implicit stratification by regions (Gaza Strip, West Bank), school type (public, private, UNWRA), and gender (boys, girls, mixed), for a total of 20 implicit strata
- All schools sampled in the “Very Large Schools” stratum

Exhibit B.36.1 **Allocation of School Sample in Palestinian Nat'l Authority – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Very Large Schools	22	0	22	0	0	0
Large Schools	128	5	123	0	0	0
Total	**150**	**5**	**145**	**0**	**0**	**0**

B.37 Philippines

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of schools in the ARMM region and very small schools (less than ten eligible students)

Sample Design

- Explicit stratification by school type (public, private), for a total of two explicit strata
- Implicit stratification by region (16 regions), for a total of 32 implicit strata

Exhibit B.37.1 **Allocation of School Sample in the Philippines – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Public	149	0	111	9	4	25
Private	11	0	11	0	0	0
Total	**160**	**0**	**122**	**9**	**4**	**25**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of schools in the ARMM Region and very small schools (less than ten eligible students)

Sample Design

- Explicit stratification by school type (public, private), for a total of two explicit strata
- Implicit stratification by region (16 regions), for a total of 32 implicit strata

Exhibit B.37.2 **Allocation of School Sample in the Philippines – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Public	126	0	100	5	0	21
Private	34	0	32	0	0	2
Total	**160**	**0**	**132**	**5**	**0**	**23**

B.38 Québec Province, Canada

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools, remote schools (northern regions), and very small schools (less than 11 eligible students in both Fourth Grade and Eighth Grade)

Sample Design

- Explicit stratification by grade (Fourth Grade only, Fourth Grade and Eighth Grade) and language (English, French, English & French), for a total of five explicit strata
- Implicit stratification by school type (public, private), for a total of nine implicit strata
- Maximum school sample overlap between Fourth Grade and Eighth Grade

Exhibit B.38.1 **Allocation of School Sample in Québec Province, Canada – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Fourth Grade Only – English	65	1	63	0	0	1
Fourth Grade Only – English & French	2	0	2	0	0	0
Fourth Grade Only – French	112	0	111	1	0	0
Fourth & Eighth Grade – English	13	1	12	0	0	0
Fourth & Eighth Grade – French	6	2	4	0	0	0
Total	**198**	**4**	**192**	**1**	**0**	**1**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools, remote schools (northern regions), and very small schools (less than 11 eligible students in both Fourth Grade and Eighth Grade)

Sample Design

- Explicit stratification by grade (Eighth Grade only, Fourth Grade and Eighth Grade) and language (English, French, English & French), for a total of five explicit strata
- Implicit stratification by school type (public, private), for a total of nine implicit strata
- All schools sampled in the “Eighth Grade Only – English” and “Eighth Grade Only – English & French” strata
- Maximum school sample overlap between Fourth Grade and Eighth Grade

Exhibit B.38.2 **Allocation of School Sample in Québec Province, Canada – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Eighth Grade Only – English	66	8	58	0	0	0
Eighth Grade Only – English & French	1	0	1	0	0	0
Eighth Grade Only – French	113	5	98	2	0	8
Fourth & Eighth Grade – English	13	0	12	0	0	1
Fourth & Eighth Grade – French	6	1	4	0	0	1
Total	**199**	**14**	**173**	**2**	**0**	**10**

B.39 Romania

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and very small schools (less than seven eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by region (42 regions) and urbanization (rural, urban), for a total of 83 implicit strata

Exhibit B.39.1 **Allocation of School Sample in Romania – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Romania	150	1	148	0	0	1
Total	**150**	**1**	**148**	**0**	**0**	**1**

B.40 Russian Federation

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of evening schools, special needs schools, atypical schools, and very small schools (less than four eligible students)

Sample Design

- Preliminary sampling of 45 regions from a frame of 89 regions, 17 regions large enough to be sampled with certainty
- No explicit stratification (the explicit strata in table B.40.1 correspond to the primary sampling units)
- Implicit stratification by town size (ten levels), for a total of 225 implicit strata
- Generally, four schools sampled per region, more schools sampled in some certainty regions
- Large schools were sampled with equal probabilities in the regions “Rasan Obl”, “Kirov Obl”, and “Omsk Obl”

Exhibit B.40.1 **Allocation of School Sample in the Russian Federation – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Sankt-Petersburg*	6	0	6	0	0	0
Archangelsk Obl	4	0	4	0	0	0
Komi	4	0	4	0	0	0
Karelia	4	0	4	0	0	0
Moscow*	10	0	10	0	0	0
Moscow Obl*	8	0	8	0	0	0
Voroneg Obl	4	0	4	0	0	0
Tula Obl	4	0	4	0	0	0
Brjansk Obl	4	0	4	0	0	0
Yaroslav Obl	4	0	4	0	0	0
Tambov Obl	4	0	4	0	0	0
Rasan Obl	4	0	4	0	0	0
Kaluga Obl	4	0	4	0	0	0
Bashkortostan*	8	0	8	0	0	0
Tatarstan*	6	0	6	0	0	0
N_Novgorod Obl*	4	0	4	0	0	0
Samara Obl*	4	0	4	0	0	0
Perm Obl*	4	0	4	0	0	0
Saratov Obl	4	1	3	0	0	0
Orenburg Obl	4	0	4	0	0	0
Udmurtia	4	0	4	0	0	0
Kirov Obl	4	0	4	0	0	0
Pensa Obl	4	0	4	0	0	0
Marii_Al	4	0	4	0	0	0
Krasnodar Kr*	6	0	6	0	0	0
Rostov Obl*	6	0	6	0	0	0
Dagestan*	6	0	6	0	0	0
Stavropol Kr*	4	0	4	0	0	0
Volgograd Obl	4	0	4	0	0	0
Alania	4	0	4	0	0	0
Sverdlovsk Obl*	6	0	5	1	0	0
Chelyabinsk Obl*	4	0	4	0	0	0
Hanty_Mansii Ok	4	0	4	0	0	0
Tumen Obl	4	0	4	0	0	0
Krasnoyarsk Obl*	4	0	4	0	0	0

Exhibit B.40.1 **Allocation of School Sample in the Russian Federation – Fourth Grade** *(...Continued)*

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Kemerovo Obl*	4	0	4	0	0	0
Irkutsk Obl*	4	0	4	0	0	0
Altay Kr	4	0	4	0	0	0
Novosibirsk Obl	4	0	4	0	0	0
Omsk Obl	4	0	4	0	0	0
Chita Obl	4	0	4	0	0	0
Tyva	4	0	4	0	0	0
Primorsk Kr	4	0	4	0	0	0
Saha	4	0	4	0	0	0
Magadan Obl	4	0	4	0	0	0
Total	**206**	**1**	**204**	**1**	**0**	**0**

Strata marked with (*) were selected with certainty

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of evening schools, special needs schools, atypical schools, and very small schools (less than five eligible students)

Sample Design

- Preliminary sampling of 45 regions from a frame of 89 regions, 19 regions large enough to be sampled with certainty
- No explicit stratification (the explicit strata in table B.40.2 correspond to the primary sampling units)
- Implicit stratification by town size (ten levels), for a total of 230 implicit strata
- Generally, four schools sampled per region, more schools sampled in some certainty regions
- Large schools were sampled with equal probabilities in the regions “Rasan Obl”, “Kirov Obl”, “Omsk Obl”, and “Tomsk Obl”

Exhibit B.40.2 **Allocation of School Sample in the Russian Federation – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Sankt-Petersburg*	6	0	6	0	0	0
Leningrad Obl	4	0	4	0	0	0
Vologda Obl	4	0	4	0	0	0
Murmansk Obl	4	0	4	0	0	0
Novgorod Obl	4	0	4	0	0	0
Moscow*	12	0	12	0	0	0
Moscow Obl*	10	0	10	0	0	0
Vladimir Obl	4	0	4	0	0	0
Tver Obl	4	0	4	0	0	0
Rasan Obl	4	0	3	0	0	1
Smolensk Obl	4	0	4	0	0	0
Orel Obl	4	0	4	0	0	0
N_Novgorod Obl*	6	0	6	0	0	0
Kirov Obl	4	0	4	0	0	0
Marii_Al	4	0	4	0	0	0
Belgorod Obl	4	0	4	0	0	0
Tambov Obl	4	0	4	0	0	0
Samara Obl*	6	0	6	0	0	0
Saratov Obl*	4	0	4	0	0	0
Volgograd Obl*	4	0	4	0	0	0
Ulianovsk Obl	4	0	4	0	0	0
Tatarstan	4	0	4	0	0	0
Kalmykia	4	0	4	0	0	0
Krasnodar Kr*	8	0	8	0	0	0
Rostov Obl*	6	0	6	0	0	0
Stavropol Kr*	4	0	4	0	0	0
Kabarda_Balkaria	4	0	4	0	0	0
Sverdlovsk Obl*	8	0	8	0	0	0
Bashkortostan*	6	0	6	0	0	0
Chelyabinsk Obl*	6	0	6	0	0	0
Perm Obl*	4	0	3	0	0	1
Orenburg Obl	4	0	4	0	0	0
Udmurtia	4	0	4	0	0	0
Kemerovo Obl*	6	0	6	0	0	0
Novosibirsk Obl*	4	0	4	0	0	0

Exhibit B.40.2 **Allocation of School Sample in the Russian Federation – Eighth Grade**
(...Continued)

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Altay Kr*	4	0	4	0	0	0
Omsk Obl	4	0	4	0	0	0
Hanty_Mansii Ok	4	0	4	0	0	0
Tomsk Obl	4	0	4	0	0	0
Krasnoyarsk Obl*	4	0	4	0	0	0
Irkutsk Obl*	4	0	4	0	0	0
Chita Obl	4	0	4	0	0	0
Primorsk Kr	4	0	4	0	0	0
Habarovsk Kr	4	0	4	0	0	0
Sahalin Obl	4	0	4	0	0	0
Total	**216**	**0**	**214**	**0**	**0**	**2**

Strata marked with (*) were selected with certainty

B.41 Saudi Arabia

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of very small schools (less than seven eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by gender (boys, girls) and school type (government, private), for a total of four implicit strata

Exhibit B.41.1 **Allocation of School Sample in Saudi Arabia – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Saudi Arabia	160	0	154	1	0	5
Total	**160**	**0**	**154**	**1**	**0**	**5**

B.42 Scotland

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted very small schools (less than seven eligible students)

Sample Design

- Explicit stratification by grade (Fourth Grade, Fourth Grade & Eighth Grade), for a total of two explicit strata
- Implicit stratification by school performance (six levels) and school type (five types), for a total of 18 implicit strata
- Maximum school sample overlap between Fourth Grade and Eighth Grade

Exhibit B.42.1 **Allocation of School Sample in Scotland – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Fourth Grade Only	146	0	90	25	6	25
Fourth Grade & Eighth Grade	4	0	4	0	0	0
Total	**150**	**0**	**94**	**25**	**6**	**25**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted very small schools (less than seven eligible students)

Sample Design

- Explicit stratification by grade (Eighth Grade only, Fourth Grade and Eighth Grade) and school size (large, other) in the "Eighth Grade Only" stratum, for a total of three explicit strata
- Implicit stratification by school performance (six levels) and school type (four types), for a total of 28 implicit strata
- Maximum school sample overlap between Fourth Grade and Eighth Grade
- Minimum school sample overlap between TIMSS and PISA

Exhibit B.42.2 **Allocation of School Sample in Scotland – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Eighth Grade Only – Large	60	0	48	5	0	7
Eighth Grade Only – Other	82	0	60	7	0	15
Fourth & Eighth Grade	8	0	7	1	0	0
Total	**150**	**0**	**115**	**13**	**0**	**22**

B.43 Serbia

EIGHTH GRADE

Coverage and Exclusions

- Coverage in Serbia was restricted to students outside Kosovo (81% of International Desired Target Grade).
- School-level exclusions consisted of schools near Kosovo, special education schools, and very small schools (less than ten eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by region (Central Serbia, Belgrade, Vojvodina) and urbanization (rural, urban), for a total of six implicit strata

Exhibit B.43.1 **Allocation of School Sample in Serbia – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Serbia	150	0	149	0	0	1
Total	**150**	**0**	**149**	**0**	**0**	**1**

B.44 Singapore

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- No school-level exclusions

Sample Design

- All schools in the sample

Exhibit B.44.1 **Allocation of School Sample in Singapore – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Singapore	182	0	182	0	0	0
Total	**182**	**0**	**182**	**0**	**0**	**0**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- No school-level exclusions

Sample Design

- All schools in the sample

Exhibit B.44.2 **Allocation of School Sample in Singapore – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Singapore	164	0	164	0	0	0
Total	**164**	**0**	**164**	**0**	**0**	**0**

B.45 Slovak Republic

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools

Sample Design

- Explicit stratification by school type (gymnasium, basic) and language (Slovak, Hungarian), for a total of four explicit strata
- Implicit stratification by regions (eight regions), for a total of 25 implicit strata
- The school measure of size was based on the number of classes in the schools

Exhibit B.45.1 **Allocation of School Sample in the Slovak Republic – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Gymnasium – Slovak	30	0	26	3	1	0
Gymnasium – Hungarian	10	1	9	0	0	0
Basic – Slovak	120	0	116	4	0	0
Basic – Hungarian	20	0	19	1	0	0
Total	**180**	**1**	**170**	**8**	**1**	**0**

B.46 Slovenia

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools, Italian schools, and very small schools (less than eight eligible students in both Fourth Grade and Eighth Grade)

Sample Design

- Explicit stratification by school structure (new system in both Fourth Grade & 2, new system in Fourth Grade, new system in Eighth Grade, old system) and school size (very large, large) in the "Old System" stratum, for a total of five explicit strata
- Implicit stratification by region (eight regions), for a total of 29 implicit strata
- All schools sampled in the "New System In Both Fourth Grade and 2", "New System In Fourth Grade", "New System In Eighth Grade", and "Old System – Very Large" strata
- Same schools sampled in Fourth Grade and Eighth Grade

Exhibit B.46.1 **Allocation of School Sample in Slovenia – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
New System In Both Fourth Grade & 2	16	0	16	0	0	0
New System in Fourth Grade	23	0	21	0	0	2
New System In Eighth Grade	15	0	15	0	0	0
Old System – Very Large	3	0	3	0	0	0
Old System - Large	120	0	114	5	0	1
Total	**177**	**0**	**169**	**5**	**0**	**3**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools, Italian schools, and very small schools (less than eight eligible students in both Fourth Grade and Eighth Grade)

Sample Design

- Explicit stratification by school structure (new system in both Fourth Grade & 2, new system in Fourth Grade, new system in Eighth Grade, old system) and school size (very large, large) in the “Old System” stratum, for a total of five explicit strata
- Implicit stratification by region (eight regions), for a total of 29 implicit strata
- All schools sampled in the “New System In Both Fourth Grade and 2”, “New System In Fourth Grade”, “New System In Eighth Grade”, and “Old System – Very Large” strata
- Same schools sampled in Fourth Grade and Eighth Grade

Exhibit B.46.2 **Allocation of School Sample in Slovenia – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
New System In Both Fourth Grade & 2	16	0	16	0	0	0
New System In Fourth Grade	23	0	21	0	0	2
New System In Eighth Grade	15	0	15	0	0	0
Old System – Very Large	3	0	3	0	0	0
Old System - Large	120	0	114	5	0	1
Total	**177**	**0**	**169**	**5**	**0**	**3**

B.47 South Africa

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of special education schools and very small schools (less than 12 eligible students)

Sample Design

- Explicit stratification by province, for a total of nine explicit strata
- Implicit stratification by language (English, Afrikaans, mixed), for a total of 19 implicit strata

Exhibit B.47.1 **Allocation of School Sample in South Africa – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Eastern Cape	33	0	29	3	1	0
Free State	25	0	24	1	0	0
Gauteng	27	0	20	3	0	4
Kwazulu Natal	48	0	43	2	1	2
Mpumalanga	25	0	23	1	0	1
North West	25	0	25	0	0	0
Northern Cape	25	0	24	1	0	0
Northern Province	32	0	31	0	0	1
Western Cape	25	0	22	1	0	2
Total	**265**	**0**	**241**	**12**	**2**	**10**

B.48 Sweden

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of very small schools (less than seven eligible students)

Sample Design

- No explicit stratification
- No implicit stratification

Exhibit B.48.1 **Allocation of School Sample in Sweden – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Sweden	160	0	155	4	0	1
Total	**160**	**0**	**155**	**4**	**0**	**1**

B.49 Syrian, Arab Republic

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of small classes (small schools) because of changes in the school system

Sample Design

- Explicit stratification by urbanization (rural, urban), for a total of two explicit strata
- Implicit stratification by school type (public, private, UNRWA) and gender (girls, boys, mixed), for a total of 16 implicit strata

Exhibit B.49.1 **Allocation of School Sample in Syrian, Arab Republic – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Rural	73	0	61	7	1	4
Urban	77	0	60	4	1	12
Total	**150**	**0**	**121**	**11**	**2**	**16**

B.50 Tunisia

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of private schools, special education schools, and very small schools (less than eight eligible students)

Sample Design

- No explicit stratification
- Implicit stratification by school type (communal, non-communal) and governates (24 provinces), for a total of 46 implicit strata

Exhibit B.50.1 **Allocation of School Sample in Tunisia – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Tunisia	150	0	150	0	0	0
Total	**150**	**0**	**150**	**0**	**0**	**0**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of private schools and special education schools

Sample Design

- No explicit stratification
- Implicit stratification by performance (high, low, unknown) and governates (24 provinces), for a total of 63 implicit strata

Exhibit B.50.2 **Allocation of School Sample in Tunisia – Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Tunisia	150	0	150	0	0	0
Total	**150**	**0**	**150**	**0**	**0**	**0**

B.51 United States

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- There were no reported school-level exclusions

Sample Design

- Explicit stratification by poverty (high, low), for a total of two explicit strata

- Implicit stratification by school type (public, private), region (four regions), urbanization (eight levels), and minority status (above 15%, below 15%), for a total of 192 implicit strata
- Small schools were sampled with probabilities proportional to size

Exhibit B.51.1 **Allocation of School Sample in the United States– Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
High Poverty	120	3	85	1	16	15
Low Poverty	190	7	127	3	16	37
Total	**310**	**10**	**212**	**4**	**32**	**52**

EIGHTH GRADE

Coverage and Exclusions

- Coverage is 100%
- There were no reported school-level exclusions

Sample Design

- No explicit stratification
- Implicit stratification by school type (public, private), region (four regions), urbanization (eight levels), and minority status (above 15%, below 15%), for a total of 128 implicit strata
- Small schools were sampled with probabilities proportional to size

Exhibit B.51.2 **Allocation of School Sample in the United States– Eighth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
United States	301	5	211	3	18	64
Total	**301**	**5**	**211**	**3**	**18**	**64**

B.52 Yemen

FOURTH GRADE

Coverage and Exclusions

- Coverage is 100%
- School level exclusions consisted of very small schools (less than 13 eligible students)

Sample Design

- Explicit stratification by urbanization (rural, urban), for a total of two explicit strata
- Implicit stratification by school type (public, private, national) and gender (girls, boys, mixed), for a total of 13 implicit strata

Exhibit B.52.1 **Allocation of School Sample in Yemen – Fourth Grade**

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Participating Schools			Non-Participating Schools
			Sampled	1st Replacement	2nd Replacement	
Rural	103	0	103	0	0	0
Urban	47	0	47	0	0	0
Total	**150**	**0**	**150**	**0**	**0**	**0**

C

Appendix C

Country Adaptations to Items and Item Scoring

C.1 Fourth Grade

C.1.1 Items to be deleted

ALL COUNTRIES

M11_04, M14_04 Mathematics (faulty distracters)

S02_08, S08_06 Science (faulty distracters)

ARMENIA

S01_03, S01_04 Science (negative discrimination)

CYPRUS

S01_04 Science (poor discrimination)

HUNGARY

M04_11 Mathematics (not administered)

M07_08 Mathematics (negative discrimination)

M14_06 Mathematics (printing error)

S10_02 Science (printing error)

IRAN

S05_06 Science (printing error)

LITHUANIA

M12_06C Mathematics (printing error)

MOLDOVA

S08_08, S10_02 Science (negative discrimination)

MOLDOVA *(Russian only)*

M09_04 Mathematics (negative discrimination)

S03_02, S13_06, S14_04 Science (negative discrimination)

MOROCCO

M11_10 Mathematics (poor discrimination)

S04_01 Science (printing error)

NETHERLANDS

M04_06 Mathematics (scorer reliability less than 70%)

S13_03 Science (scorer reliability less than 70%)

SLOVENIA

M03_06 Mathematics (Slovenia administered a different item than the International version)

TUNISIA

S05_06 Science (printing error)

S07_10 Science (poor discrimination)

YEMEN

M03_08 Mathematics (printing error)

S07_10 Science (translation error)

C.1.2. Items needing options changed

MOLDOVA *(Russian only)*

M08_05 Mathematics (printing error, recode D to B and B to D)

C.1.3 Constructed-response items needing category recoding

ALL COUNTRIES

M11_12 Mathematics (recode 20 to 10, 10 to 71)

S08_08 Science (recode 20 to 10, 10 to 73, 11 to 74, 12 to 75)

C.2 GRADE 8

C.2.1 Items to be deleted

BAHRAIN

S05_08 Science (negative discrimination)

BOTSWANA

S02_01, S05_01, S11_09 Science (negative discrimination)

CANADA – ONTARIO AND QUEBEC *(French only)*

M05_07 Mathematics (printing error)

EGYPT

M01_12 Mathematics (Booklets 1 and Booklet 12 for Arabic version only)

M04_12 Mathematics (all Booklets for French version only)

S07_10, S09_13 Science (negative discrimination)

S08_10A, S08_10B Science (Booklets 2 for Arabic version only)

S14_01 Science (all Booklets for Arabic version only)

GHANA

M01_01 Mathematics (data moved from M01_01 to M01_02 in Booklet 6 and Booklet 12)

M01_02 Mathematics (Booklet 6 and Booklet 12)

M01_05, M01_12, M01_14, M02_06, M02_07, M02_12, M02_14, M02_15, M03_13, M04_01, M05_09, M06_13, M07_02, M07_10, M08_02, M08_05, M08_07, M08_09, M09_05, M10_02, M10_03, M10_04, M10_05, M11_02, M11_05, M11_06, M11_10, M11_12, M12_04, M12_05, M12_09, M12_10, M13_02, M13_09, M13_10, M13_11, M14_01, M14_03, M14_09 Mathematics (printing error)

M10_07, M10_08, M10_09 Mathematics (printing error in Booklet 6 only)

M12_03 Mathematics (not administered)

S01_03, S09_03 Science (printing error)

S14_01 Science (negative discrimination)

HUNGARY

M07_01 Mathematics (printing error)

INDONESIA
M02_04 Mathematics (negative discrimination)

S11_09 Science (negative discrimination)

ITALY
M06_10 Mathematics (printing error with fractions)

JORDAN
M08_06, M09_01 Mathematics (translation error)

M14_05 Mathematics (poor discrimination)

M14_08 Mathematics (negative discrimination)

S05_08, S07_08, S09_05, S09_06 Science (translation error)

KOREA
S02_11 Science (printing error)

LATVIA
M04_11A Mathematics (printing error)

M11_03 Mathematics (scorer reliability less than 70%)

M14_01 Mathematics (poor discrimination)

LITHUANIA
S14_09 Science (scorer reliability less than 70%)

LEBANON
M01_06 Mathematics (negative discrimination)

MACEDONIA
M14_05 Mathematics (negative discrimination)

S01_10 Science (negative discrimination)

S05_02 Science (translation error)

MOLDOVA
M06_10 (negative discrimination)

MOROCCO
M03_01, M03_05 Mathematics (printing error)

S01_08, S04_10 Science (negative discrimination)

S01_12 Science (translation error)

NETHERLANDS

S11_05 Science (negative discrimination)

PALESTINIAN NATIONAL AUTHORITY

M14_08 Mathematics (negative discrimination)

S01_06, S01_16, S05_08, S09_05, S09_06 Science (negative discrimination)

S07_08 Science (translation error)

ROMANIA

M12_11 Mathematics (printing error)

SAUDI ARABIA

M13_06 Mathematics (negative discrimination)

M13_07 Mathematics (poor discrimination)

S01_08, S05_08 Science (negative discrimination)

SLOVAK REPUBLIC

M06_12 Mathematics (printing error in Booklets 2, 5, and 6)

M12_13A Mathematics (negative discrimination)

S11_01 Science (printing error)

SLOVENIA

S14_08B, S14_09 Science (scorer reliability less than 70%)

TUNISIA

M03_05, M03_10, M08_06 Mathematics (translation error)

S04_10, S13_01 Science (negative discrimination)

C 2.2. Constructed-response items needing category recoding

ALL COUNTRIES

M07_05, M13_04 Mathematics (recode 20 to 10, 10 to 70)

S07_11, S08_10B, S09_03 Science (recode 20 to 10, 10 to 12, 29 to 19)

S10_06 Science (recode 20 to 10, 21 to 11, 22 to 12, 29 to 19, 10 to 71, 11 to 72, 19 to 79)

S13_10 Science (recode 20 to 10, 29 to 19, 10 to 70, 70 to 71, 71 to 72, 19 to 79)

D

Appendix D

Item Parameters for IRT Analyses of TIMSS 2003 Data

Exhibit D.1 **IRT Parameters for TIMSS Joint 1999-2003 Eighth-Grade Mathematics**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M012001	1.466	0.025	0.030	0.011	0.154	0.005				
M012002	0.585	0.013	-1.304	0.053	0.000	0.021				
M012003	0.911	0.016	-0.197	0.017	0.073	0.007				
M012004	1.139	0.030	0.643	0.017	0.319	0.006				
M012005	0.662	0.016	-0.315	0.038	0.171	0.013				
M012006	0.583	0.017	-0.998	0.073	0.248	0.023				
M012007	0.837	0.031	0.020	0.040	0.234	0.014				
M012008	0.493	0.021	-0.945	0.113	0.137	0.035				
M012009	0.893	0.037	0.332	0.038	0.324	0.012				
M012010	1.335	0.028	0.191	0.012	0.027	0.004				
M012011	1.016	0.028	-0.156	0.025	0.123	0.011				
M012012	1.412	0.037	-0.385	0.019	0.152	0.001				
M012013	1.231	0.035	0.136	0.020	0.193	0.009				
M012014	0.864	0.027	-0.777	0.045	0.205	0.019				
M012015	0.868	0.021	-0.545	0.028	0.058	0.012				
M012016	1.369	0.064	0.891	0.024	0.409	0.007				
M012017	0.733	0.024	0.077	0.036	0.127	0.013				
M012018	0.651	0.031	-0.302	0.077	0.241	0.025				
M012019	0.717	0.027	-0.385	0.052	0.121	0.019				
M012020	1.197	0.051	-0.053	0.037	0.389	0.013				
M012021	1.311	0.039	-0.235	0.022	0.134	0.011				
M012022	0.634	0.044	0.953	0.058	0.282	0.016				
M012023	0.623	0.027	-1.485	0.116	0.186	0.042				
M012024	0.873	0.040	-0.045	0.052	0.338	0.017				
M012025	0.706	0.021	-0.611	0.047	0.106	0.018				
M012026	1.108	0.032	0.260	0.021	0.167	0.009				
M012027	1.235	0.038	0.231	0.021	0.234	0.009				
M012028	1.013	0.027	-0.498	0.029	0.154	0.013				
M012029	1.017	0.028	0.040	0.023	0.137	0.001				
M012030	1.275	0.036	0.491	0.016	0.135	0.006				
M012031	1.370	0.052	0.907	0.019	0.140	0.006				
M012032	0.453	0.027	-0.000	0.114	0.149	0.031				
M012033	1.093	0.043	0.089	0.034	0.287	0.013				
M012034	0.644	0.029	0.419	0.047	0.123	0.016				
M012035	1.563	0.051	0.389	0.017	0.158	0.007				
M012036	0.817	0.041	0.712	0.039	0.243	0.013				
M012037	0.562	0.030	0.600	0.060	0.226	0.017				
M012038	0.983	0.035	-0.299	0.040	0.355	0.014				
M012039	1.006	0.029	0.055	0.025	0.167	0.010				
M012040	1.044	0.031	-0.416	0.032	0.236	0.014				
M012041	1.046	0.027	-0.112	0.023	0.122	0.001				
M012042	1.116	0.029	0.037	0.021	0.132	0.009				
M012043	0.892	0.039	0.057	0.046	0.295	0.016				
M012044	1.169	0.036	-0.405	0.029	0.160	0.014				
M012045	0.829	0.028	-1.268	0.060	0.131	0.027				
M012046	1.433	0.046	0.298	0.019	0.156	0.008				
M012047	1.160	0.044	0.074	0.031	0.278	0.012				
M012048	0.925	0.032	-0.617	0.045	0.190	0.019				
M022002	1.627	0.082	1.121	0.021	0.142	0.006				
M022004	1.440	0.067	0.462	0.026	0.289	0.010				
M022005	1.060	0.069	1.079	0.036	0.275	0.010				
M022008	0.531	0.017	0.756	0.034	0.000	0.000				

Exhibit D.1 **IRT Parameters for TIMSS Joint 1999-2003 Eighth-Grade Mathematics**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M022010	0.790	0.031	-0.596	0.053	0.118	0.022				
M022012	0.518	0.015	-0.567	0.029	0.000	0.000				
M022016	0.986	0.059	1.078	0.034	0.204	0.010				
M022021	1.579	0.059	0.387	0.019	0.141	0.008				
M022022	1.536	0.096	0.632	0.031	0.183	0.012				
M022026	0.860	0.031	0.054	0.027	0.000	0.000				
M022030	1.053	0.037	-0.701	0.028	0.000	0.000				
M022031	1.148	0.074	0.612	0.039	0.164	0.015				
M022033	0.468	0.040	-0.407	0.182	0.159	0.050				
M022037	0.819	0.055	0.001	0.071	0.185	0.026				
M022038	0.706	0.053	-0.034	0.091	0.183	0.031				
M022041	0.568	0.058	0.047	0.160	0.305	0.042				
M022042	2.013	0.102	0.325	0.022	0.110	0.010				
M022043	0.615	0.023	-0.839	0.068	0.086	0.025				
M022046	0.705	0.018	-0.732	0.024	0.000	0.000				
M022049	0.558	0.034	-0.121	0.104	0.288	0.028				
M022050	0.841	0.035	0.831	0.028	0.102	0.009				
M022055	1.186	0.024	0.370	0.013	0.000	0.000				
M022057	0.405	0.027	-0.538	0.184	0.184	0.045				
M022062	0.978	0.035	0.552	0.024	0.010	0.009				
M022066	1.330	0.034	-0.234	0.017	0.053	0.008				
M022070	0.726	0.039	-1.079	0.090	0.097	0.035				
M022073	0.660	0.048	-0.400	0.117	0.196	0.039				
M022078	0.832	0.064	0.456	0.068	0.231	0.023				
M022079	0.677	0.037	-0.740	0.085	0.087	0.031				
M022083	1.229	0.085	0.967	0.036	0.145	0.012				
M022085	1.038	0.064	0.587	0.041	0.130	0.015				
M022089	0.920	0.032	0.093	0.026	0.000	0.000				
M022093	1.299	0.092	0.486	0.045	0.299	0.017				
M022097	0.979	0.029	-0.382	0.029	0.086	0.012				
M022101	0.721	0.029	-0.549	0.062	0.183	0.023				
M022104	0.833	0.026	-0.674	0.040	0.086	0.017				
M022105	0.555	0.027	0.540	0.052	0.086	0.017				
M022106	0.878	0.019	0.551	0.017	0.000	0.000				
M022108	0.791	0.031	-0.112	0.046	0.182	0.017				
M022110	0.407	0.014	0.018	0.034	0.000	0.000				
M022113	0.758	0.044	-0.813	0.092	0.137	0.036				
M022116	0.965	0.063	0.655	0.046	0.141	0.016				
M022118	1.013	0.034	-0.195	0.025	0.000	0.000				
M022121	1.503	0.094	0.123	0.041	0.290	0.018				
M022124	0.604	0.038	-0.338	0.097	0.101	0.032				
M022126	1.341	0.127	1.155	0.045	0.311	0.013				
M022127	1.330	0.076	1.157	0.027	0.182	0.008				
M022128	1.036	0.082	1.056	0.045	0.173	0.014				
M022132	1.209	0.040	-0.001	0.022	0.000	0.000				
M022135	0.701	0.030	0.702	0.033	0.044	0.010				
M022139	1.236	0.059	0.945	0.025	0.157	0.008				
M022142	1.244	0.051	0.370	0.025	0.185	0.010				
M022144	0.660	0.043	0.653	0.060	0.222	0.019				
M022146	1.570	0.057	0.026	0.021	0.183	0.010				
M022148	0.967	0.022	-0.131	0.017	0.000	0.000				

Exhibit D.1 **IRT Parameters for TIMSS Joint 1999-2003 Eighth-Grade Mathematics**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M022154	0.830	0.039	0.224	0.044	0.182	0.016				
M022156	1.195	0.027	0.079	0.014	0.000	0.000				
M022159	1.506	0.113	1.244	0.034	0.111	0.008				
M022160	1.557	0.117	0.939	0.034	0.211	0.011				
M022165	1.111	0.075	0.112	0.055	0.279	0.021				
M022166	1.140	0.066	0.230	0.041	0.153	0.017				
M022168	0.620	0.052	0.734	0.077	0.125	0.024				
M022169	0.871	0.057	-0.295	0.076	0.218	0.030				
M022172	0.999	0.066	0.155	0.056	0.219	0.022				
M022173	1.458	0.100	0.562	0.037	0.251	0.014				
M022176	0.673	0.059	-0.486	0.153	0.344	0.045				
M022178	1.045	0.037	0.307	0.025	0.000	0.000				
M022181	1.064	0.039	-0.838	0.046	0.278	0.020				
M022185	0.858	0.041	0.050	0.048	0.233	0.018				
M022188	0.755	0.048	0.830	0.049	0.241	0.015				
M022189	0.822	0.030	-0.768	0.051	0.104	0.021				
M022191	0.801	0.036	-0.272	0.056	0.202	0.021				
M022194	0.819	0.038	0.181	0.044	0.173	0.016				
M022196	1.217	0.038	-0.357	0.024	0.082	0.011				
M022198	1.163	0.057	0.652	0.029	0.234	0.010				
M022199	1.352	0.059	0.510	0.025	0.211	0.001				
M022202	0.713	0.019	0.660	0.025	0.000	0.000				
M022204	0.486	0.036	-1.574	0.225	0.178	0.067				
M022206	1.707	0.144	0.737	0.039	0.378	0.013				
M022207	1.017	0.046	0.080	0.040	0.198	0.016				
M022208	0.680	0.050	0.118	0.085	0.156	0.029				
M022210	1.276	0.071	0.538	0.032	0.101	0.012				
M022213	0.803	0.051	0.335	0.055	0.109	0.020				
M022219	1.005	0.036	0.620	0.028	0.000	0.000				
M022222	0.994	0.034	-0.034	0.025	0.000	0.000				
M022227A	1.038	0.024	-0.444	0.017	0.000	0.000				
M022227B	1.362	0.032	0.402	0.014	0.000	0.000				
M022227C	1.256	0.032	0.803	0.017	0.000	0.000				
M022228	0.641	0.011	0.411	0.016	0.000	0.000	-1.738	0.053	1.738	0.055
M022231A	1.309	0.034	0.775	0.017	0.000	0.000				
M022231B	1.542	0.049	1.329	0.021	0.000	0.000				
M022232	0.521	0.008	1.498	0.020	0.000	0.000	-2.520	0.060	2.520	0.065
M022234A	0.742	0.001	0.612	0.011	0.000	0.000	-0.632	0.023	0.632	0.025
M022234B	0.815	0.011	0.925	0.011	0.000	0.000	-1.549	0.038	1.549	0.040
M022237	0.941	0.023	0.270	0.019	0.000	0.000				
M022241	1.160	0.046	0.380	0.026	0.118	0.011				
M022243	1.080	0.019	0.306	0.012	0.000	0.000				
M022244	1.176	0.029	0.049	0.016	0.000	0.000				
M022245	0.661	0.040	-1.141	0.118	0.132	0.044				
M022246	0.680	0.047	0.057	0.081	0.133	0.028				
M022249	1.136	0.073	0.566	0.042	0.185	0.016				
M022251	0.826	0.053	1.272	0.040	0.158	0.010				
M022252	1.163	0.051	-0.020	0.036	0.285	0.015				
M022253	1.118	0.025	-0.192	0.015	0.000	0.000				
M022256	0.666	0.014	0.548	0.017	0.000	0.000	0.019	0.028	-0.019	0.032
M022257	1.334	0.050	0.366	0.022	0.226	0.009				

Exhibit D.1 **IRT Parameters for TIMSS Joint 1999-2003 Eighth-Grade Mathematics**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M022258	0.717	0.051	0.023	0.085	0.168	0.030				
M022260	0.996	0.060	-1.049	0.083	0.203	0.038				
M022261A	1.105	0.025	0.222	0.015	0.000	0.000				
M022261B	1.175	0.029	0.717	0.017	0.000	0.000				
M022261C	0.706	0.013	1.001	0.017	0.000	0.000	-2.048	0.064	2.048	0.067
M022262A	0.961	0.024	-0.616	0.021	0.000	0.000				
M022262B	0.955	0.024	-0.218	0.019	0.000	0.000				
M022262C	0.614	0.011	0.553	0.017	0.000	0.000	-1.662	0.054	1.662	0.056
M032036	1.081	0.084	0.141	0.057	0.190	0.024				
M032044	1.104	0.068	0.531	0.041	0.231	0.015				
M032046	1.204	0.077	1.071	0.034	0.140	0.001				
M032047	1.337	0.182	1.104	0.067	0.441	0.017				
M032064	1.229	0.053	0.504	0.028	0.000	0.000				
M032079	0.997	0.071	1.005	0.043	0.184	0.013				
M032094	1.196	0.092	-0.016	0.062	0.270	0.026				
M032097	1.187	0.127	1.244	0.054	0.185	0.015				
M032100	0.730	0.050	0.002	0.072	0.082	0.025				
M032116	0.842	0.080	0.607	0.074	0.202	0.025				
M032132	0.542	0.053	0.216	0.131	0.138	0.039				
M032142	1.204	0.138	0.857	0.066	0.366	0.019				
M032160	1.760	0.161	1.127	0.036	0.152	0.011				
M032163	1.304	0.101	0.415	0.047	0.211	0.019				
M032166	0.957	0.076	-0.181	0.078	0.216	0.031				
M032198	0.786	0.060	0.199	0.073	0.121	0.027				
M032205	0.519	0.047	-0.278	0.158	0.139	0.047				
M032208	1.317	0.076	0.377	0.036	0.255	0.014				
M032210	1.453	0.084	0.643	0.030	0.212	0.012				
M032228	1.175	0.061	0.136	0.038	0.191	0.016				
M032233	0.948	0.033	1.000	0.028	0.000	0.000	-0.586	0.052	0.586	0.062
M032261	0.871	0.054	0.454	0.049	0.173	0.018				
M032271	1.352	0.077	0.360	0.034	0.246	0.014				
M032273	1.095	0.087	-0.232	0.076	0.296	0.030				
M032294	0.793	0.056	-0.430	0.088	0.132	0.033				
M032295	1.220	0.086	-0.694	0.070	0.237	0.034				
M032307	1.339	0.034	0.761	0.017	0.000	0.000				
M032324	1.062	0.081	0.546	0.049	0.138	0.018				
M032331	1.823	0.186	1.209	0.039	0.196	0.011				
M032344	1.140	0.048	0.362	0.028	0.000	0.000				
M032352	1.248	0.120	0.416	0.064	0.376	0.022				
M032381	0.904	0.039	0.104	0.033	0.000	0.000				
M032397	1.149	0.099	0.660	0.052	0.211	0.019				
M032398	1.557	0.145	0.806	0.044	0.280	0.016				
M032402	0.686	0.074	0.654	0.098	0.206	0.031				
M032403	0.936	0.029	-0.318	0.023	0.000	0.000				
M032414	1.106	0.048	0.472	0.030	0.000	0.000				
M032416	1.086	0.070	0.612	0.039	0.058	0.012				
M032419	1.211	0.113	0.717	0.054	0.267	0.019				
M032424	0.939	0.071	0.258	0.062	0.154	0.024				
M032447	1.237	0.097	0.391	0.047	0.189	0.019				
M032477	1.188	0.087	0.290	0.050	0.180	0.020				
M032489	0.858	0.051	-0.697	0.086	0.262	0.033				

Exhibit D.1 **IRT Parameters for TIMSS Joint 1999-2003 Eighth-Grade Mathematics**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M032507	1.685	0.148	0.989	0.037	0.173	0.012				
M032523	1.747	0.088	0.950	0.021	0.161	0.007				
M032525	0.902	0.046	-0.097	0.051	0.139	0.021				
M032529	1.414	0.106	0.754	0.037	0.132	0.013				
M032533	1.315	0.065	0.157	0.032	0.176	0.014				
M032538	1.135	0.047	0.063	0.028	0.000	0.000				
M032540	0.675	0.062	-0.209	0.131	0.224	0.042				
M032545	1.040	0.033	0.714	0.024	0.000	0.000				
M032557	1.102	0.035	0.787	0.024	0.000	0.000				
M032570	1.434	0.088	0.289	0.038	0.334	0.015				
M032575	2.152	0.166	0.463	0.031	0.236	0.015				
M032579	1.089	0.041	-0.335	0.034	0.129	0.016				
M032588	0.781	0.047	-0.174	0.074	0.200	0.027				
M032595	1.149	0.077	0.022	0.048	0.118	0.021				
M032609	0.799	0.054	-0.707	0.088	0.114	0.035				
M032612	0.887	0.054	0.655	0.043	0.133	0.015				
M032623	1.526	0.102	0.492	0.033	0.117	0.013				
M032626	0.719	0.054	0.126	0.073	0.086	0.026				
M032637A	0.919	0.039	-0.484	0.035	0.000	0.000				
M032637B	1.283	0.052	-0.253	0.026	0.000	0.000				
M032637C	1.204	0.050	0.226	0.027	0.000	0.000				
M032640	0.530	0.021	1.506	0.055	0.000	0.000	-0.861	0.078	0.861	0.101
M032643	1.133	0.064	0.539	0.036	0.170	0.014				
M032647	0.857	0.095	1.342	0.067	0.334	0.016				
M032649A	0.949	0.029	0.258	0.023	0.000	0.000				
M032649B	1.174	0.040	1.009	0.026	0.000	0.000				
M032652	1.240	0.040	0.797	0.022	0.000	0.000				
M032662	1.326	0.126	1.352	0.047	0.099	0.010				
M032670	0.750	0.051	-1.667	0.130	0.130	0.052				
M032671	0.779	0.025	-0.804	0.031	0.000	0.000				
M032673	1.380	0.103	0.336	0.042	0.184	0.018				
M032678	1.469	0.060	0.213	0.022	0.066	0.009				
M032679	0.972	0.075	0.046	0.069	0.199	0.027				
M032681A	0.598	0.031	-0.793	0.054	0.000	0.000				
M032681B	0.475	0.029	0.844	0.070	0.000	0.000				
M032681C	0.958	0.042	0.435	0.033	0.000	0.000				
M032683	0.560	0.018	0.628	0.032	0.000	0.000	-1.006	0.072	1.006	0.078
M032688	0.744	0.036	0.584	0.042	0.000	0.000				
M032689	0.662	0.069	1.012	0.084	0.296	0.023				
M032690	0.814	0.080	0.711	0.070	0.163	0.023				
M032691	0.783	0.021	0.102	0.021	0.000	0.000				
M032692	0.687	0.022	0.835	0.030	0.000	0.000	-1.287	0.077	1.287	0.085
M032693	0.674	0.024	0.505	0.032	0.000	0.000				
M032695	0.494	0.015	-0.519	0.034	0.000	0.000	-1.239	0.086	1.239	0.081
M032698	1.019	0.073	0.282	0.054	0.125	0.021				
M032699	0.639	0.048	-0.846	0.163	0.332	0.048				
M032701	0.897	0.046	-1.480	0.087	0.150	0.040				
M032704	0.993	0.053	-0.307	0.054	0.196	0.023				
M032721	0.704	0.108	1.288	0.109	0.260	0.026				
M032725	1.091	0.049	0.700	0.033	0.000	0.000				
M032727	1.412	0.111	0.379	0.043	0.217	0.018				

Exhibit D.1 **IRT Parameters for TIMSS Joint 1999-2003 Eighth-Grade Mathematics**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M032728	1.363	0.132	0.754	0.048	0.258	0.017				
M032732	0.881	0.081	0.086	0.087	0.260	0.032				
M032734	0.660	0.032	-0.742	0.048	0.000	0.000				
M032738	1.160	0.085	-0.378	0.069	0.253	0.030				
M032743	0.575	0.030	-0.288	0.047	0.000	0.000				
M032744	0.826	0.039	0.402	0.037	0.000	0.000				
M032745	0.499	0.025	2.207	0.104	0.000	0.000	-1.288	0.110	1.288	0.157
M032753A	1.066	0.035	0.648	0.021	0.000	0.000	-0.250	0.039	0.250	0.044
M032753B	1.089	0.039	0.820	0.023	0.000	0.000	-0.016	0.035	0.016	0.043
M032753C	0.851	0.038	0.342	0.035	0.000	0.000				
M032754	0.695	0.034	-0.803	0.047	0.000	0.000				
M032755	1.038	0.038	1.116	0.027	0.000	0.000	-0.237	0.043	0.237	0.054
M032756	0.685	0.033	0.140	0.040	0.000	0.000				
M032757	0.465	0.014	-0.402	0.033	0.000	0.000	-2.368	0.118	2.368	0.115
M032760A	0.772	0.023	0.554	0.025	0.000	0.000	-1.484	0.082	1.484	0.085
M032760B	1.232	0.059	0.918	0.035	0.000	0.000				
M032760C	1.463	0.077	1.157	0.036	0.000	0.000				
M032761	1.297	0.053	1.131	0.025	0.000	0.000	-0.134	0.037	0.134	0.048
M032762	0.365	0.009	1.097	0.037	0.000	0.000	-2.628	0.096	2.628	0.104
M032763	0.839	0.024	1.590	0.030	0.000	0.000	-0.694	0.047	0.694	0.061
M032764	0.842	0.025	1.460	0.028	0.000	0.000	-0.313	0.037	0.313	0.051
MC22046	0.820	0.038	-0.934	0.043	0.000	0.000				
MC22110	0.509	0.029	-1.382	0.077	0.000	0.000				
MC32525	1.036	0.069	-0.284	0.062	0.146	0.027				
MC32701	1.147	0.080	-1.537	0.089	0.163	0.046				
MC32704	1.159	0.082	-0.439	0.066	0.210	0.030				
MF12001	1.608	0.101	0.017	0.036	0.142	0.017				
MF12002	0.715	0.045	-0.760	0.088	0.083	0.032				
MF12003	1.062	0.058	-0.137	0.042	0.054	0.016				
MF12004	1.155	0.095	0.511	0.053	0.218	0.020				
MF12005	0.793	0.053	-0.163	0.070	0.090	0.026				
MF12006	0.757	0.055	-0.453	0.096	0.138	0.036				
MF12013	0.967	0.059	0.049	0.050	0.077	0.019				
MF12014	0.965	0.068	-0.522	0.081	0.190	0.034				
MF12015	0.988	0.055	-0.245	0.047	0.056	0.018				
MF12016	0.918	0.092	0.659	0.074	0.254	0.025				
MF12017	0.857	0.056	0.213	0.054	0.070	0.019				
MF12025	0.797	0.052	-0.381	0.075	0.094	0.029				
MF12026	1.044	0.080	0.454	0.052	0.145	0.020				
MF12027	1.383	0.098	0.254	0.043	0.182	0.019				
MF12028	1.156	0.075	-0.081	0.051	0.139	0.023				
MF12029	1.270	0.075	0.152	0.037	0.073	0.015				
MF12030	1.277	0.087	0.529	0.038	0.104	0.014				
MF12037	0.587	0.048	0.227	0.095	0.091	0.030				
MF12038	1.017	0.068	-0.419	0.069	0.163	0.030				
MF12039	1.069	0.072	0.165	0.050	0.123	0.020				
MF12040	1.272	0.081	-0.277	0.050	0.152	0.023				
MF12041	1.355	0.080	0.090	0.036	0.090	0.016				
MF12042	1.532	0.091	0.170	0.032	0.092	0.014				
MF22002	1.298	0.120	1.267	0.047	0.108	0.011				
MF22004	1.197	0.114	0.625	0.058	0.300	0.021				

Exhibit D.1 **IRT Parameters for TIMSS Joint 1999-2003 Eighth-Grade Mathematics**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
MF22005	0.897	0.119	1.276	0.082	0.271	0.021				
MF22008	0.688	0.037	1.028	0.057	0.000	0.000				
MF22010	0.897	0.063	-0.120	0.069	0.131	0.028				
MF22012	0.785	0.036	-0.053	0.036	0.000	0.000				
MF22016	0.747	0.073	1.031	0.070	0.107	0.020				
MF22021	1.762	0.134	0.643	0.033	0.177	0.014				
MF22043	0.677	0.044	-0.640	0.093	0.087	0.032				
MF22046	0.722	0.033	-0.621	0.043	0.000	0.000				
MF22049	0.542	0.051	-0.366	0.167	0.170	0.050				
MF22050	0.667	0.064	0.930	0.075	0.095	0.022				
MF22055	1.181	0.050	0.422	0.028	0.000	0.000				
MF22057	0.510	0.049	-0.428	0.184	0.167	0.053				
MF22062	1.113	0.083	0.711	0.044	0.108	0.015				
MF22066	1.259	0.067	-0.046	0.035	0.052	0.013				
MF22097	1.095	0.067	-0.482	0.057	0.116	0.026				
MF22101	0.855	0.058	-0.596	0.085	0.138	0.035				
MF22104	0.853	0.052	-0.532	0.069	0.086	0.028				
MF22105	0.593	0.051	0.611	0.084	0.080	0.025				
MF22106	0.981	0.044	0.590	0.034	0.000	0.000				
MF22108	0.851	0.055	-0.233	0.066	0.093	0.026				
MF22110	0.507	0.028	-0.013	0.051	0.000	0.000				
MF22127	1.529	0.150	1.355	0.045	0.112	0.001				
MF22135	0.731	0.056	0.925	0.058	0.043	0.014				
MF22139	1.274	0.104	0.949	0.041	0.115	0.013				
MF22142	1.401	0.090	0.426	0.035	0.099	0.014				
MF22144	0.688	0.063	0.564	0.081	0.116	0.027				
MF22146	1.186	0.068	0.328	0.035	0.050	0.012				
MF22148	1.232	0.051	0.244	0.026	0.000	0.000				
MF22154	1.176	0.086	0.461	0.046	0.145	0.018				
MF22156	1.490	0.062	0.341	0.023	0.000	0.000				
MF22181	1.131	0.070	-0.737	0.061	0.126	0.030				
MF22185	0.907	0.067	0.010	0.067	0.153	0.026				
MF22188	0.702	0.066	0.741	0.076	0.124	0.025				
MF22189	1.022	0.062	-0.310	0.055	0.096	0.023				
MF22191	1.012	0.062	-0.021	0.050	0.086	0.020				
MF22194	0.956	0.069	0.348	0.054	0.117	0.020				
MF22196	1.630	0.092	0.052	0.030	0.078	0.013				
MF22198	1.052	0.084	0.729	0.049	0.131	0.017				
MF22199	1.323	0.091	0.609	0.037	0.106	0.014				
MF22202	0.856	0.042	0.874	0.043	0.000	0.000				
MF22227A	1.186	0.050	0.144	0.027	0.000	0.000				
MF22227B	1.641	0.072	0.645	0.024	0.000	0.000				
MF22227C	1.401	0.068	1.043	0.033	0.000	0.000				
MF22232	0.542	0.021	1.653	0.054	0.000	0.000	-2.536	0.153	2.536	0.167
MF22234A	0.916	0.030	0.771	0.024	0.000	0.000	-0.432	0.046	0.432	0.052
MF22234B	0.972	0.033	1.044	0.025	0.000	0.000	-1.169	0.076	1.169	0.081
MF22243	1.206	0.052	0.503	0.028	0.000	0.000				
MF22251	0.977	0.111	1.379	0.067	0.157	0.016				
MF22252	1.125	0.083	0.150	0.054	0.182	0.023				
MF22253	1.373	0.056	0.165	0.024	0.000	0.000				
MF22257	1.343	0.102	0.456	0.044	0.202	0.018				

Exhibit D.1 **IRT Parameters for TIMSS Joint 1999-2003 Eighth-Grade Mathematics**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
MF22261A	1.326	0.057	0.578	0.027	0.000	0.000				
MF22261B	1.615	0.076	0.941	0.028	0.000	0.000				
MF22261C	0.950	0.035	1.184	0.028	0.000	0.000	-1.538	0.104	1.538	0.109
MF32036	1.016	0.079	0.395	0.057	0.173	0.021				
MF32047	1.548	0.202	1.069	0.056	0.380	0.015				
MF32064	1.645	0.071	0.348	0.022	0.000	0.000				
MF32094	1.310	0.087	-0.067	0.045	0.159	0.021				
MF32097	1.117	0.125	1.303	0.060	0.169	0.014				
MF32100	0.953	0.072	0.350	0.054	0.121	0.021				
MF32116	1.114	0.101	0.601	0.054	0.216	0.020				
MF32132	0.773	0.072	0.752	0.066	0.116	0.022				
MF32142	2.347	0.262	0.856	0.039	0.359	0.014				
MF32160	1.847	0.154	0.999	0.032	0.121	0.001				
MF32163	1.518	0.146	0.939	0.044	0.230	0.014				
MF32166	1.022	0.075	-0.070	0.066	0.187	0.027				
MF32198	0.885	0.066	0.218	0.063	0.124	0.024				
MF32205	0.556	0.067	0.612	0.135	0.195	0.039				
MF32233	1.016	0.037	1.210	0.029	0.000	0.000	-0.541	0.053	0.541	0.063
MF32273	1.162	0.083	-0.294	0.064	0.228	0.028				
MF32294	1.239	0.087	0.059	0.050	0.182	0.021				
MF32295	1.124	0.067	-0.756	0.058	0.110	0.028				
MF32307	1.279	0.057	0.786	0.030	0.000	0.000				
MF32324	1.430	0.110	0.771	0.036	0.116	0.012				
MF32331	1.947	0.235	1.322	0.044	0.196	0.011				
MF32344	1.368	0.058	0.379	0.025	0.000	0.000				
MF32352	1.292	0.118	0.253	0.060	0.350	0.022				
MF32381	1.009	0.043	0.147	0.030	0.000	0.000				
MF32397	1.401	0.119	0.734	0.041	0.180	0.015				
MF32398	1.614	0.148	0.762	0.040	0.241	0.015				
MF32402	0.940	0.099	0.798	0.066	0.228	0.022				
MF32414	1.156	0.049	0.492	0.030	0.000	0.000				
MF32416	1.118	0.074	0.574	0.040	0.071	0.013				
MF32419	1.536	0.145	0.823	0.042	0.238	0.015				
MF32424	0.976	0.078	0.470	0.054	0.138	0.020				
MF32447	1.282	0.094	0.630	0.041	0.135	0.015				
MF32477	1.567	0.121	0.528	0.037	0.183	0.015				
MF32507	1.930	0.166	0.884	0.033	0.155	0.011				
MF32523	1.413	0.126	1.036	0.043	0.151	0.013				
MF32525	1.080	0.066	-0.040	0.047	0.091	0.020				
MF32529	1.700	0.148	0.893	0.037	0.185	0.012				
MF32538	1.213	0.051	0.261	0.027	0.000	0.000				
MF32540	0.897	0.085	0.193	0.088	0.296	0.030				
MF32570	1.403	0.100	0.190	0.044	0.203	0.019				
MF32575	1.818	0.138	0.429	0.035	0.219	0.015				
MF32579	1.028	0.062	-0.135	0.050	0.087	0.021				
MF32595	1.128	0.073	0.177	0.046	0.104	0.019				
MF32609	0.963	0.057	-0.796	0.066	0.095	0.028				
MF32623	1.715	0.115	0.429	0.030	0.119	0.013				
MF32626	0.877	0.073	0.428	0.068	0.169	0.025				
MF32637A	0.886	0.038	-0.024	0.033	0.000	0.000				
MF32637B	1.054	0.044	-0.043	0.029	0.000	0.000				

Exhibit D.1 **IRT Parameters for TIMSS Joint 1999-2003 Eighth-Grade Mathematics**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
MF32637C	1.076	0.046	0.409	0.031	0.000	0.000				
MF32640	0.569	0.022	1.234	0.047	0.000	0.000	-0.587	0.067	0.587	0.086
MF32643	1.126	0.081	0.542	0.045	0.122	0.017				
MF32662	1.756	0.165	1.230	0.039	0.094	0.009				
MF32670	0.635	0.044	-0.981	0.119	0.109	0.041				
MF32673	1.325	0.097	0.447	0.044	0.179	0.018				
MF32679	0.954	0.072	0.115	0.061	0.149	0.024				
MF32681A	0.541	0.030	-0.237	0.050	0.000	0.000				
MF32681B	0.607	0.034	0.904	0.062	0.000	0.000				
MF32681C	0.977	0.044	0.511	0.034	0.000	0.000				
MF32683	0.579	0.018	0.565	0.031	0.000	0.000	-1.461	0.083	1.461	0.088
MF32688	0.878	0.041	0.765	0.041	0.000	0.000				
MF32690	0.869	0.079	0.758	0.064	0.157	0.021				
MF32691	0.902	0.040	0.306	0.033	0.000	0.000				
MF32692	0.608	0.020	1.176	0.037	0.000	0.000	-1.551	0.093	1.551	0.103
MF32693	0.861	0.039	0.436	0.036	0.000	0.000				
MF32695	0.517	0.016	0.215	0.031	0.000	0.000	-1.485	0.085	1.485	0.087
MF32698	1.134	0.079	0.366	0.046	0.108	0.018				
MF32701	1.095	0.060	-0.976	0.054	0.063	0.023				
MF32704	1.118	0.072	-0.074	0.050	0.123	0.022				
MF32721	0.608	0.095	1.404	0.123	0.244	0.030				
MF32725	1.101	0.051	0.695	0.034	0.000	0.000				
MF32727	1.432	0.096	0.361	0.037	0.133	0.015				
MF32728	1.111	0.111	1.080	0.055	0.191	0.016				
MF32732	0.742	0.072	0.471	0.090	0.208	0.029				
MF32734	0.809	0.037	-0.052	0.035	0.000	0.000				
MF32738	1.086	0.074	-0.418	0.064	0.171	0.029				
MF32743	0.603	0.030	0.105	0.045	0.000	0.000				
MF32744	0.835	0.040	0.757	0.042	0.000	0.000				
MF32745	0.579	0.029	2.309	0.095	0.000	0.000	-1.169	0.109	1.169	0.155
MF32753A	0.807	0.028	0.979	0.030	0.000	0.000	-0.504	0.053	0.504	0.064
MF32753B	0.885	0.033	1.132	0.032	0.000	0.000	-0.309	0.048	0.309	0.062
MF32753C	0.746	0.038	0.905	0.051	0.000	0.000				
MF32754	0.710	0.033	-0.454	0.041	0.000	0.000				
MF32755	0.956	0.038	1.320	0.034	0.000	0.000	-0.272	0.048	0.272	0.065
MF32756	0.556	0.032	0.792	0.062	0.000	0.000				
MF32757	0.413	0.012	-0.107	0.036	0.000	0.000	-2.750	0.132	2.750	0.132
MF32760A	0.766	0.023	0.798	0.027	0.000	0.000	-1.360	0.080	1.360	0.084
MF32760B	1.383	0.070	1.214	0.036	0.000	0.000				
MF32760C	1.859	0.104	1.340	0.031	0.000	0.000				
MF32761	1.236	0.049	1.300	0.026	0.000	0.000	-0.312	0.045	0.312	0.054

Exhibit D.2 **IRT Parameters for TIMSS Joint 1999-2003 Eighth-Grade Science**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S012001	0.587	0.016	-0.757	0.045	0.178	0.015				
S012002	0.587	0.024	-0.014	0.051	0.332	0.014				
S012003	1.006	0.021	-0.877	0.025	0.248	0.012				
S012004	0.612	0.022	-0.308	0.052	0.332	0.015				
S012005	0.748	0.024	-0.071	0.032	0.281	0.011				
S012006	0.906	0.023	-0.115	0.022	0.222	0.009				
S012007	1.149	0.055	-0.528	0.048	0.555	0.015				
S012008	0.631	0.048	0.525	0.062	0.361	0.018				
S012009	1.350	0.056	0.720	0.018	0.134	0.006				
S012010	0.911	0.028	-1.818	0.050	0.162	0.021				
S012011	0.789	0.061	0.917	0.044	0.323	0.012				
S012012	1.072	0.068	-0.218	0.059	0.661	0.013				
S012013	0.685	0.038	0.878	0.035	0.125	0.011				
S012014	1.154	0.038	-0.724	0.035	0.330	0.016				
S012015	0.852	0.030	-0.543	0.042	0.247	0.017				
S012016	0.718	0.037	-0.570	0.076	0.443	0.021				
S012017	1.528	0.055	0.306	0.016	0.240	0.008				
S012018	0.508	0.036	0.079	0.097	0.342	0.024				
S012019	0.728	0.035	0.295	0.036	0.121	0.014				
S012020	0.959	0.034	-0.759	0.041	0.207	0.018				
S012021	1.233	0.055	0.524	0.021	0.149	0.008				
S012022	0.668	0.034	-0.236	0.060	0.220	0.021				
S012023	1.001	0.044	-0.402	0.043	0.320	0.017				
S012024	0.872	0.038	-0.647	0.054	0.304	0.021				
S012025	0.770	0.069	1.142	0.053	0.349	0.012				
S012026	0.691	0.037	-0.593	0.084	0.473	0.021				
S012027	0.769	0.021	-1.214	0.037	0.075	0.014				
S012028	0.829	0.028	-0.046	0.027	0.113	0.011				
S012029	0.687	0.048	0.390	0.058	0.398	0.016				
S012030	0.678	0.035	0.234	0.045	0.211	0.016				
S012031	1.008	0.044	-0.489	0.045	0.324	0.019				
S012032	1.309	0.042	-0.506	0.025	0.178	0.014				
S012033	0.491	0.030	-0.708	0.116	0.211	0.032				
S012034	0.885	0.031	-0.712	0.041	0.159	0.018				
S012035	0.963	0.035	-1.032	0.048	0.237	0.021				
S012036	1.221	0.045	-0.432	0.031	0.253	0.015				
S012037	0.672	0.027	-1.679	0.091	0.331	0.028				
S012038	1.174	0.048	0.070	0.027	0.344	0.011				
S012039	0.977	0.044	-0.382	0.048	0.465	0.015				
S012040	1.668	0.060	0.247	0.016	0.286	0.008				
S012041	0.607	0.035	0.209	0.058	0.264	0.018				
S012042	0.977	0.047	0.310	0.031	0.307	0.012				
S012043	0.908	0.039	-0.458	0.048	0.267	0.019				
S012044	0.539	0.024	-1.601	0.104	0.138	0.032				
S012045	1.344	0.062	-0.604	0.043	0.489	0.017				
S012046	0.575	0.050	0.762	0.067	0.261	0.020				
S012047	1.099	0.080	1.232	0.042	0.152	0.008				
S012048	0.908	0.037	-0.121	0.034	0.167	0.015				
S022002	1.009	0.046	0.079	0.032	0.220	0.014				
S022007	0.769	0.048	-0.579	0.075	0.134	0.030				
S022009	0.937	0.063	-1.373	0.102	0.310	0.041				
S022012	1.457	0.119	0.612	0.033	0.175	0.013				

Exhibit D.2 **IRT Parameters for TIMSS Joint 1999-2003 Eighth-Grade Science**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S022014	0.382	0.045	-0.467	0.246	0.198	0.056				
S022017	1.165	0.045	0.331	0.024	0.000	0.000				
S022019	1.046	0.046	-0.446	0.043	0.330	0.018				
S022022	0.761	0.019	-0.255	0.017	0.000	0.000				
S022030	0.832	0.057	-0.712	0.089	0.213	0.037				
S022035	0.351	0.015	-0.207	0.038	0.000	0.000				
S022040	0.685	0.032	-0.542	0.060	0.151	0.022				
S022041	0.816	0.033	-1.115	0.057	0.168	0.023				
S022042	1.169	0.043	-0.095	0.025	0.167	0.012				
S022043	0.978	0.044	0.621	0.035	0.000	0.000				
S022048	0.973	0.045	0.631	0.034	0.000	0.000				
S022049	0.874	0.036	0.131	0.027	0.000	0.000				
S022054	1.053	0.047	0.062	0.030	0.226	0.014				
S022058	0.820	0.052	-0.138	0.064	0.348	0.021				
S022064	0.468	0.061	0.977	0.118	0.127	0.030				
S022069	1.005	0.020	0.045	0.012	0.000	0.000				
S022073	0.966	0.067	-0.724	0.081	0.296	0.034				
S022074	1.168	0.059	0.113	0.031	0.253	0.014				
S022078	1.204	0.025	-0.287	0.012	0.000	0.000				
S022081	0.917	0.036	-0.818	0.030	0.000	0.000				
S022082	1.393	0.120	0.859	0.038	0.117	0.011				
S022086	1.035	0.026	-0.181	0.015	0.000	0.000				
S022088A	0.865	0.019	-0.864	0.018	0.000	0.000				
S022088B	0.576	0.016	-0.213	0.020	0.000	0.000				
S022090	0.600	0.021	-0.368	0.025	0.000	0.000	-0.133	0.050	0.133	0.046
S022094	0.722	0.072	0.842	0.066	0.107	0.019				
S022099	0.763	0.076	0.410	0.071	0.213	0.026				
S022106	0.752	0.053	1.159	0.048	0.093	0.010				
S022115	0.991	0.039	-0.267	0.035	0.197	0.016				
S022117	0.776	0.049	0.358	0.045	0.201	0.017				
S022118	1.848	0.135	0.337	0.028	0.253	0.014				
S022123	1.093	0.104	0.373	0.056	0.328	0.022				
S022126	0.555	0.037	0.186	0.071	0.195	0.023				
S022131	0.719	0.049	-0.861	0.101	0.190	0.038				
S022132	1.227	0.127	0.783	0.046	0.225	0.015				
S022137	1.197	0.099	0.531	0.040	0.200	0.017				
S022140	0.827	0.033	-0.455	0.028	0.000	0.000				
S022141	1.053	0.045	0.609	0.031	0.000	0.000				
S022145	0.742	0.047	-0.297	0.066	0.105	0.026				
S022150	0.955	0.045	0.134	0.033	0.212	0.014				
S022152	1.063	0.026	-0.132	0.014	0.000	0.000				
S022154	0.649	0.020	-0.682	0.025	0.000	0.000				
S022157	1.079	0.096	0.373	0.049	0.249	0.020				
S022158	1.171	0.044	0.003	0.021	0.000	0.000				
S022160	0.670	0.021	0.398	0.026	0.000	0.000				
S022161	0.607	0.020	0.213	0.026	0.000	0.000				
S022165D	0.733	0.024	-0.343	0.020	0.000	0.000	0.049	0.037	-0.049	0.035
S022172A	0.735	0.024	-1.302	0.035	0.000	0.000				
S022172B	0.588	0.022	-1.447	0.048	0.000	0.000				
S022174	0.696	0.031	-0.191	0.031	0.000	0.000				
S022178	1.186	0.082	0.142	0.041	0.190	0.020				

Exhibit D.2 **IRT Parameters for TIMSS Joint 1999-2003 Eighth-Grade Science**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S022181	0.978	0.053	0.379	0.032	0.238	0.013				
S022183	1.347	0.069	0.574	0.022	0.221	0.009				
S022187	0.597	0.043	0.631	0.054	0.132	0.018				
S022188	1.239	0.110	0.752	0.039	0.436	0.011				
S022191	0.661	0.013	-0.756	0.016	0.000	0.000	-0.259	0.031	0.259	0.027
S022194	1.008	0.084	0.537	0.044	0.148	0.017				
S022198	1.452	0.104	0.814	0.029	0.269	0.009				
S022202	0.787	0.052	0.441	0.044	0.206	0.017				
S022206	0.752	0.054	0.592	0.046	0.199	0.016				
S022208	1.150	0.070	0.597	0.029	0.284	0.011				
S022213	0.835	0.041	0.764	0.045	0.000	0.000				
S022217A	1.068	0.042	0.191	0.024	0.000	0.000				
S022217D	0.711	0.025	0.440	0.025	0.000	0.000	-0.048	0.038	0.048	0.045
S022222	1.258	0.059	0.224	0.024	0.183	0.011				
S022225	1.068	0.079	1.195	0.045	0.107	0.008				
S022235	1.170	0.097	0.564	0.041	0.450	0.013				
S022238	0.760	0.079	0.509	0.073	0.220	0.027				
S022240	1.420	0.123	1.070	0.038	0.269	0.008				
S022244	1.166	0.028	0.563	0.016	0.000	0.000				
S022245	0.800	0.097	0.788	0.072	0.230	0.023				
S022249D	0.828	0.019	-0.156	0.015	0.000	0.000				
S022254	1.534	0.134	0.667	0.035	0.200	0.013				
S022258	0.910	0.037	0.093	0.026	0.000	0.000				
S022264	0.902	0.113	0.972	0.071	0.207	0.019				
S022268	0.562	0.015	0.268	0.022	0.000	0.000				
S022275	1.347	0.082	0.817	0.026	0.156	0.008				
S022276	0.780	0.045	0.094	0.050	0.262	0.018				
S022277D	0.547	0.019	-0.445	0.025	0.000	0.000	-0.092	0.050	0.092	0.046
S022278	1.166	0.074	-0.271	0.048	0.214	0.024				
S022279	0.698	0.021	-0.178	0.020	0.000	0.000				
S022280	1.868	0.135	0.181	0.030	0.292	0.015				
S022281	0.557	0.019	0.740	0.035	0.000	0.000				
S022282	1.030	0.048	1.566	0.060	0.000	0.000				
S022283	0.876	0.023	-0.935	0.022	0.000	0.000				
S022284	1.130	0.044	0.262	0.024	0.000	0.000				
S022286	0.854	0.026	0.982	0.029	0.000	0.000				
S022288	0.710	0.019	0.888	0.025	0.000	0.000	-0.299	0.030	0.299	0.041
S022289	0.818	0.014	0.430	0.011	0.000	0.000	0.669	0.014	-0.669	0.020
S022290	1.296	0.056	0.061	0.025	0.272	0.012				
S022292	0.731	0.019	-0.088	0.017	0.000	0.000				
S022293	1.181	0.102	0.562	0.042	0.214	0.017				
S022294	1.147	0.055	-0.117	0.036	0.363	0.015				
S022295	0.872	0.067	-0.126	0.069	0.216	0.028				
S032007	0.877	0.044	0.045	0.033	0.000	0.000				
S032008	0.952	0.071	-0.119	0.066	0.322	0.026				
S032015	0.815	0.045	0.403	0.041	0.000	0.000				
S032019A	1.032	0.061	0.841	0.050	0.000	0.000				
S032019B	1.196	0.087	1.314	0.076	0.000	0.000				
S032024	1.058	0.152	0.901	0.073	0.252	0.021				
S032035	1.203	0.054	0.050	0.026	0.144	0.013				
S032055	0.989	0.064	-1.266	0.098	0.403	0.038				

Exhibit D.2 **IRT Parameters for TIMSS Joint 1999-2003 Eighth-Grade Science**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S032056	0.841	0.044	0.117	0.035	0.000	0.000				
S032057	1.284	0.046	0.558	0.022	0.000	0.000				
S032060	0.733	0.039	-0.993	0.048	0.000	0.000				
S032063	0.769	0.025	0.907	0.030	0.000	0.000	-0.148	0.032	0.148	0.047
S032083	0.850	0.076	0.888	0.052	0.108	0.015				
S032087	0.610	0.099	0.937	0.122	0.214	0.033				
S032115	1.240	0.091	0.052	0.041	0.122	0.020				
S032120A	0.839	0.030	0.882	0.034	0.000	0.000				
S032120B	1.076	0.042	1.177	0.039	0.000	0.000				
S032122	0.618	0.040	0.647	0.063	0.000	0.000				
S032126	0.610	0.035	-0.504	0.045	0.000	0.000				
S032131	0.950	0.031	-0.574	0.022	0.000	0.000				
S032141	1.559	0.154	0.598	0.039	0.187	0.016				
S032150	0.594	0.049	-0.242	0.110	0.218	0.035				
S032151	1.034	0.091	0.283	0.051	0.148	0.022				
S032156	1.291	0.121	0.409	0.044	0.196	0.019				
S032158	0.870	0.107	0.164	0.097	0.350	0.033				
S032160	0.927	0.127	0.291	0.096	0.415	0.031				
S032184	0.501	0.080	0.683	0.155	0.221	0.042				
S032202	0.589	0.019	-0.441	0.021	0.000	0.000	0.278	0.039	-0.278	0.036
S032206	1.112	0.043	0.716	0.029	0.000	0.000				
S032238	1.305	0.104	0.162	0.042	0.166	0.020				
S032242	0.657	0.031	0.768	0.047	0.000	0.000				
S032257	1.461	0.171	0.661	0.047	0.253	0.017				
S032258	0.883	0.042	-0.423	0.049	0.175	0.021				
S032272	0.899	0.056	1.007	0.063	0.000	0.000				
S032273	0.680	0.152	1.447	0.184	0.264	0.028				
S032279	0.766	0.107	0.958	0.090	0.163	0.024				
S032281	1.368	0.076	-0.212	0.037	0.244	0.019				
S032301	1.554	0.116	0.582	0.029	0.220	0.012				
S032306	0.462	0.015	0.159	0.033	0.000	0.000	-1.558	0.088	1.558	0.092
S032310D	0.593	0.024	-0.305	0.028	0.000	0.000	-0.064	0.055	0.064	0.054
S032315	0.906	0.097	0.147	0.079	0.263	0.031				
S032369	0.651	0.028	0.364	0.031	0.000	0.000	-0.111	0.049	0.111	0.058
S032375	0.634	0.021	0.371	0.029	0.000	0.000	-1.087	0.068	1.087	0.074
S032385	0.854	0.050	-0.362	0.064	0.299	0.024				
S032386	1.110	0.084	0.782	0.038	0.118	0.012				
S032392	0.485	0.047	-2.016	0.262	0.192	0.068				
S032394	1.107	0.115	0.219	0.064	0.291	0.026				
S032403	1.073	0.142	0.640	0.065	0.279	0.023				
S032422	1.298	0.099	-0.132	0.048	0.198	0.024				
S032425	1.013	0.114	0.357	0.067	0.263	0.026				
S032437	1.004	0.109	0.681	0.056	0.351	0.018				
S032446	1.026	0.090	0.263	0.056	0.347	0.021				
S032451	0.540	0.017	-0.410	0.028	0.000	0.000	-1.457	0.083	1.457	0.080
S032463	1.957	0.194	0.162	0.041	0.407	0.019				
S032465	0.996	0.104	-0.223	0.093	0.362	0.035				
S032502	0.875	0.077	0.327	0.056	0.106	0.022				
S032510	0.976	0.085	-0.691	0.101	0.325	0.041				
S032514	0.878	0.132	0.705	0.087	0.314	0.027				
S032516	0.707	0.038	-0.706	0.043	0.000	0.000				

Exhibit D.2 **IRT Parameters for TIMSS Joint 1999-2003 Eighth-Grade Science**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S032519	0.736	0.023	0.090	0.022	0.000	0.000				
S032530D	0.492	0.025	0.099	0.037	0.000	0.000	0.590	0.059	-0.590	0.066
S032532	0.760	0.039	-0.490	0.037	0.000	0.000				
S032542	1.345	0.148	0.402	0.053	0.322	0.021				
S032555	1.023	0.051	0.339	0.033	0.000	0.000				
S032562	0.747	0.024	-0.134	0.022	0.000	0.000	-0.504	0.050	0.504	0.051
S032564	1.734	0.141	0.743	0.029	0.211	0.010				
S032565	0.815	0.047	0.539	0.046	0.000	0.000				
S032570	0.753	0.043	0.484	0.047	0.000	0.000				
S032574	1.079	0.133	0.415	0.069	0.325	0.025				
S032579	0.924	0.149	0.912	0.090	0.283	0.024				
S032595	1.052	0.106	0.993	0.053	0.164	0.013				
S032606	0.890	0.070	-1.400	0.124	0.269	0.051				
S032607	0.760	0.050	-0.497	0.081	0.220	0.031				
S032611	1.015	0.141	0.902	0.073	0.222	0.021				
S032614	0.646	0.037	-0.346	0.041	0.000	0.000				
S032620	0.553	0.096	1.251	0.146	0.150	0.031				
S032625A	0.828	0.030	-0.027	0.024	0.000	0.000				
S032625B	1.107	0.038	0.317	0.022	0.000	0.000				
S032626	1.270	0.040	-0.006	0.017	0.000	0.000				
S032637	0.811	0.075	0.663	0.055	0.192	0.020				
S032640	0.473	0.032	-0.389	0.054	0.000	0.000				
S032645	1.072	0.137	0.608	0.066	0.295	0.023				
S032650D	0.484	0.022	-0.093	0.033	0.000	0.000	-0.174	0.065	0.174	0.067
S032651A	1.420	0.061	-0.030	0.022	0.000	0.000				
S032651B	0.975	0.055	0.781	0.048	0.000	0.000				
S032652	0.826	0.057	0.093	0.054	0.170	0.022				
S032654	0.994	0.102	0.245	0.065	0.236	0.027				
S032656	0.914	0.047	-0.238	0.042	0.099	0.018				
S032660	1.330	0.225	1.133	0.082	0.260	0.016				
S032663	0.458	0.089	1.137	0.185	0.219	0.043				
S032665A	0.915	0.046	0.163	0.033	0.000	0.000				
S032665B	0.990	0.057	0.834	0.050	0.000	0.000				
S032665C	0.921	0.053	0.761	0.050	0.000	0.000				
S032672	0.320	0.048	-0.422	0.362	0.207	0.070				
S032679	0.882	0.057	1.097	0.070	0.000	0.000				
S032680	0.612	0.025	-0.661	0.031	0.000	0.000	-0.001	0.058	0.001	0.050
S032682	1.361	0.117	0.691	0.036	0.260	0.013				
S032683	0.996	0.061	0.378	0.035	0.200	0.015				
S032693A	0.934	0.045	-0.266	0.030	0.000	0.000				
S032693B	0.813	0.035	0.383	0.028	0.000	0.000	0.635	0.036	-0.635	0.049
S032695	0.694	0.029	0.347	0.030	0.000	0.000	-0.085	0.047	0.085	0.056
S032697D	0.912	0.035	0.258	0.022	0.000	0.000	-0.055	0.037	0.055	0.042
S032704	0.820	0.045	0.344	0.041	0.000	0.000				
S032705A	1.056	0.050	0.087	0.029	0.000	0.000				
S032705B	1.037	0.048	-0.220	0.028	0.000	0.000				
S032706A	0.986	0.049	0.258	0.033	0.000	0.000				
S032706B	1.188	0.057	0.315	0.029	0.000	0.000				
S032707	1.557	0.093	0.950	0.041	0.000	0.000				
S032709	1.446	0.049	0.485	0.019	0.000	0.000				
S032711	0.878	0.022	0.617	0.019	0.000	0.000	-0.515	0.033	0.515	0.040

Exhibit D.2 **IRT Parameters for TIMSS Joint 1999-2003 Eighth-Grade Science**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S032712A	0.929	0.034	0.310	0.025	0.000	0.000				
S032712B	1.211	0.050	0.899	0.033	0.000	0.000				
S032713A	1.070	0.045	0.897	0.036	0.000	0.000				
S032713B	1.006	0.055	1.456	0.067	0.000	0.000				
S032714	1.735	0.153	-0.378	0.055	0.423	0.026				
SF12001	0.817	0.050	-0.398	0.050	0.048	0.018				
SF12002	0.802	0.048	-0.580	0.053	0.048	0.018				
SF12003	1.350	0.073	-0.557	0.037	0.066	0.018				
SF12004	0.977	0.062	-0.325	0.049	0.077	0.021				
SF12005	0.994	0.062	-0.127	0.041	0.054	0.017				
SF12006	1.135	0.077	-0.002	0.041	0.090	0.019				
SF12013	0.697	0.073	0.722	0.073	0.081	0.021				
SF12014	1.344	0.078	-0.728	0.045	0.106	0.024				
SF12015	0.962	0.063	-0.480	0.057	0.098	0.025				
SF12016	0.978	0.059	-0.874	0.062	0.094	0.027				
SF12017	1.306	0.089	0.086	0.035	0.095	0.017				
SF12018	0.618	0.049	-0.391	0.086	0.085	0.029				
SF12025	0.495	0.061	0.504	0.120	0.122	0.035				
SF12026	0.908	0.072	-0.911	0.101	0.250	0.042				
SF12027	0.987	0.058	-0.829	0.056	0.081	0.024				
SF12028	0.879	0.062	0.063	0.048	0.064	0.018				
SF12029	0.848	0.089	0.234	0.075	0.206	0.030				
SF12030	0.585	0.064	0.417	0.093	0.113	0.030				
SF12037	0.994	0.057	-1.076	0.058	0.067	0.022				
SF12038	0.811	0.059	-0.309	0.064	0.095	0.026				
SF12039	1.000	0.059	-0.805	0.054	0.076	0.023				
SF12040	1.253	0.088	0.028	0.039	0.110	0.019				
SF12041	0.777	0.056	-0.091	0.055	0.065	0.021				
SF12042	0.856	0.062	-0.104	0.054	0.082	0.022				
SF22002	1.382	0.099	0.141	0.035	0.119	0.017				
SF22019	1.207	0.081	-0.351	0.049	0.143	0.025				
SF22022	0.944	0.046	0.066	0.031	0.000	0.000				
SF22035	0.534	0.035	0.115	0.052	0.000	0.000				
SF22040	1.272	0.080	-0.017	0.033	0.066	0.015				
SF22041	1.143	0.066	-0.336	0.038	0.061	0.017				
SF22042	1.378	0.097	0.134	0.035	0.114	0.017				
SF22054	1.232	0.099	0.115	0.046	0.167	0.022				
SF22058	0.845	0.062	-0.295	0.065	0.112	0.027				
SF22069	1.443	0.065	0.282	0.024	0.000	0.000				
SF22074	1.105	0.083	0.051	0.046	0.123	0.021				
SF22078	1.410	0.060	-0.052	0.022	0.000	0.000				
SF22086	1.362	0.061	0.101	0.023	0.000	0.000				
SF22088A	1.389	0.060	-0.111	0.022	0.000	0.000				
SF22088B	1.065	0.053	0.269	0.030	0.000	0.000				
SF22106	0.744	0.085	1.150	0.093	0.055	0.015				
SF22115	1.077	0.071	-0.200	0.046	0.094	0.021				
SF22117	0.713	0.067	0.276	0.072	0.109	0.026				
SF22126	0.788	0.074	0.190	0.071	0.133	0.028				
SF22150	1.087	0.085	0.204	0.045	0.112	0.020				
SF22152	1.348	0.060	0.089	0.023	0.000	0.000				
SF22154	1.043	0.050	0.034	0.028	0.000	0.000				

Exhibit D.2 **IRT Parameters for TIMSS Joint 1999-2003 Eighth-Grade Science**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
SF22160	0.871	0.048	0.549	0.044	0.000	0.000				
SF22161	0.821	0.047	0.489	0.045	0.000	0.000				
SF22181	1.061	0.097	0.282	0.053	0.168	0.023				
SF22183	1.248	0.115	0.477	0.043	0.157	0.018				
SF22187	0.787	0.079	0.504	0.063	0.099	0.023				
SF22188	0.849	0.094	0.359	0.074	0.208	0.028				
SF22191	0.969	0.033	-0.146	0.019	0.000	0.000	-0.158	0.037	0.158	0.037
SF22198	0.907	0.103	0.733	0.064	0.133	0.021				
SF22202	0.792	0.069	0.309	0.059	0.087	0.022				
SF22206	0.692	0.066	0.473	0.068	0.081	0.022				
SF22208	1.140	0.109	0.406	0.049	0.174	0.021				
SF22222	1.467	0.108	0.334	0.031	0.084	0.013				
SF22225	1.008	0.114	1.064	0.072	0.076	0.014				
SF22235	0.763	0.088	0.375	0.083	0.190	0.031				
SF22240	0.917	0.127	1.023	0.082	0.162	0.020				
SF22244	1.676	0.082	0.578	0.026	0.000	0.000				
SF22249D	1.326	0.065	0.442	0.029	0.000	0.000				
SF22268	0.960	0.052	0.594	0.042	0.000	0.000				
SF22275	1.090	0.098	0.701	0.048	0.070	0.014				
SF22276	0.931	0.078	0.002	0.062	0.146	0.027				
SF22279	0.942	0.049	0.287	0.034	0.000	0.000				
SF22281	0.843	0.050	0.763	0.053	0.000	0.000				
SF22283	1.260	0.055	-0.323	0.024	0.000	0.000				
SF22286	1.304	0.089	1.167	0.063	0.000	0.000				
SF22289	1.105	0.046	0.581	0.024	0.000	0.000	0.505	0.028	-0.505	0.044
SF22290	1.302	0.108	0.040	0.049	0.220	0.023				
SF22292	0.714	0.040	0.143	0.040	0.000	0.000				
SF22294	1.257	0.097	-0.107	0.050	0.202	0.025				
SF32007	1.031	0.049	0.047	0.029	0.000	0.000				
SF32015	0.999	0.051	0.428	0.036	0.000	0.000				
SF32019A	1.083	0.066	0.989	0.054	0.000	0.000				
SF32019B	1.244	0.092	1.383	0.077	0.000	0.000				
SF32024	1.297	0.158	0.753	0.053	0.222	0.017				
SF32035	1.320	0.091	0.010	0.036	0.101	0.017				
SF32056	0.987	0.051	0.361	0.034	0.000	0.000				
SF32060	1.210	0.053	-0.445	0.025	0.000	0.000				
SF32087	0.532	0.072	0.737	0.115	0.128	0.033				
SF32115	1.316	0.092	0.201	0.034	0.090	0.016				
SF32120A	1.198	0.068	0.872	0.045	0.000	0.000				
SF32120B	1.459	0.093	1.098	0.050	0.000	0.000				
SF32122	0.788	0.047	0.763	0.057	0.000	0.000				
SF32126	0.812	0.042	-0.011	0.034	0.000	0.000				
SF32131	1.414	0.060	-0.165	0.022	0.000	0.000				
SF32141	1.718	0.169	0.635	0.036	0.168	0.014				
SF32151	1.444	0.118	0.306	0.036	0.149	0.017				
SF32156	0.983	0.096	0.377	0.055	0.153	0.023				
SF32158	0.868	0.089	0.032	0.081	0.232	0.032				
SF32160	0.869	0.095	0.033	0.089	0.277	0.034				
SF32184	0.622	0.095	0.880	0.110	0.183	0.032				
SF32202	0.910	0.034	0.069	0.021	0.000	0.000	0.064	0.036	-0.064	0.039
SF32238	1.234	0.094	0.172	0.039	0.117	0.018				

Exhibit D.2 **IRT Parameters for TIMSS Joint 1999-2003 Eighth-Grade Science**

(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
SF32257	1.325	0.139	0.528	0.047	0.218	0.018				
SF32258	1.150	0.079	-0.187	0.048	0.125	0.023				
SF32272	1.243	0.072	0.914	0.047	0.000	0.000				
SF32273	0.728	0.142	1.326	0.144	0.234	0.026				
SF32279	0.992	0.128	0.878	0.068	0.159	0.019				
SF32306	0.577	0.019	0.161	0.029	0.000	0.000	-1.155	0.071	1.155	0.075
SF32310D	0.617	0.023	-0.326	0.027	0.000	0.000	-0.261	0.055	0.261	0.054
SF32315	0.971	0.087	0.045	0.064	0.194	0.027				
SF32369	0.786	0.030	0.272	0.025	0.000	0.000	-0.156	0.042	0.156	0.049
SF32375	0.593	0.020	0.477	0.031	0.000	0.000	-1.155	0.071	1.155	0.078
SF32385	1.134	0.082	-0.265	0.054	0.158	0.027				
SF32392	0.592	0.047	-1.484	0.148	0.138	0.046				
SF32394	1.014	0.103	0.236	0.063	0.229	0.026				
SF32403	1.142	0.131	0.587	0.055	0.238	0.021				
SF32422	1.318	0.100	-0.009	0.046	0.185	0.023				
SF32425	1.022	0.103	0.308	0.059	0.211	0.024				
SF32451	0.668	0.020	-0.203	0.024	0.000	0.000	-0.991	0.063	0.991	0.063
SF32463	1.377	0.104	0.077	0.040	0.163	0.019				
SF32465	0.935	0.075	-0.488	0.079	0.204	0.034				
SF32502	0.909	0.090	0.544	0.056	0.112	0.020				
SF32510	0.862	0.069	-0.688	0.095	0.208	0.039				
SF32514	1.032	0.123	0.617	0.061	0.218	0.022				
SF32516	0.822	0.041	-0.323	0.034	0.000	0.000				
SF32519	1.071	0.052	0.293	0.031	0.000	0.000				
SF32530D	0.618	0.028	0.121	0.031	0.000	0.000	0.584	0.047	-0.584	0.055
SF32532	0.905	0.044	-0.384	0.031	0.000	0.000				
SF32542	1.273	0.125	0.245	0.051	0.263	0.022				
SF32555	1.156	0.058	0.501	0.034	0.000	0.000				
SF32562	0.778	0.025	0.000	0.022	0.000	0.000	-0.483	0.048	0.483	0.049
SF32565	0.864	0.051	0.746	0.051	0.000	0.000				
SF32570	0.947	0.051	0.586	0.043	0.000	0.000				
SF32574	1.000	0.128	0.458	0.075	0.315	0.028				
SF32579	1.166	0.183	0.913	0.075	0.291	0.019				
SF32595	1.364	0.137	0.804	0.046	0.087	0.012				
SF32606	1.059	0.067	-1.233	0.072	0.143	0.035				
SF32611	0.949	0.102	0.595	0.057	0.138	0.021				
SF32614	0.714	0.038	-0.423	0.038	0.000	0.000				
SF32620	0.825	0.119	1.133	0.095	0.135	0.020				
SF32625A	1.713	0.078	0.324	0.022	0.000	0.000				
SF32625B	2.133	0.106	0.529	0.021	0.000	0.000				
SF32640	0.717	0.038	-0.243	0.037	0.000	0.000				
SF32645	1.210	0.167	0.762	0.061	0.281	0.019				
SF32650D	0.727	0.026	0.109	0.025	0.000	0.000	-0.311	0.047	0.311	0.051
SF32651A	1.550	0.066	0.042	0.021	0.000	0.000				
SF32651B	1.351	0.070	0.643	0.033	0.000	0.000				
SF32654	0.847	0.079	0.242	0.062	0.126	0.025				
SF32656	1.669	0.120	0.284	0.028	0.098	0.013				
SF32660	0.769	0.121	1.138	0.107	0.173	0.023				
SF32663	0.657	0.098	0.950	0.105	0.168	0.028				
SF32665A	1.031	0.053	0.457	0.035	0.000	0.000				
SF32665B	1.162	0.068	0.869	0.047	0.000	0.000				

Exhibit D.2 **IRT Parameters for TIMSS Joint 1999-2003 Eighth-Grade Science**

(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
SF32665C	1.053	0.063	0.894	0.051	0.000	0.000				
SF32672	0.598	0.093	0.376	0.151	0.325	0.042				
SF32679	1.143	0.066	0.830	0.045	0.000	0.000				
SF32680	0.752	0.026	-0.547	0.024	0.000	0.000	-0.221	0.049	0.221	0.045
SF32683	1.402	0.106	0.347	0.034	0.102	0.015				
SF32693A	0.936	0.046	0.019	0.030	0.000	0.000				
SF32693B	0.743	0.034	0.659	0.033	0.000	0.000	0.647	0.038	-0.647	0.059
SF32695	0.750	0.031	0.594	0.032	0.000	0.000	-0.180	0.046	0.180	0.058
SF32697D	0.842	0.033	0.609	0.029	0.000	0.000	-0.213	0.042	0.213	0.053
SF32704	0.816	0.046	0.565	0.046	0.000	0.000				
SF32705A	1.101	0.052	0.218	0.029	0.000	0.000				
SF32705B	1.185	0.052	-0.058	0.025	0.000	0.000				
SF32706A	0.884	0.048	0.505	0.041	0.000	0.000				
SF32706B	1.122	0.057	0.580	0.035	0.000	0.000				
SF32707	1.619	0.010	1.049	0.043	0.000	0.000				
SF32714	1.424	0.108	-0.482	0.059	0.293	0.031				

Exhibit D.3 **IRT Parameters for TIMSS Joint 1995-2003 Fourth-Grade Mathematics**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M011001	0.869	0.033	-0.841	0.064	0.398	0.023				
M011002	0.839	0.033	0.418	0.035	0.274	0.012				
M011003	0.646	0.025	-0.184	0.060	0.206	0.020				
M011004	0.725	0.026	-1.152	0.081	0.298	0.029				
M011005	0.474	0.026	-1.508	0.205	0.349	0.051				
M011006	0.409	0.021	-0.135	0.099	0.069	0.027				
M011007	0.896	0.036	-1.500	0.079	0.237	0.037				
M011008	1.277	0.052	0.100	0.030	0.294	0.013				
M011009	1.046	0.040	-1.431	0.058	0.146	0.031				
M011010	1.119	0.048	-0.154	0.040	0.242	0.018				
M011011	1.261	0.055	-0.705	0.048	0.320	0.023				
M011012	0.761	0.028	-1.452	0.067	0.083	0.028				
M011013	0.759	0.052	0.542	0.064	0.326	0.020				
M011014	0.646	0.028	-2.047	0.115	0.119	0.044				
M011015	0.785	0.038	-0.084	0.058	0.190	0.022				
M011016	1.009	0.049	0.339	0.038	0.241	0.015				
M011017	0.651	0.028	-0.707	0.078	0.111	0.030				
M011018	0.715	0.028	-1.191	0.075	0.101	0.031				
M011019	0.881	0.035	-0.463	0.050	0.136	0.022				
M011020	1.267	0.082	1.146	0.031	0.281	0.001				
M011021	0.754	0.035	-0.484	0.071	0.186	0.028				
M011022	0.393	0.023	-1.290	0.183	0.117	0.048				
M011023	0.466	0.033	-0.896	0.202	0.225	0.055				
M011024	0.697	0.034	-2.233	0.135	0.167	0.057				
M011025	0.762	0.047	0.614	0.053	0.241	0.018				
M011026	0.605	0.037	-0.303	0.109	0.247	0.034				
M011027	0.707	0.032	-0.741	0.081	0.167	0.032				
M011028	0.626	0.032	-0.506	0.093	0.172	0.033				
M011029	0.549	0.055	0.036	0.135	0.113	0.042				
M011030	1.400	0.146	0.918	0.050	0.210	0.018				
M011031	0.927	0.083	0.111	0.080	0.167	0.033				
M011032	1.243	0.107	-1.368	0.111	0.251	0.059				
M011033	0.731	0.074	0.270	0.102	0.153	0.037				
M011034	0.730	0.053	-1.037	0.109	0.096	0.040				
M011035	1.337	0.094	-0.179	0.050	0.123	0.025				
M011036	0.425	0.085	0.702	0.290	0.263	0.068				
M011037	0.809	0.070	-1.923	0.174	0.190	0.069				
M011038	0.620	0.076	-0.281	0.199	0.273	0.060				
M011039	0.962	0.073	-0.815	0.091	0.137	0.041				
M011040	1.134	0.099	0.521	0.052	0.142	0.021				
M011041	1.273	0.133	0.787	0.054	0.195	0.020				
M011042	0.614	0.055	-0.739	0.155	0.145	0.052				
M011043	1.437	0.139	0.150	0.062	0.337	0.027				
M011044	1.361	0.113	0.283	0.049	0.182	0.023				
M011045	0.794	0.079	-0.761	0.161	0.283	0.059				
M011046	0.466	0.048	-0.722	0.209	0.139	0.057				
M011047	0.604	0.056	-1.536	0.219	0.198	0.073				
M011048	0.753	0.076	0.265	0.098	0.153	0.036				
M011049	0.934	0.084	0.320	0.071	0.126	0.029				
M011050	0.443	0.047	-1.434	0.290	0.183	0.075				
M011051	0.753	0.073	-0.253	0.127	0.198	0.047				
M011052	1.089	0.085	-0.659	0.088	0.197	0.042				

Exhibit D.3 **IRT Parameters for TIMSS Joint 1995-2003 Fourth-Grade Mathematics**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M011053	1.131	0.112	0.821	0.053	0.146	0.019				
M011054	1.203	0.057	-0.063	0.029	0.000	0.000				
M011055	0.816	0.096	-1.557	0.251	0.448	0.081				
M011056	0.654	0.088	0.924	0.105	0.155	0.034				
M011057	0.645	0.084	0.349	0.141	0.206	0.046				
M011058	1.204	0.096	0.358	0.049	0.125	0.021				
M011059	0.742	0.078	-0.166	0.136	0.226	0.049				
M011060	0.393	0.044	-1.688	0.342	0.188	0.079				
M011061	0.688	0.038	-0.205	0.046	0.000	0.000				
M011062	0.643	0.073	-0.198	0.171	0.252	0.054				
M011063	0.571	0.036	0.232	0.053	0.000	0.000				
M011064	1.317	0.166	1.119	0.060	0.263	0.019				
M011065	1.043	0.091	-0.503	0.098	0.264	0.042				
M011066	0.705	0.104	1.088	0.105	0.209	0.032				
M011067	0.820	0.072	-1.707	0.172	0.220	0.069				
M011068	0.807	0.077	-0.477	0.132	0.245	0.049				
M011069	0.640	0.016	0.018	0.021	0.000	0.000	-0.722	0.050	0.722	0.049
M011070	1.058	0.038	-0.434	0.026	0.000	0.000				
M011071	0.771	0.022	0.637	0.021	0.000	0.000	-0.313	0.038	0.313	0.043
M011072	0.858	0.034	0.016	0.028	0.000	0.000				
M011073	0.514	0.026	-0.111	0.044	0.000	0.000				
M011074A	1.000	0.039	-1.007	0.036	0.000	0.000				
M011074B	0.683	0.020	0.141	0.021	0.000	0.000	-0.165	0.042	0.165	0.042
M011075	0.608	0.027	0.093	0.036	0.000	0.000				
M011076	0.794	0.031	0.012	0.029	0.000	0.000				
M011077A	0.838	0.038	1.114	0.043	0.000	0.000				
M011077B	1.229	0.055	1.232	0.035	0.000	0.000				
M011078	0.624	0.029	-0.950	0.052	0.000	0.000				
M011079	0.377	0.010	-0.382	0.033	0.000	0.000	-1.620	0.085	1.620	0.081
M011080A	1.014	0.036	-0.454	0.026	0.000	0.000				
M011080B	1.183	0.041	0.143	0.021	0.000	0.000				
M011080C	0.607	0.029	-1.490	0.067	0.000	0.000				
M011081	0.730	0.029	-0.552	0.036	0.000	0.000				
M011082	0.623	0.031	-1.741	0.077	0.000	0.000				
M011083	0.591	0.021	0.225	0.025	0.000	0.000	0.414	0.042	-0.414	0.044
M011084	0.994	0.035	0.048	0.024	0.000	0.000				
M011085	0.830	0.032	-0.688	0.034	0.000	0.000				
M011086A	0.357	0.011	-0.012	0.046	0.000	0.000	1.964	0.073	-1.964	0.072
M011086B	0.929	0.034	-0.087	0.026	0.000	0.000				
M011087	0.455	0.024	-0.206	0.049	0.000	0.000				
M012023	0.765	0.040	-0.301	0.076	0.265	0.027				
M012030	1.698	0.230	1.434	0.054	0.161	0.012				
M012044	1.131	0.052	0.190	0.035	0.251	0.015				
M012048	0.807	0.040	0.127	0.053	0.189	0.020				
M012054	0.386	0.022	-0.588	0.061	0.000	0.000				
M012065	0.900	0.053	0.864	0.038	0.191	0.014				
M012069	0.412	0.062	1.630	0.137	0.294	0.033				
M012078	0.710	0.028	-1.008	0.071	0.102	0.029				
M012080	0.971	0.080	0.903	0.051	0.049	0.014				
M012081	0.798	0.066	-0.744	0.121	0.162	0.049				
M012088	0.888	0.081	0.176	0.079	0.154	0.032				

Exhibit D.3 **IRT Parameters for TIMSS Joint 1995-2003 Fourth-Grade Mathematics**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M012117	0.946	0.045	0.597	0.033	0.213	0.013				
M012119	0.684	0.047	0.172	0.087	0.334	0.026				
M012126	0.711	0.027	-0.592	0.060	0.115	0.024				
M012139	0.856	0.078	-0.297	0.110	0.203	0.044				
M031004	0.804	0.128	1.254	0.102	0.140	0.028				
M031006	0.474	0.056	-1.538	0.296	0.182	0.077				
M031008	1.155	0.138	1.410	0.060	0.201	0.016				
M031009	0.813	0.056	0.541	0.052	0.000	0.000				
M031011	0.825	0.038	0.188	0.033	0.000	0.000				
M031016	1.143	0.075	0.857	0.046	0.000	0.000				
M031023	0.575	0.074	0.191	0.194	0.283	0.055				
M031029	1.188	0.176	0.299	0.110	0.455	0.036				
M031030	0.845	0.071	1.548	0.094	0.000	0.000				
M031038	0.614	0.077	-0.597	0.227	0.235	0.071				
M031041	0.647	0.027	0.082	0.034	0.000	0.000				
M031043	1.274	0.123	0.179	0.062	0.154	0.029				
M031045	1.122	0.064	-0.446	0.060	0.175	0.030				
M031050	1.324	0.097	0.638	0.042	0.288	0.017				
M031051	0.853	0.060	-0.587	0.095	0.150	0.040				
M031064	1.271	0.166	0.771	0.068	0.253	0.026				
M031065	1.107	0.046	0.192	0.027	0.000	0.000				
M031068	1.149	0.040	0.334	0.021	0.000	0.000				
M031071	1.199	0.148	0.895	0.066	0.188	0.025				
M031079B	0.937	0.061	-0.899	0.060	0.000	0.000				
M031079C	0.609	0.048	0.543	0.067	0.000	0.000				
M031083	0.908	0.096	-0.387	0.126	0.191	0.052				
M031085	0.615	0.119	0.791	0.181	0.256	0.053				
M031088	0.561	0.070	-0.874	0.252	0.211	0.075				
M031093	0.953	0.197	1.079	0.123	0.401	0.033				
M031097	1.071	0.130	0.605	0.079	0.200	0.032				
M031098	1.130	0.101	0.074	0.066	0.117	0.029				
M031106	0.876	0.033	0.324	0.026	0.000	0.000				
M031108	1.110	0.076	0.333	0.047	0.116	0.021				
M031109	0.520	0.073	-0.391	0.256	0.210	0.072				
M031128	0.420	0.040	-1.409	0.146	0.000	0.000				
M031130	0.897	0.058	-0.470	0.052	0.000	0.000				
M031133	0.601	0.048	-1.478	0.114	0.000	0.000				
M031134	0.439	0.027	1.392	0.081	0.000	0.000				
M031135	0.994	0.087	-0.886	0.108	0.151	0.048				
M031155	1.205	0.136	0.068	0.086	0.267	0.038				
M031159	0.754	0.084	-0.394	0.148	0.187	0.054				
M031162	0.475	0.030	-1.106	0.083	0.000	0.000				
M031172	1.206	0.122	-0.185	0.087	0.233	0.041				
M031173	1.173	0.010	-0.382	0.075	0.130	0.036				
M031178	0.870	0.098	0.777	0.076	0.094	0.027				
M031183	0.549	0.034	-0.195	0.047	0.000	0.000	0.537	0.084	-0.537	0.075
M031185	1.425	0.154	0.431	0.062	0.231	0.028				
M031187	1.764	0.279	0.403	0.082	0.553	0.027				
M031190	1.148	0.090	0.342	0.055	0.197	0.025				
M031210	1.746	0.250	0.897	0.060	0.329	0.022				
M031216	0.716	0.066	-0.611	0.162	0.270	0.057				

Exhibit D.3 **IRT Parameters for TIMSS Joint 1995-2003 Fourth-Grade Mathematics**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M031218	1.270	0.144	0.346	0.072	0.248	0.031				
M031219	0.346	0.068	0.519	0.343	0.180	0.073				
M031220	0.797	0.054	-1.043	0.110	0.132	0.046				
M031227	0.976	0.044	1.413	0.043	0.000	0.000				
M031235	0.723	0.029	0.463	0.032	0.000	0.000				
M031240	0.686	0.030	-1.015	0.050	0.000	0.000				
M031242A	0.923	0.057	-0.368	0.047	0.000	0.000				
M031242B	1.046	0.063	0.198	0.039	0.000	0.000				
M031242C	0.849	0.107	0.102	0.130	0.252	0.048				
M031245	1.597	0.180	1.053	0.048	0.115	0.015				
M031247	0.481	0.030	1.308	0.076	0.000	0.000	-0.325	0.094	0.325	0.125
M031249	0.905	0.053	1.556	0.063	0.000	0.000				
M031251	1.466	0.198	0.852	0.065	0.285	0.024				
M031252	0.841	0.088	-0.382	0.127	0.181	0.050				
M031254	1.049	0.118	0.288	0.085	0.210	0.035				
M031255	1.057	0.089	0.481	0.063	0.350	0.023				
M031258	0.942	0.037	0.853	0.030	0.000	0.000				
M031264	1.029	0.050	-1.239	0.049	0.000	0.000				
M031265	0.584	0.033	-0.217	0.049	0.000	0.000				
M031267	0.514	0.031	0.316	0.050	0.000	0.000				
M031269	0.305	0.011	-1.126	0.065	0.000	0.000	-2.009	0.141	2.009	0.123
M031271	0.559	0.029	-1.819	0.089	0.000	0.000				
M031272A	0.811	0.042	-1.231	0.060	0.000	0.000				
M031272B	0.717	0.045	-2.066	0.108	0.000	0.000				
M031272C	0.918	0.041	0.120	0.031	0.000	0.000				
M031274	0.708	0.029	-0.662	0.040	0.000	0.000				
M031276	1.274	0.130	0.072	0.073	0.219	0.033				
M031282	0.697	0.018	0.946	0.023	0.000	0.000	-1.013	0.053	1.013	0.059
M031285	0.742	0.031	0.800	0.036	0.000	0.000				
M031286	0.909	0.034	0.457	0.026	0.000	0.000				
M031294	1.153	0.120	0.012	0.082	0.215	0.037				
M031297	0.529	0.044	0.427	0.073	0.000	0.000				
M031298	0.833	0.041	0.845	0.040	0.000	0.000				
M031299	1.270	0.043	0.115	0.020	0.000	0.000				
M031301	0.948	0.035	-0.639	0.031	0.000	0.000				
M031303	1.461	0.148	-0.266	0.077	0.274	0.039				
M031304	0.972	0.043	-0.356	0.033	0.000	0.000				
M031305	0.712	0.035	-0.757	0.049	0.000	0.000				
M031306	0.759	0.036	-0.227	0.038	0.000	0.000				
M031309	1.089	0.064	-0.208	0.040	0.000	0.000				
M031310	1.380	0.098	-0.475	0.064	0.251	0.033				
M031313	0.587	0.047	-1.303	0.109	0.000	0.000				
M031315	0.915	0.073	0.155	0.076	0.172	0.032				
M031316	0.470	0.048	-2.688	0.246	0.000	0.000				
M031317	1.273	0.153	0.750	0.062	0.194	0.025				
M031322	0.420	0.029	-1.458	0.111	0.000	0.000				
M031325	0.926	0.064	0.827	0.053	0.000	0.000				
M031327	0.430	0.028	0.132	0.058	0.000	0.000				
M031330	0.472	0.043	-1.889	0.165	0.000	0.000				
M031332	1.014	0.124	0.252	0.010	0.266	0.039				
M031333	0.897	0.101	0.562	0.081	0.123	0.032				

Exhibit D.3 **IRT Parameters for TIMSS Joint 1995-2003 Fourth-Grade Mathematics**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M031334	1.098	0.083	0.794	0.045	0.211	0.017				
M031335	0.959	0.059	-0.037	0.061	0.168	0.027				
M031338	0.612	0.071	0.151	0.165	0.249	0.051				
M031341	0.789	0.061	-0.755	0.124	0.194	0.050				
M031344A	0.510	0.043	0.364	0.073	0.000	0.000				
M031344B	0.766	0.052	0.257	0.050	0.000	0.000				
M031344C	0.456	0.019	-0.123	0.046	0.000	0.000	-1.983	0.145	1.983	0.141
M031345A	0.734	0.050	-0.412	0.058	0.000	0.000				
M031345B	0.650	0.047	-0.254	0.061	0.000	0.000				
M031345C	0.653	0.061	1.700	0.126	0.000	0.000				
M031346A	1.222	0.071	-0.395	0.040	0.000	0.000				
M031346B	1.272	0.076	0.545	0.036	0.000	0.000				
M031346C	0.814	0.044	0.325	0.033	0.000	0.000	0.380	0.055	-0.380	0.058
M031347A	0.620	0.033	0.068	0.042	0.000	0.000				
M031347B	0.571	0.032	0.234	0.046	0.000	0.000				
M031347C	0.852	0.040	0.470	0.034	0.000	0.000				
M031348A	0.660	0.036	0.585	0.045	0.000	0.000				
M031348B	0.559	0.029	1.455	0.053	0.000	0.000	0.690	0.051	-0.690	0.087
M031350A	0.865	0.033	0.562	0.028	0.000	0.000				
M031350B	0.843	0.032	0.019	0.028	0.000	0.000				
M031350C	0.638	0.029	0.914	0.043	0.000	0.000				
M031351	0.626	0.083	0.085	0.169	0.188	0.054				
M031379	0.784	0.057	1.044	0.066	0.000	0.000				
M031380	0.872	0.067	1.365	0.076	0.000	0.000				
MF11001	1.921	0.201	-0.114	0.061	0.276	0.035				
MF11002	1.412	0.136	0.514	0.049	0.129	0.023				
MF11003	1.037	0.084	0.191	0.055	0.057	0.020				
MF11004	1.300	0.097	-0.235	0.051	0.066	0.022				
MF11005	1.147	0.096	-0.275	0.069	0.108	0.031				
MF11006	0.787	0.080	0.281	0.087	0.095	0.033				
MF11007	2.351	0.240	-0.238	0.052	0.284	0.033				
MF11008	1.893	0.158	0.212	0.038	0.106	0.020				
MF11009	1.598	0.151	-1.263	0.083	0.160	0.047				
MF11010	1.068	0.100	-0.252	0.086	0.157	0.038				
MF11011	1.445	0.120	-0.854	0.071	0.124	0.036				
MF11012	1.193	0.095	-1.009	0.080	0.094	0.035				
MF11013	0.734	0.076	0.077	0.103	0.110	0.037				
MF11014	1.160	0.091	-0.968	0.079	0.088	0.033				
MF11015	0.882	0.080	0.056	0.076	0.087	0.029				
MF11016	0.963	0.103	0.419	0.076	0.137	0.030				
MF11017	0.919	0.071	-0.742	0.082	0.069	0.028				
MF11018	0.879	0.078	-1.020	0.121	0.131	0.048				
MF11019	0.976	0.077	-0.593	0.080	0.083	0.031				
MF11020	0.753	0.107	0.914	0.099	0.139	0.032				
MF11021	0.996	0.078	-0.171	0.064	0.065	0.024				
MF11022	0.937	0.073	-0.200	0.065	0.058	0.023				
MF11023	0.695	0.063	-0.877	0.133	0.107	0.044				
MF11024	1.251	0.109	-1.323	0.095	0.127	0.045				
MF11025	1.102	0.105	0.596	0.055	0.081	0.022				
MF11026	1.529	0.141	0.379	0.047	0.135	0.024				
MF11027	1.806	0.159	0.142	0.044	0.140	0.024				

Exhibit D.3 **IRT Parameters for TIMSS Joint 1995-2003 Fourth-Grade Mathematics**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
MF11028	1.439	0.126	0.208	0.050	0.118	0.025				
MF12023	1.010	0.083	-0.170	0.055	0.056	0.021				
MF12044	0.728	0.070	-0.224	0.107	0.101	0.037				
MF12048	0.836	0.070	-0.289	0.081	0.074	0.029				
MF12065	0.753	0.093	0.690	0.091	0.111	0.031				
MF12069	0.782	0.110	1.070	0.092	0.115	0.030				
MF12078	1.291	0.093	-0.427	0.052	0.055	0.020				
MF12117	1.164	0.109	0.519	0.054	0.091	0.023				
MF12119	0.885	0.077	0.022	0.073	0.077	0.027				
MF12126	1.169	0.109	-0.126	0.079	0.163	0.038				
MF31004	0.808	0.105	1.073	0.086	0.105	0.026				
MF31006	0.870	0.113	0.162	0.131	0.289	0.046				
MF31009	0.780	0.054	0.723	0.056	0.000	0.000				
MF31016	1.259	0.081	0.917	0.043	0.000	0.000				
MF31029	0.767	0.117	0.389	0.153	0.313	0.048				
MF31030	0.748	0.063	1.616	0.103	0.000	0.000				
MF31038	1.172	0.116	0.046	0.077	0.191	0.034				
MF31041	0.683	0.048	0.123	0.055	0.000	0.000				
MF31043	1.254	0.124	0.442	0.060	0.151	0.026				
MF31045	1.236	0.101	-0.243	0.064	0.112	0.030				
MF31050	1.508	0.178	0.487	0.064	0.295	0.027				
MF31051	1.199	0.095	-0.245	0.062	0.089	0.028				
MF31064	0.972	0.116	0.864	0.073	0.137	0.026				
MF31065	1.312	0.077	0.310	0.033	0.000	0.000				
MF31068	1.288	0.074	0.256	0.034	0.000	0.000				
MF31071	1.024	0.126	0.977	0.072	0.142	0.025				
MF31079B	1.150	0.068	-0.528	0.044	0.000	0.000				
MF31079C	0.738	0.053	0.858	0.063	0.000	0.000				
MF31083	1.056	0.112	0.028	0.096	0.197	0.042				
MF31085	1.188	0.201	1.207	0.082	0.251	0.025				
MF31088	0.748	0.084	-0.105	0.139	0.176	0.051				
MF31093	0.474	0.084	0.819	0.204	0.162	0.056				
MF31097	1.611	0.174	0.951	0.047	0.117	0.016				
MF31098	1.650	0.139	0.222	0.043	0.108	0.020				
MF31106	0.907	0.056	0.146	0.044	0.000	0.000				
MF31109	1.181	0.115	0.315	0.063	0.140	0.028				
MF31128	0.617	0.046	-0.557	0.071	0.000	0.000				
MF31130	0.871	0.056	0.157	0.045	0.000	0.000				
MF31133	0.900	0.057	-0.379	0.049	0.000	0.000				
MF31134	0.501	0.046	1.245	0.115	0.000	0.000				
MF31135	0.961	0.090	0.041	0.081	0.123	0.033				
MF31155	1.279	0.127	0.198	0.068	0.190	0.031				
MF31159	1.608	0.133	0.199	0.042	0.096	0.021				
MF31172	1.355	0.132	0.395	0.056	0.159	0.026				
MF31173	1.392	0.103	0.136	0.042	0.054	0.017				
MF31178	1.758	0.214	1.143	0.049	0.128	0.015				
MF31183	0.989	0.052	0.330	0.029	0.000	0.000	0.374	0.046	-0.374	0.049
MF31185	1.618	0.151	0.345	0.049	0.157	0.024				
MF31187	1.302	0.125	0.037	0.068	0.182	0.033				
MF31210	1.302	0.155	0.730	0.065	0.227	0.026				
MF31218	1.924	0.145	0.318	0.031	0.048	0.012				

Exhibit D.3 **IRT Parameters for TIMSS Joint 1995-2003 Fourth-Grade Mathematics**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
MF31219	0.890	0.131	0.901	0.094	0.207	0.033				
MF31220	1.143	0.096	-0.371	0.075	0.117	0.035				
MF31227	1.005	0.078	1.394	0.073	0.000	0.000				
MF31235	0.758	0.051	0.356	0.052	0.000	0.000				
MF31240	0.768	0.052	-0.727	0.063	0.000	0.000				
MF31242A	1.109	0.066	0.179	0.037	0.000	0.000				
MF31242B	1.158	0.071	0.565	0.038	0.000	0.000				
MF31242C	1.588	0.179	0.560	0.056	0.257	0.026				
MF31245	1.864	0.192	1.028	0.040	0.093	0.013				
MF31247	0.592	0.037	1.556	0.076	0.000	0.000	-0.357	0.085	0.357	0.120
MF31251	1.571	0.166	0.744	0.049	0.163	0.020				
MF31252	0.805	0.079	-0.251	0.115	0.135	0.044				
MF31254	1.730	0.172	0.525	0.046	0.184	0.021				
MF31255	0.905	0.109	0.055	0.120	0.251	0.046				
MF31258	0.914	0.061	0.752	0.051	0.000	0.000				
MF31264	1.315	0.077	-0.359	0.038	0.000	0.000				
MF31265	0.878	0.059	0.352	0.047	0.000	0.000				
MF31269	0.492	0.021	-0.092	0.044	0.000	0.000	-1.400	0.120	1.400	0.116
MF31271	0.755	0.052	-1.086	0.076	0.000	0.000				
MF31274	0.914	0.058	-0.617	0.052	0.000	0.000				
MF31276	1.271	0.130	0.388	0.063	0.183	0.028				
MF31282	0.763	0.033	0.870	0.037	0.000	0.000	-0.999	0.088	0.999	0.097
MF31285	0.855	0.057	0.664	0.052	0.000	0.000				
MF31286	1.074	0.065	0.399	0.039	0.000	0.000				
MF31294	1.682	0.139	0.174	0.041	0.010	0.021				
MF31297	1.186	0.072	0.552	0.038	0.000	0.000				
MF31298	1.124	0.073	0.898	0.045	0.000	0.000				
MF31299	1.427	0.081	0.091	0.032	0.000	0.000				
MF31301	1.101	0.065	-0.473	0.043	0.000	0.000				
MF31303	1.399	0.132	-0.246	0.074	0.215	0.038				
MF31305	0.782	0.052	-0.586	0.061	0.000	0.000				
MF31309	1.402	0.080	-0.029	0.033	0.000	0.000				
MF31310	1.609	0.143	-0.288	0.060	0.176	0.034				
MF31313	0.573	0.045	-0.785	0.086	0.000	0.000				
MF31316	0.753	0.055	-1.385	0.092	0.000	0.000				
MF31317	1.170	0.116	0.562	0.059	0.123	0.024				
MF31322	0.751	0.052	-0.386	0.059	0.000	0.000				
MF31325	1.123	0.075	1.020	0.049	0.000	0.000				
MF31327	0.628	0.048	0.584	0.063	0.000	0.000				
MF31330	0.675	0.046	-0.493	0.064	0.000	0.000				
MF31332	1.145	0.122	0.358	0.073	0.196	0.031				
MF31333	1.547	0.185	1.052	0.053	0.142	0.017				
MF31334	1.055	0.135	0.809	0.076	0.193	0.028				
MF31335	1.178	0.110	0.010	0.071	0.155	0.032				
MF31344A	0.673	0.053	0.934	0.075	0.000	0.000				
MF31344B	1.299	0.079	0.619	0.037	0.000	0.000				
MF31344C	0.675	0.027	0.422	0.034	0.000	0.000	-1.394	0.105	1.394	0.108
MF31345A	0.828	0.054	0.224	0.047	0.000	0.000				
MF31345B	0.734	0.051	0.357	0.053	0.000	0.000				
MF31345C	0.836	0.076	1.760	0.115	0.000	0.000				
MF31346A	1.193	0.070	-0.307	0.039	0.000	0.000				

Exhibit D.3 **IRT Parameters for TIMSS Joint 1995-2003 Fourth-Grade Mathematics**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
MF31346B	1.172	0.072	0.727	0.040	0.000	0.000				
MF31346C	0.825	0.046	0.594	0.035	0.000	0.000	0.444	0.052	-0.444	0.061
MF31350A	1.213	0.072	0.441	0.037	0.000	0.000				
MF31350B	1.234	0.071	0.001	0.035	0.000	0.000				
MF31350C	0.995	0.064	0.706	0.047	0.000	0.000				
MF31351	1.143	0.135	0.814	0.067	0.174	0.025				
MF31379	0.978	0.069	1.214	0.061	0.000	0.000				
MF31380	1.120	0.083	1.448	0.065	0.000	0.000				

Exhibit D.4 **IRT Parameters for TIMSS Joint 1995-2003 Fourth-Grade Science**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S011001	0.704	0.024	-1.337	0.087	0.195	0.038				
S011002	0.894	0.032	-0.919	0.071	0.251	0.034				
S011003	1.500	0.057	0.474	0.022	0.397	0.011				
S011004	0.786	0.029	-0.184	0.054	0.180	0.024				
S011005	0.836	0.028	-0.515	0.056	0.181	0.026				
S011006	0.624	0.037	-0.413	0.124	0.257	0.043				
S011007	0.803	0.039	-0.303	0.078	0.232	0.034				
S011008	1.171	0.060	0.281	0.042	0.348	0.019				
S011009	0.573	0.033	-0.509	0.124	0.152	0.045				
S011010	0.733	0.042	-1.754	0.161	0.292	0.067				
S011011	1.194	0.075	0.942	0.030	0.242	0.014				
S011012	0.739	0.034	-1.246	0.098	0.140	0.045				
S011013	0.859	0.077	1.105	0.051	0.283	0.020				
S011014	1.076	0.056	0.212	0.046	0.240	0.023				
S011015	0.697	0.046	-0.244	0.119	0.292	0.043				
S011016	0.769	0.040	-1.458	0.130	0.265	0.056				
S011017	0.747	0.042	-0.161	0.085	0.197	0.036				
S011018	0.711	0.043	-1.650	0.172	0.352	0.064				
S011019	0.636	0.023	-1.034	0.047	0.000	0.000				
S011020	0.426	0.044	1.022	0.119	0.104	0.035				
S011021	0.709	0.050	-0.720	0.164	0.428	0.052				
S011022	1.031	0.074	0.451	0.060	0.428	0.023				
S011023	1.293	0.065	0.232	0.039	0.302	0.020				
S011024	1.198	0.062	-0.173	0.053	0.219	0.029				
S011025	0.770	0.039	-0.964	0.107	0.227	0.048				
S011026	0.356	0.025	-2.699	0.294	0.143	0.063				
S011027	1.200	0.067	0.115	0.052	0.379	0.024				
S011029	1.023	0.045	-1.168	0.078	0.203	0.044				
S011030	0.590	0.036	-0.913	0.159	0.237	0.056				
S011031	0.986	0.048	-1.048	0.093	0.275	0.049				
S011032	0.774	0.024	0.546	0.019	0.000	0.000				
S011033	0.505	0.105	2.249	0.175	0.324	0.028				
S011034	1.212	0.146	0.905	0.060	0.232	0.028				
S011035	0.977	0.091	-0.209	0.113	0.211	0.055				
S011036	0.962	0.118	0.813	0.075	0.190	0.035				
S011037	0.990	0.202	1.679	0.120	0.222	0.025				
S011038	0.749	0.078	-0.167	0.148	0.195	0.061				
S011039	0.908	0.113	0.912	0.075	0.158	0.033				
S011040	1.193	0.129	0.598	0.066	0.239	0.033				
S011041	1.038	0.102	0.185	0.089	0.207	0.045				
S011042	1.212	0.149	1.002	0.056	0.206	0.025				
S011043	0.866	0.075	-0.201	0.105	0.147	0.048				
S011044	0.585	0.077	-0.454	0.277	0.310	0.084				
S011045	1.301	0.153	0.667	0.064	0.311	0.030				
S011046	0.896	0.052	0.373	0.034	0.000	0.000				
S011047	1.339	0.119	-0.758	0.111	0.289	0.068				
S011048	1.345	0.107	-0.763	0.092	0.197	0.059				
S011049	1.635	0.271	1.421	0.063	0.259	0.018				
S011050	0.576	0.046	1.021	0.069	0.000	0.000				
S011051	1.579	0.156	0.937	0.039	0.135	0.019				
S011052	1.918	0.199	0.933	0.037	0.227	0.019				
S011053	1.235	0.113	0.417	0.059	0.170	0.033				

Exhibit D.4 **IRT Parameters for TIMSS Joint 1995-2003 Fourth-Grade Science**

(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S011054	1.103	0.096	-0.012	0.082	0.179	0.043				
S011055	0.766	0.091	0.503	0.107	0.166	0.045				
S011056	0.569	0.068	-0.603	0.254	0.247	0.081				
S011057	1.314	0.181	1.061	0.061	0.287	0.025				
S011058	1.282	0.131	0.747	0.051	0.183	0.026				
S011059	1.231	0.127	0.438	0.067	0.242	0.035				
S011060	0.875	0.119	1.011	0.079	0.170	0.034				
S011061	0.348	0.063	0.693	0.340	0.199	0.073				
S011062	1.239	0.099	-0.214	0.077	0.171	0.044				
S011063	0.703	0.089	0.352	0.143	0.202	0.055				
S011064	0.467	0.092	1.328	0.204	0.184	0.056				
S011065	0.980	0.108	0.431	0.087	0.206	0.041				
S011066	0.519	0.109	1.094	0.223	0.280	0.062				
S011067	1.128	0.160	1.066	0.069	0.255	0.029				
S011068	0.478	0.021	-0.121	0.032	0.000	0.000	0.607	0.057	-0.607	0.047
S011069	0.902	0.039	0.490	0.025	0.000	0.000				
S011070	1.255	0.048	0.173	0.020	0.000	0.000				
S011071D	0.773	0.030	0.435	0.019	0.000	0.000	0.164	0.036	-0.164	0.035
S011072	1.011	0.031	-0.773	0.021	0.000	0.000	1.121	0.055	-1.121	0.022
S011073	0.841	0.039	-0.911	0.051	0.000	0.000				
S011074	0.791	0.028	0.458	0.019	0.000	0.000	0.450	0.032	-0.450	0.033
S011075	1.043	0.049	-1.092	0.052	0.000	0.000				
S011076	0.500	0.059	0.107	0.220	0.221	0.065				
S011077D	0.617	0.025	0.121	0.024	0.000	0.000	0.455	0.044	-0.455	0.039
S011078D	0.580	0.025	0.031	0.027	0.000	0.000	0.234	0.049	-0.234	0.042
S011079	0.957	0.039	-0.266	0.030	0.000	0.000				
S011080	0.701	0.024	0.211	0.020	0.000	0.000	-0.012	0.039	0.012	0.036
S012007	0.633	0.040	-0.615	0.140	0.259	0.050				
S012010	1.023	0.050	0.068	0.050	0.207	0.025				
S012016	0.641	0.068	0.067	0.142	0.145	0.053				
S012020	1.181	0.131	0.617	0.067	0.239	0.033				
S012024	1.335	0.164	0.992	0.054	0.238	0.025				
S012033	0.580	0.042	0.217	0.114	0.223	0.038				
S012045	0.632	0.076	-0.203	0.203	0.233	0.070				
S012049	0.575	0.059	-0.245	0.159	0.135	0.054				
S012077	0.902	0.051	-0.054	0.038	0.000	0.000				
S012089	0.694	0.046	0.284	0.043	0.000	0.000				
S012096	0.732	0.047	-0.276	0.052	0.000	0.000				
S012097	0.772	0.110	0.956	0.096	0.178	0.039				
S012099	0.614	0.055	1.680	0.115	0.000	0.000				
S012104	1.158	0.047	0.858	0.024	0.000	0.000				
S012106	0.745	0.040	1.143	0.044	0.000	0.000				
S012123	0.533	0.145	1.671	0.251	0.385	0.050				
S012128A	0.789	0.035	-0.268	0.035	0.000	0.000				
S012128B	0.623	0.039	1.541	0.072	0.000	0.000				
S031001	0.696	0.080	-0.820	0.204	0.208	0.074				
S031003	0.624	0.063	-0.335	0.183	0.202	0.067				
S031005	0.873	0.058	1.467	0.063	0.000	0.000				
S031009	0.755	0.035	0.127	0.030	0.000	0.000				
S031017	0.709	0.061	-0.591	0.153	0.177	0.063				
S031026	0.564	0.017	0.042	0.024	0.000	0.000	-0.596	0.052	0.596	0.048

Exhibit D.4 **IRT Parameters for TIMSS Joint 1995-2003 Fourth-Grade Science**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S031035	0.792	0.085	-0.210	0.174	0.314	0.068				
S031038	0.500	0.056	-0.945	0.287	0.237	0.084				
S031044	0.544	0.053	0.485	0.065	0.000	0.000				
S031047	0.634	0.054	0.232	0.057	0.000	0.000				
S031053	0.613	0.025	0.143	0.027	0.000	0.000	-0.189	0.054	0.189	0.050
S031060	0.927	0.179	1.614	0.106	0.257	0.027				
S031061	0.613	0.076	-0.441	0.201	0.171	0.070				
S031068	1.746	0.242	0.898	0.054	0.298	0.027				
S031072	0.721	0.044	0.126	0.039	0.000	0.000	0.720	0.067	-0.720	0.058
S031075	0.435	0.010	0.972	0.337	0.299	0.076				
S031076	0.793	0.065	0.873	0.056	0.000	0.000				
S031077	0.384	0.061	-0.966	0.408	0.235	0.089				
S031078	1.112	0.119	0.662	0.076	0.395	0.032				
S031081	0.778	0.062	-0.538	0.078	0.000	0.000				
S031082	0.463	0.057	-0.417	0.284	0.218	0.079				
S031088D	0.271	0.014	0.853	0.010	0.000	0.000	2.420	0.147	-2.420	0.177
S031190	1.113	0.079	0.836	0.041	0.000	0.000				
S031193	0.645	0.088	0.099	0.183	0.183	0.066				
S031197D	0.412	0.028	-0.904	0.094	0.000	0.000	-0.571	0.143	0.571	0.108
S031204	0.362	0.046	1.131	0.137	0.000	0.000				
S031205	0.689	0.061	0.221	0.114	0.173	0.046				
S031212	0.691	0.082	-0.063	0.191	0.291	0.068				
S031218	0.698	0.042	-0.051	0.044	0.000	0.000				
S031229	1.330	0.105	0.814	0.039	0.232	0.020				
S031230	0.553	0.073	-1.442	0.314	0.235	0.090				
S031233	0.316	0.041	-0.623	0.155	0.000	0.000				
S031235A	1.280	0.048	0.652	0.019	0.000	0.000				
S031235B	1.302	0.050	0.791	0.020	0.000	0.000				
S031236	0.621	0.074	-1.203	0.239	0.195	0.075				
S031239	1.227	0.150	0.687	0.080	0.573	0.025				
S031240D	0.571	0.020	-0.164	0.029	0.000	0.000	0.984	0.052	-0.984	0.039
S031241D	0.656	0.029	0.711	0.029	0.000	0.000	0.750	0.043	-0.750	0.051
S031246	0.965	0.054	1.057	0.039	0.000	0.000				
S031251	0.608	0.044	1.226	0.067	0.000	0.000				
S031252	0.595	0.032	-0.861	0.056	0.000	0.000	0.346	0.080	-0.346	0.048
S031254	1.567	0.417	1.349	0.107	0.517	0.025				
S031255	1.177	0.092	0.297	0.065	0.307	0.032				
S031264	0.888	0.097	-0.119	0.129	0.169	0.059				
S031266	2.311	0.340	0.910	0.046	0.370	0.024				
S031269	0.824	0.129	0.996	0.105	0.306	0.041				
S031270	0.496	0.038	2.050	0.123	0.000	0.000				
S031273	1.523	0.282	0.988	0.078	0.447	0.029				
S031275	1.298	0.302	1.625	0.119	0.242	0.024				
S031278	0.576	0.033	-0.527	0.059	0.000	0.000				
S031281	0.504	0.066	-1.665	0.329	0.209	0.083				
S031283	0.792	0.136	0.118	0.223	0.440	0.069				
S031284	0.816	0.151	1.594	0.108	0.194	0.031				
S031287	0.772	0.074	0.044	0.124	0.187	0.052				
S031291	0.899	0.106	-0.570	0.184	0.281	0.079				
S031298	1.065	0.220	1.486	0.111	0.221	0.030				
S031299	0.509	0.051	0.790	0.081	0.000	0.000				

Exhibit D.4 **IRT Parameters for TIMSS Joint 1995-2003 Fourth-Grade Science**

(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S031306	1.028	0.116	1.044	0.056	0.164	0.025				
S031311	3.396	0.574	0.777	0.042	0.531	0.022				
S031313	1.389	0.194	1.265	0.055	0.292	0.022				
S031317	1.289	0.210	0.512	0.108	0.479	0.042				
S031319	1.404	0.112	1.057	0.033	0.168	0.016				
S031325	0.509	0.052	0.746	0.077	0.000	0.000				
S031326D	0.390	0.025	0.439	0.039	0.000	0.000	0.077	0.076	-0.077	0.076
S031330	0.738	0.044	-0.648	0.062	0.000	0.000				
S031338	0.881	0.089	-0.138	0.143	0.301	0.062				
S031340	1.150	0.197	1.148	0.082	0.249	0.032				
S031346	0.876	0.081	1.451	0.086	0.000	0.000				
S031347	0.571	0.073	-0.548	0.236	0.190	0.076				
S031349	0.563	0.055	-1.355	0.244	0.206	0.079				
S031356	0.485	0.072	-1.619	0.415	0.295	0.103				
S031361	0.938	0.171	0.916	0.115	0.317	0.044				
S031370	0.806	0.046	0.315	0.033	0.000	0.000				
S031371	2.038	0.333	1.090	0.054	0.340	0.024				
S031372A	1.140	0.052	0.035	0.027	0.000	0.000				
S031372B	0.781	0.032	1.059	0.028	0.000	0.000	-0.275	0.042	0.275	0.052
S031376	1.270	0.239	1.363	0.086	0.226	0.027				
S031379	0.736	0.073	0.101	0.128	0.179	0.053				
S031382	0.651	0.042	0.343	0.039	0.000	0.000				
S031383	0.722	0.078	0.974	0.075	0.094	0.031				
S031384A	0.924	0.043	-0.906	0.052	0.000	0.000				
S031384B	0.901	0.038	-0.216	0.032	0.000	0.000				
S031387	0.676	0.134	1.498	0.144	0.142	0.039				
S031389	1.664	0.326	1.340	0.075	0.290	0.023				
S031390D	0.504	0.038	0.647	0.047	0.000	0.000	0.153	0.081	-0.153	0.089
S031391D	0.489	0.033	0.529	0.045	0.000	0.000	-0.224	0.088	0.224	0.091
S031393	0.998	0.045	-0.786	0.046	0.000	0.000				
S031396D	0.498	0.037	-1.096	0.010	0.000	0.000	0.072	0.134	-0.072	0.084
S031398	0.719	0.103	0.343	0.155	0.188	0.061				
S031399A	1.294	0.048	0.515	0.018	0.000	0.000				
S031399B	1.227	0.046	0.262	0.019	0.000	0.000				
S031401	1.502	0.138	0.944	0.040	0.330	0.019				
S031406A	0.916	0.049	-0.477	0.047	0.000	0.000				
S031406B	1.044	0.063	1.352	0.049	0.000	0.000				
S031409	1.209	0.156	0.377	0.091	0.269	0.047				
S031410	0.439	0.067	-0.265	0.285	0.191	0.074				
S031414A	1.333	0.048	-0.078	0.021	0.000	0.000				
S031414B	1.146	0.044	-0.127	0.025	0.000	0.000				
S031418	1.452	0.265	1.209	0.074	0.307	0.027				
S031420	1.016	0.148	1.290	0.070	0.226	0.028				
S031421	0.409	0.046	-0.200	0.101	0.000	0.000				
S031422	1.011	0.113	-0.755	0.168	0.294	0.078				
S031426	0.872	0.123	0.092	0.164	0.313	0.065				
S031427	0.714	0.091	-0.102	0.176	0.213	0.067				
S031431	1.009	0.300	1.999	0.230	0.227	0.026				
S031439A	1.022	0.091	1.383	0.075	0.000	0.000				
S031439B	0.762	0.062	0.143	0.051	0.000	0.000				
S031440	0.962	0.078	0.991	0.053	0.000	0.000				

Exhibit D.4 **IRT Parameters for TIMSS Joint 1995-2003 Fourth-Grade Science**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S031441A	1.416	0.089	0.020	0.034	0.000	0.000				
S031441B	1.032	0.057	0.727	0.030	0.000	0.000	0.629	0.041	-0.629	0.051
S031442	1.311	0.088	0.438	0.031	0.000	0.000				
S031443	1.004	0.089	1.210	0.064	0.000	0.000				
S031445A	1.585	0.098	0.570	0.027	0.000	0.000				
S031445B	1.226	0.080	-0.385	0.049	0.000	0.000				
S031446A	1.005	0.073	0.786	0.042	0.000	0.000				
S031446B	0.824	0.069	1.071	0.062	0.000	0.000				
S031446C	0.782	0.060	0.095	0.051	0.000	0.000				
S031447	0.475	0.040	1.115	0.067	0.000	0.000	0.385	0.082	-0.385	0.106
SF11001	1.802	0.136	-0.273	0.049	0.080	0.028				
SF11003	1.406	0.108	0.070	0.047	0.061	0.023				
SF11004	1.540	0.112	0.108	0.040	0.046	0.018				
SF11005	2.511	0.182	-0.019	0.030	0.045	0.015				
SF11006	1.038	0.095	0.058	0.074	0.095	0.035				
SF11007	1.283	0.099	0.049	0.050	0.057	0.023				
SF11008	1.356	0.104	0.095	0.047	0.056	0.022				
SF11009	1.906	0.130	0.292	0.027	0.027	0.011				
SF11010	3.287	0.256	0.007	0.026	0.055	0.017				
SF11011	1.330	0.115	0.811	0.038	0.034	0.013				
SF11012	3.009	0.217	0.081	0.024	0.035	0.013				
SF11013	1.412	0.120	0.822	0.036	0.032	0.012				
SF11014	2.267	0.156	0.509	0.022	0.023	0.009				
SF11015	2.065	0.149	0.324	0.027	0.041	0.014				
SF11016	4.185	0.332	0.174	0.018	0.021	0.007				
SF11017	1.150	0.102	0.286	0.056	0.074	0.029				
SF11018	2.575	0.196	-0.205	0.037	0.078	0.026				
SF11019	1.731	0.103	0.176	0.026	0.000	0.000				
SF11021	1.700	0.122	-0.099	0.043	0.056	0.021				
SF11022	0.978	0.091	0.160	0.073	0.088	0.033				
SF11023	1.786	0.130	0.160	0.034	0.046	0.017				
SF11025	3.022	0.211	0.010	0.023	0.018	0.007				
SF11026	2.683	0.181	0.140	0.024	0.017	0.007				
SF11027	2.381	0.163	0.361	0.022	0.017	0.007				
SF11029	5.523	0.469	0.180	0.015	0.013	0.005				
SF11030	1.875	0.128	0.084	0.033	0.033	0.013				
SF11031	2.715	0.190	0.005	0.028	0.033	0.012				
SF11032	1.578	0.104	0.771	0.029	0.000	0.000				
SF11033	0.944	0.073	0.867	0.049	0.000	0.004				
SF12007	2.544	0.173	0.274	0.022	0.016	0.006				
SF12010	1.717	0.126	0.214	0.034	0.046	0.018				
SF12033	1.005	0.087	0.349	0.053	0.047	0.021				
SF31001	1.168	0.102	-0.354	0.087	0.126	0.047				
SF31005	1.277	0.105	1.375	0.061	0.000	0.000				
SF31009	1.236	0.080	0.426	0.032	0.000	0.000				
SF31017	1.164	0.118	-0.153	0.103	0.196	0.057				
SF31026	0.701	0.033	0.227	0.032	0.000	0.000	-0.611	0.074	0.611	0.071
SF31044	0.974	0.071	0.798	0.044	0.000	0.000				
SF31047	0.892	0.064	0.476	0.042	0.000	0.000				
SF31053	0.995	0.050	0.404	0.024	0.000	0.000	-0.087	0.048	0.087	0.047
SF31061	0.773	0.105	0.005	0.177	0.244	0.070				

Exhibit D.4 **IRT Parameters for TIMSS Joint 1995-2003 Fourth-Grade Science**

(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
SF31068	1.403	0.163	0.869	0.051	0.149	0.026				
SF31072	0.918	0.054	0.276	0.031	0.000	0.000	0.525	0.053	-0.525	0.047
SF31075	0.872	0.137	0.745	0.119	0.234	0.051				
SF31076	1.093	0.077	0.862	0.041	0.000	0.000				
SF31077	3.166	0.343	0.691	0.028	0.238	0.021				
SF31078	1.226	0.119	0.388	0.058	0.099	0.031				
SF31081	0.786	0.061	0.070	0.053	0.000	0.000				
SF31082	1.094	0.120	0.444	0.075	0.140	0.040				
SF31088D	0.995	0.060	1.046	0.034	0.000	0.000	0.506	0.040	-0.506	0.062
SF31190	1.460	0.093	0.833	0.031	0.000	0.000				
SF31193	0.871	0.111	0.532	0.107	0.168	0.049				
SF31197D	0.715	0.035	0.020	0.035	0.000	0.000	-0.422	0.074	0.422	0.066
SF31204	1.214	0.088	0.962	0.042	0.000	0.000				
SF31205	0.965	0.092	0.390	0.066	0.080	0.030				
SF31229	1.061	0.150	0.843	0.078	0.200	0.037				
SF31230	0.991	0.101	-0.468	0.135	0.200	0.065				
SF31233	1.070	0.073	0.487	0.036	0.000	0.000				
SF31235A	1.457	0.091	0.609	0.029	0.000	0.000				
SF31235B	1.673	0.104	0.686	0.027	0.000	0.000				
SF31236	0.840	0.091	-0.754	0.173	0.207	0.076				
SF31239	0.662	0.075	-0.314	0.158	0.148	0.057				
SF31240D	0.665	0.039	-0.040	0.042	0.000	0.000	0.809	0.074	-0.809	0.059
SF31246	1.283	0.092	1.033	0.042	0.000	0.000				
SF31251	0.836	0.075	1.282	0.076	0.000	0.000				
SF31254	1.541	0.207	0.748	0.060	0.288	0.032				
SF31255	1.147	0.114	0.183	0.075	0.142	0.040				
SF31264	1.151	0.120	0.369	0.075	0.148	0.041				
SF31266	1.630	0.152	0.634	0.038	0.096	0.022				
SF31270	0.658	0.071	1.719	0.137	0.000	0.000				
SF31273	2.559	0.226	0.693	0.026	0.103	0.017				
SF31275	1.025	0.196	1.575	0.109	0.139	0.026				
SF31278	0.881	0.063	0.033	0.046	0.000	0.000				
SF31281	3.948	0.430	0.522	0.026	0.292	0.023				
SF31283	0.777	0.082	-0.213	0.136	0.148	0.058				
SF31287	1.149	0.144	0.415	0.092	0.238	0.048				
SF31291	1.405	0.134	-0.206	0.085	0.200	0.050				
SF31298	0.997	0.209	1.522	0.114	0.205	0.032				
SF31299	1.319	0.095	1.008	0.041	0.000	0.000				
SF31306	1.071	0.137	1.049	0.063	0.094	0.026				
SF31311	2.711	0.267	0.727	0.028	0.160	0.019				
SF31317	1.101	0.124	0.111	0.107	0.224	0.055				
SF31319	1.296	0.145	0.944	0.050	0.097	0.022				
SF31325	0.853	0.066	0.799	0.049	0.000	0.000				
SF31340	1.063	0.141	0.851	0.074	0.161	0.036				
SF31346	1.281	0.104	1.433	0.061	0.000	0.000				
SF31347	0.975	0.108	0.192	0.104	0.171	0.051				
SF31356	1.321	0.156	-0.328	0.128	0.369	0.064				
SF31361	0.814	0.130	0.625	0.140	0.248	0.057				
SF31371	1.493	0.152	0.857	0.042	0.095	0.021				
SF31372A	1.934	0.114	0.487	0.023	0.000	0.000				
SF31372B	1.610	0.088	1.192	0.027	0.000	0.000	-0.091	0.036	0.091	0.049

Exhibit D.4 **IRT Parameters for TIMSS Joint 1995-2003 Fourth-Grade Science**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
SF31376	1.166	0.216	1.408	0.088	0.189	0.027				
SF31384A	1.626	0.097	-0.117	0.033	0.000	0.000				
SF31384B	1.592	0.096	0.308	0.027	0.000	0.000				
SF31387	1.018	0.170	1.410	0.087	0.125	0.027				
SF31389	1.860	0.212	1.150	0.041	0.076	0.014				
SF31390D	1.062	0.059	0.728	0.026	0.000	0.000	0.178	0.041	-0.178	0.047
SF31391D	0.705	0.044	0.699	0.035	0.000	0.000	0.075	0.059	-0.075	0.066
SF31393	1.273	0.081	-0.324	0.043	0.000	0.000				
SF31396D	0.550	0.027	-0.134	0.045	0.000	0.000	-0.869	0.100	0.869	0.090
SF31398	0.963	0.109	0.428	0.085	0.133	0.041				
SF31399A	1.744	0.105	0.575	0.025	0.000	0.000				
SF31399B	1.555	0.094	0.396	0.027	0.000	0.000				
SF31401	1.467	0.168	0.808	0.050	0.162	0.026				
SF31409	2.019	0.202	0.280	0.048	0.233	0.034				
SF31410	0.727	0.099	0.212	0.163	0.201	0.065				
SF31414A	3.216	0.200	0.244	0.017	0.000	0.000				
SF31414B	2.413	0.144	0.233	0.021	0.000	0.000				
SF31418	1.051	0.121	0.773	0.063	0.106	0.030				
SF31421	0.627	0.055	0.095	0.060	0.000	0.000				
SF31422	1.602	0.171	-0.049	0.084	0.297	0.052				
SF31426	1.298	0.140	0.278	0.079	0.210	0.044				
SF31427	1.064	0.128	0.446	0.092	0.196	0.046				
SF31431	0.931	0.201	1.796	0.147	0.124	0.026				
SF31439A	1.230	0.097	1.298	0.057	0.000	0.000				
SF31439B	0.894	0.067	0.494	0.042	0.000	0.000				
SF31440	1.221	0.090	1.089	0.046	0.000	0.000				
SF31441A	1.430	0.089	0.224	0.030	0.000	0.000				
SF31441B	1.101	0.063	0.837	0.028	0.000	0.000	0.465	0.037	-0.465	0.049
SF31442	1.320	0.087	0.668	0.032	0.000	0.000				
SF31443	1.321	0.101	1.248	0.051	0.000	0.000				
SF31445A	1.575	0.098	0.737	0.028	0.000	0.000				
SF31445B	1.492	0.092	-0.001	0.033	0.000	0.000				
SF31446A	1.187	0.083	0.919	0.039	0.000	0.000				
SF31446B	0.993	0.077	1.063	0.051	0.000	0.000				
SF31446C	1.022	0.071	0.433	0.037	0.000	0.000				
SF31447	0.631	0.047	1.241	0.057	0.000	0.000	0.274	0.063	-0.274	0.089

Exhibit D.5 **IRT Parameters for TIMSS 2003 Eighth-Grade Mathematics - Number**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M012001	1.719	0.062	0.278	0.018	0.138	0.009				
M012004	1.370	0.075	0.830	0.029	0.290	0.010				
M012016	1.546	0.111	1.045	0.032	0.405	0.009				
M012027	1.378	0.064	0.465	0.027	0.261	0.011				
M012028	1.087	0.043	-0.127	0.034	0.147	0.016				
M012041	1.198	0.043	0.102	0.025	0.102	0.012				
M022004	1.575	0.080	0.678	0.025	0.282	0.001				
M022010	0.848	0.034	-0.363	0.047	0.090	0.021				
M022012	0.522	0.017	-0.334	0.029	0.000	0.000				
M022043	0.685	0.027	-0.383	0.062	0.111	0.024				
M022046	0.884	0.022	-0.314	0.019	0.000	0.000				
M022057	0.471	0.031	-0.151	0.144	0.192	0.040				
M022066	1.296	0.036	0.162	0.017	0.066	0.008				
M022104	0.863	0.027	-0.459	0.038	0.072	0.017				
M022106	0.944	0.020	0.849	0.018	0.000	0.000				
M022110	0.451	0.016	0.349	0.033	0.000	0.000				
M022127	1.596	0.104	1.390	0.027	0.175	0.007				
M022139	1.389	0.074	1.109	0.025	0.165	0.008				
M022144	0.711	0.055	0.937	0.060	0.259	0.019				
M022156	1.222	0.029	0.348	0.015	0.000	0.000				
M022191	0.939	0.047	0.106	0.048	0.247	0.019				
M022194	0.808	0.039	0.302	0.043	0.122	0.017				
M022198	1.153	0.059	0.738	0.031	0.219	0.011				
M022199	1.349	0.069	0.851	0.026	0.210	0.001				
M022234B	0.817	0.013	1.223	0.015	0.000	0.000	-1.492	0.044	1.492	0.047
M032064	1.094	0.040	0.693	0.026	0.000	0.000				
M032079	1.169	0.069	1.148	0.031	0.189	0.009				
M032094	1.467	0.107	0.406	0.043	0.364	0.017				
M032142	2.439	0.251	1.084	0.033	0.392	0.010				
M032160	1.933	0.154	1.236	0.029	0.158	0.008				
M032166	1.084	0.072	0.160	0.052	0.231	0.022				
M032228	1.539	0.063	0.362	0.022	0.193	0.010				
M032233	1.105	0.032	1.205	0.021	0.000	0.000	-0.433	0.037	0.433	0.045
M032307	1.494	0.032	0.980	0.013	0.000	0.000				
M032352	1.297	0.010	0.519	0.048	0.360	0.017				
M032381	0.976	0.034	0.263	0.024	0.000	0.000				
M032416	1.255	0.072	0.805	0.030	0.080	0.010				
M032447	1.138	0.073	0.665	0.038	0.149	0.015				
M032523	1.956	0.087	1.176	0.016	0.156	0.005				
M032525	0.905	0.038	0.167	0.037	0.102	0.016				
M032529	1.313	0.088	0.949	0.033	0.137	0.011				
M032533	1.718	0.072	0.452	0.020	0.214	0.009				
M032570	1.638	0.080	0.491	0.025	0.321	0.011				
M032609	0.882	0.045	-0.319	0.053	0.073	0.023				
M032612	0.881	0.050	1.023	0.036	0.143	0.012				
M032626	0.812	0.054	0.401	0.055	0.112	0.021				
M032643	1.364	0.067	0.863	0.025	0.195	0.009				
M032652	1.325	0.035	0.957	0.018	0.000	0.000				
M032662	1.459	0.120	1.464	0.039	0.103	0.008				
M032670	0.841	0.046	-1.116	0.093	0.122	0.043				
M032671	0.920	0.023	-0.411	0.019	0.000	0.000				
M032690	0.784	0.067	0.961	0.059	0.153	0.020				

Exhibit D.5 **IRT Parameters for TIMSS 2003 Eighth-Grade Mathematics - Number**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M032701	1.095	0.044	-0.939	0.052	0.149	0.028				
M032704	1.076	0.044	-0.070	0.035	0.150	0.016				
M032725	1.117	0.042	0.920	0.028	0.000	0.000				
M032727	1.619	0.101	0.607	0.030	0.211	0.013				
M032755	0.749	0.022	1.353	0.031	0.000	0.000	-0.657	0.049	0.657	0.062
MC22046	0.805	0.029	-0.578	0.032	0.000	0.000				
MC22110	0.591	0.025	-0.960	0.047	0.000	0.000				
MC32525	0.986	0.055	-0.016	0.049	0.113	0.022				
MC32701	1.311	0.068	-1.081	0.055	0.125	0.032				
MC32704	1.258	0.069	-0.187	0.043	0.168	0.021				
MF12001	1.527	0.076	0.176	0.027	0.113	0.013				
MF12004	1.154	0.076	0.592	0.041	0.192	0.016				
MF12016	0.998	0.085	0.835	0.054	0.262	0.018				
MF12027	1.196	0.066	0.274	0.036	0.133	0.016				
MF12028	1.125	0.061	0.107	0.039	0.115	0.018				
MF12041	1.441	0.069	0.259	0.026	0.081	0.012				
MF22004	0.926	0.067	0.557	0.054	0.192	0.020				
MF22010	0.845	0.051	0.091	0.057	0.111	0.023				
MF22012	0.807	0.029	0.018	0.028	0.000	0.000				
MF22043	0.695	0.038	-0.450	0.076	0.076	0.029				
MF22046	0.912	0.032	-0.322	0.026	0.000	0.000				
MF22057	0.564	0.041	-0.309	0.124	0.118	0.041				
MF22066	1.469	0.068	0.260	0.025	0.068	0.011				
MF22104	0.808	0.050	-0.324	0.077	0.134	0.032				
MF22106	1.074	0.039	0.784	0.027	0.000	0.000				
MF22110	0.592	0.025	0.244	0.036	0.000	0.000				
MF22127	1.302	0.120	1.578	0.048	0.124	0.009				
MF22139	1.494	0.101	1.029	0.031	0.132	0.001				
MF22144	0.724	0.062	0.824	0.067	0.149	0.023				
MF22156	1.583	0.054	0.527	0.018	0.000	0.000				
MF22191	1.071	0.056	0.155	0.038	0.088	0.016				
MF22194	0.943	0.053	0.435	0.039	0.073	0.015				
MF22198	0.963	0.059	0.744	0.039	0.081	0.014				
MF22199	1.160	0.066	0.738	0.032	0.078	0.011				
MF22234B	0.977	0.028	1.213	0.022	0.000	0.000	-1.132	0.061	1.132	0.067
MF32064	1.711	0.059	0.598	0.018	0.000	0.000				
MF32094	1.336	0.075	0.220	0.035	0.179	0.016				
MF32142	2.974	0.331	1.204	0.028	0.378	0.001				
MF32160	2.116	0.159	1.229	0.026	0.124	0.007				
MF32166	1.049	0.065	0.214	0.049	0.175	0.021				
MF32233	1.089	0.033	1.361	0.024	0.000	0.000	-0.463	0.041	0.463	0.051
MF32307	1.498	0.056	0.988	0.024	0.000	0.000				
MF32352	1.272	0.098	0.496	0.047	0.346	0.017				
MF32381	1.136	0.039	0.435	0.023	0.000	0.000				
MF32416	1.053	0.060	0.841	0.034	0.055	0.010				
MF32447	1.481	0.096	0.916	0.031	0.138	0.011				
MF32523	1.733	0.121	1.094	0.029	0.132	0.009				
MF32525	1.079	0.056	0.221	0.036	0.081	0.016				
MF32529	1.824	0.140	1.128	0.030	0.183	0.009				
MF32570	1.476	0.078	0.262	0.030	0.147	0.014				
MF32609	1.055	0.051	-0.390	0.045	0.080	0.021				

Exhibit D.5 **IRT Parameters for TIMSS 2003 Eighth-Grade Mathematics - Number**

(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
MF32626	0.864	0.063	0.693	0.054	0.163	0.020				
MF32643	1.090	0.068	0.734	0.037	0.114	0.014				
MF32662	1.820	0.155	1.502	0.035	0.098	0.007				
MF32670	0.714	0.043	-0.500	0.090	0.111	0.035				
MF32690	1.087	0.088	1.057	0.045	0.192	0.014				
MF32701	1.537	0.068	-0.628	0.031	0.061	0.016				
MF32704	1.358	0.067	0.090	0.030	0.098	0.014				
MF32725	1.204	0.044	0.833	0.026	0.000	0.000				
MF32727	1.674	0.092	0.593	0.026	0.132	0.011				
MF32755	0.843	0.027	1.540	0.033	0.000	0.000	-0.486	0.046	0.486	0.062

Exhibit D.6 **IRT Parameters for TIMSS 2003 Eighth-Grade Mathematics - Measurement**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M012003	1.079	0.035	0.037	0.024	0.042	0.011				
M012013	1.304	0.055	0.328	0.027	0.184	0.012				
M012030	0.997	0.041	0.559	0.027	0.072	0.010				
M012038	0.981	0.046	-0.324	0.056	0.231	0.025				
M022005	0.948	0.069	1.242	0.043	0.251	0.012				
M022021	1.373	0.058	0.618	0.022	0.138	0.009				
M022055	1.339	0.026	0.600	0.012	0.000	0.000				
M022097	1.009	0.029	-0.203	0.027	0.064	0.013				
M022148	0.910	0.023	0.067	0.018	0.000	0.000				
M022188	0.741	0.054	1.023	0.055	0.239	0.017				
M022227A	1.300	0.030	-0.140	0.014	0.000	0.000				
M022227B	2.013	0.050	0.572	0.012	0.000	0.000				
M022227C	1.530	0.040	0.947	0.017	0.000	0.000				
M022232	0.543	0.009	1.722	0.025	0.000	0.000	-2.064	0.058	2.064	0.066
M022234A	0.760	0.011	0.932	0.013	0.000	0.000	-0.639	0.026	0.639	0.030
M022243	1.246	0.025	0.654	0.013	0.000	0.000				
M032097	0.878	0.081	1.412	0.058	0.144	0.015				
M032100	1.089	0.058	0.167	0.040	0.094	0.018				
M032116	0.693	0.067	0.903	0.082	0.203	0.027				
M032324	1.081	0.075	0.813	0.042	0.155	0.015				
M032331	2.999	0.292	1.351	0.026	0.221	0.008				
M032344	1.007	0.036	0.655	0.027	0.000	0.000				
M032575	1.419	0.079	0.416	0.031	0.138	0.014				
M032623	1.494	0.091	0.760	0.029	0.137	0.011				
M032647	0.801	0.064	1.276	0.054	0.284	0.015				
M032649A	0.989	0.024	0.535	0.018	0.000	0.000				
M032649B	1.246	0.034	1.216	0.021	0.000	0.000				
M032678	1.165	0.040	0.490	0.021	0.045	0.008				
M032699	0.698	0.038	-0.840	0.119	0.209	0.048				
M032732	0.759	0.062	0.163	0.092	0.214	0.034				
M032754	0.890	0.032	-0.397	0.027	0.000	0.000				
MF12003	1.264	0.062	0.152	0.031	0.078	0.014				
MF12013	1.147	0.063	0.262	0.038	0.109	0.017				
MF12030	1.414	0.085	0.771	0.030	0.128	0.011				
MF12038	1.176	0.062	-0.185	0.044	0.124	0.022				
MF22005	0.771	0.090	1.511	0.079	0.252	0.020				
MF22021	1.398	0.090	0.897	0.033	0.159	0.012				
MF22055	1.338	0.045	0.587	0.021	0.000	0.000				
MF22097	1.928	0.103	-0.109	0.027	0.182	0.016				
MF22148	1.228	0.041	0.384	0.021	0.000	0.000				
MF22188	0.745	0.071	1.141	0.069	0.178	0.021				
MF22227A	1.561	0.051	0.337	0.018	0.000	0.000				
MF22227B	2.641	0.101	0.811	0.015	0.000	0.000				
MF22227C	1.922	0.075	1.118	0.021	0.000	0.000				
MF22232	0.540	0.017	1.770	0.046	0.000	0.000	-2.413	0.118	2.413	0.130
MF22234A	0.891	0.023	0.935	0.021	0.000	0.000	-0.502	0.039	0.502	0.046
MF22243	1.304	0.044	0.623	0.022	0.000	0.000				
MF32097	0.895	0.086	1.583	0.064	0.136	0.013				
MF32100	0.841	0.054	0.525	0.049	0.087	0.019				
MF32116	1.400	0.107	0.845	0.038	0.245	0.013				
MF32324	1.047	0.075	1.089	0.042	0.100	0.012				
MF32331	1.976	0.218	1.607	0.041	0.210	0.008				

Exhibit D.6 **IRT Parameters for TIMSS2003 Eighth-Grade Mathematics - Measurement**

(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
MF32344	1.450	0.050	0.676	0.021	0.000	0.000				
MF32575	1.629	0.010	0.604	0.029	0.177	0.012				
MF32623	1.426	0.084	0.724	0.029	0.123	0.011				
MF32732	0.615	0.061	0.795	0.103	0.198	0.032				
MF32754	0.760	0.028	-0.089	0.029	0.000	0.000				

Exhibit D.7 **IRT Parameters for TIMSS 2003 Eighth-Grade Mathematics - Data**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M012006	0.524	0.031	-0.715	0.145	0.148	0.047				
M012014	0.723	0.034	-0.793	0.084	0.142	0.034				
M012037	0.498	0.032	0.315	0.092	0.095	0.029				
M022101	0.859	0.029	-0.340	0.041	0.119	0.018				
M022135	0.729	0.035	0.857	0.034	0.048	0.011				
M022146	1.223	0.047	0.238	0.026	0.144	0.012				
M022181	1.176	0.038	-0.700	0.035	0.181	0.018				
M022189	0.760	0.029	-0.709	0.055	0.067	0.023				
M022252	1.181	0.058	0.276	0.037	0.314	0.015				
M022257	1.106	0.053	0.792	0.028	0.267	0.001				
M032132	0.815	0.054	0.366	0.058	0.123	0.022				
M032271	1.500	0.063	0.474	0.023	0.204	0.001				
M032507	2.710	0.237	1.193	0.023	0.209	0.008				
M032595	1.533	0.093	0.595	0.032	0.194	0.014				
M032637C	1.408	0.046	0.387	0.019	0.000	0.000				
M032681A	0.707	0.027	-0.398	0.033	0.000	0.000				
M032681B	0.755	0.030	0.773	0.035	0.000	0.000				
M032681C	1.974	0.065	0.497	0.016	0.000	0.000				
M032688	0.898	0.033	0.676	0.029	0.000	0.000				
M032695	0.609	0.015	-0.226	0.022	0.000	0.000	-0.951	0.055	0.951	0.053
M032721	0.619	0.085	1.470	0.095	0.228	0.025				
M032753A	3.430	0.118	0.741	0.009	0.000	0.000	0.124	0.014	-0.124	0.014
M032753B	2.991	0.103	0.883	0.001	0.000	0.000	0.189	0.014	-0.189	0.015
M032753C	1.405	0.046	0.459	0.019	0.000	0.000				
M032756	0.965	0.033	0.335	0.025	0.000	0.000				
M032762	0.513	0.009	1.010	0.021	0.000	0.000	-1.843	0.056	1.843	0.061
M032763	1.747	0.047	1.438	0.013	0.000	0.000	-0.152	0.020	0.152	0.026
M032764	1.460	0.037	1.382	0.014	0.000	0.000	-0.030	0.019	0.030	0.026
MF12006	0.598	0.052	-0.067	0.139	0.202	0.044				
MF12014	1.021	0.059	-0.382	0.061	0.174	0.028				
MF12037	0.511	0.045	0.540	0.106	0.101	0.033				
MF22101	0.730	0.043	-0.402	0.084	0.109	0.033				
MF22135	0.809	0.051	0.944	0.043	0.044	0.012				
MF22146	0.969	0.054	0.530	0.037	0.070	0.014				
MF22181	0.936	0.055	-0.504	0.073	0.172	0.032				
MF22189	1.084	0.055	-0.103	0.043	0.105	0.019				
MF22252	0.921	0.064	0.371	0.059	0.208	0.022				
MF22257	0.988	0.068	0.599	0.047	0.175	0.018				
MF32132	1.695	0.117	0.952	0.028	0.170	0.010				
MF32507	2.668	0.213	1.201	0.021	0.163	0.008				
MF32595	1.112	0.065	0.525	0.038	0.124	0.016				
MF32637C	1.270	0.043	0.615	0.022	0.000	0.000				
MF32681A	0.775	0.028	0.078	0.029	0.000	0.000				
MF32681B	0.878	0.034	0.929	0.033	0.000	0.000				
MF32681C	1.632	0.054	0.643	0.018	0.000	0.000				
MF32688	1.138	0.041	0.860	0.027	0.000	0.000				
MF32695	0.608	0.014	0.506	0.022	0.000	0.000	-1.157	0.057	1.157	0.060
MF32721	0.666	0.089	1.492	0.089	0.227	0.023				
MF32753A	3.170	0.118	0.958	0.001	0.000	0.000	0.100	0.015	-0.100	0.016
MF32753B	4.421	0.195	1.060	0.008	0.000	0.000	0.162	0.012	-0.162	0.014
MF32753C	1.463	0.052	0.800	0.022	0.000	0.000				
MF32756	0.961	0.036	0.742	0.029	0.000	0.000				

Exhibit D.8 **IRT Parameters for TIMSS 2003 Eighth-Grade Mathematics - Geometry**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M012005	1.027	0.052	0.273	0.041	0.238	0.017				
M012015	0.947	0.039	-0.187	0.042	0.118	0.019				
M012026	1.380	0.064	0.596	0.024	0.206	0.010				
M012039	1.256	0.057	0.394	0.028	0.201	0.012				
M022016	0.650	0.050	1.258	0.056	0.154	0.017				
M022049	0.665	0.045	0.462	0.076	0.351	0.022				
M022062	1.052	0.040	0.735	0.022	0.110	0.008				
M022105	0.705	0.034	0.757	0.038	0.105	0.014				
M022108	0.794	0.033	0.065	0.046	0.160	0.018				
M022142	1.517	0.070	0.639	0.022	0.201	0.009				
M022154	1.045	0.052	0.499	0.034	0.194	0.014				
M022202	0.761	0.022	0.892	0.027	0.000	0.000				
M032205	0.620	0.049	-0.066	0.117	0.151	0.041				
M032261	0.971	0.048	0.633	0.032	0.141	0.013				
M032294	0.845	0.053	-0.218	0.074	0.143	0.031				
M032397	1.422	0.106	0.837	0.035	0.226	0.013				
M032398	1.712	0.133	0.881	0.032	0.263	0.012				
M032402	0.632	0.067	0.908	0.093	0.199	0.030				
M032403	0.958	0.024	-0.121	0.017	0.000	0.000				
M032414	1.400	0.048	0.550	0.020	0.000	0.000				
M032489	0.949	0.046	-0.384	0.059	0.238	0.026				
M032579	1.232	0.037	-0.151	0.024	0.118	0.012				
M032588	0.934	0.044	0.027	0.044	0.169	0.019				
M032679	1.139	0.077	0.367	0.046	0.226	0.019				
M032689	0.841	0.071	1.219	0.054	0.327	0.014				
M032691	0.987	0.020	0.323	0.014	0.000	0.000				
M032692	0.677	0.018	1.090	0.027	0.000	0.000	-1.162	0.059	1.162	0.067
M032693	0.839	0.023	0.628	0.022	0.000	0.000				
M032734	0.872	0.031	-0.339	0.027	0.000	0.000				
M032743	0.674	0.027	0.029	0.032	0.000	0.000				
M032745	0.567	0.023	2.318	0.079	0.000	0.000	-1.079	0.083	1.079	0.122
MF12005	0.908	0.055	0.162	0.053	0.112	0.022				
MF12015	1.116	0.061	0.098	0.041	0.115	0.019				
MF12026	1.123	0.066	0.499	0.036	0.110	0.015				
MF12039	1.147	0.060	0.279	0.033	0.080	0.014				
MF22016	0.814	0.080	1.333	0.064	0.167	0.017				
MF22049	0.533	0.049	-0.085	0.162	0.172	0.050				
MF22062	1.336	0.076	0.679	0.028	0.091	0.011				
MF22105	0.689	0.056	0.859	0.063	0.101	0.022				
MF22108	0.891	0.053	0.027	0.056	0.109	0.024				
MF22142	1.647	0.091	0.568	0.025	0.115	0.011				
MF22154	1.274	0.081	0.643	0.034	0.154	0.014				
MF22202	0.859	0.035	1.024	0.038	0.000	0.000				
MF32205	0.580	0.063	0.815	0.108	0.194	0.034				
MF32294	1.667	0.101	0.397	0.030	0.222	0.014				
MF32397	1.233	0.091	0.981	0.037	0.155	0.013				
MF32398	1.741	0.131	0.930	0.030	0.226	0.011				
MF32402	0.931	0.091	1.138	0.058	0.233	0.017				
MF32414	1.426	0.050	0.710	0.021	0.000	0.000				
MF32579	1.394	0.070	0.108	0.030	0.109	0.015				
MF32679	1.424	0.093	0.532	0.034	0.220	0.015				
MF32691	1.238	0.042	0.454	0.021	0.000	0.000				

Exhibit D.8 **IRT Parameters for TIMSS 2003 Eighth-Grade Mathematics - Geometry**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
MF32692	0.754	0.021	1.241	0.027	0.000	0.000	-1.232	0.063	1.232	0.071
MF32693	0.907	0.034	0.617	0.029	0.000	0.000				
MF32734	0.962	0.034	0.260	0.024	0.000	0.000				
MF32743	0.669	0.027	0.309	0.034	0.000	0.000				
MF32745	0.655	0.027	2.295	0.073	0.000	0.000	-1.041	0.084	1.041	0.122

Exhibit D.9 **IRT Parameters for TIMSS 2003 Eighth-Grade Mathematics - Algebra**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M012002	0.637	0.034	-0.402	0.088	0.148	0.031				
M012017	0.681	0.036	0.325	0.053	0.121	0.019				
M012025	0.786	0.038	-0.283	0.061	0.167	0.024				
M012029	1.089	0.046	0.244	0.029	0.137	0.012				
M012040	1.127	0.049	-0.198	0.039	0.237	0.017				
M012042	1.255	0.051	0.318	0.025	0.152	0.011				
M022002	1.594	0.111	1.468	0.028	0.155	0.006				
M022008	0.586	0.019	0.960	0.034	0.000	0.000				
M022050	0.870	0.039	1.139	0.028	0.103	0.008				
M022185	0.870	0.049	0.394	0.047	0.253	0.017				
M022196	1.305	0.051	-0.027	0.026	0.158	0.012				
M022251	0.875	0.063	1.495	0.047	0.164	0.010				
M022253	1.146	0.028	0.129	0.015	0.000	0.000				
M022261A	1.332	0.032	0.473	0.014	0.000	0.000				
M022261B	1.744	0.046	0.927	0.014	0.000	0.000				
M022261C	0.977	0.020	1.197	0.015	0.000	0.000	-1.324	0.049	1.324	0.052
M032036	0.824	0.060	0.383	0.061	0.171	0.023				
M032044	0.619	0.037	0.583	0.059	0.127	0.020				
M032046	0.811	0.046	1.324	0.038	0.077	0.009				
M032047	1.232	0.146	1.173	0.058	0.428	0.015				
M032163	1.402	0.086	0.514	0.033	0.191	0.014				
M032198	0.977	0.075	0.682	0.051	0.225	0.018				
M032208	1.318	0.059	0.399	0.027	0.228	0.011				
M032210	1.424	0.065	0.721	0.023	0.175	0.009				
M032273	1.101	0.078	0.023	0.061	0.335	0.023				
M032295	1.624	0.090	-0.402	0.038	0.218	0.020				
M032419	1.392	0.108	0.887	0.038	0.259	0.013				
M032424	0.802	0.052	0.401	0.053	0.109	0.020				
M032477	1.121	0.069	0.449	0.040	0.163	0.016				
M032538	1.281	0.044	0.276	0.020	0.000	0.000				
M032540	1.260	0.121	0.746	0.055	0.448	0.017				
M032545	0.965	0.027	0.998	0.023	0.000	0.000				
M032557	1.220	0.033	0.962	0.019	0.000	0.000				
M032637A	1.631	0.054	-0.225	0.018	0.000	0.000				
M032637B	2.091	0.071	-0.025	0.015	0.000	0.000				
M032640	0.496	0.017	1.787	0.056	0.000	0.000	-0.980	0.069	0.980	0.093
M032673	1.227	0.084	0.778	0.037	0.194	0.015				
M032683	0.667	0.017	0.847	0.024	0.000	0.000	-0.921	0.052	0.921	0.058
M032698	1.376	0.092	0.677	0.035	0.204	0.014				
M032728	1.087	0.097	1.022	0.049	0.250	0.015				
M032738	1.454	0.080	-0.203	0.038	0.204	0.019				
M032744	0.872	0.034	0.645	0.030	0.000	0.000				
M032757	0.596	0.014	-0.016	0.022	0.000	0.000	-1.612	0.069	1.612	0.068
M032760A	1.115	0.028	0.818	0.015	0.000	0.000	-0.897	0.046	0.897	0.048
M032760B	2.007	0.077	1.069	0.018	0.000	0.000				
M032760C	2.366	0.109	1.304	0.019	0.000	0.000				
M032761	1.214	0.041	1.401	0.022	0.000	0.000	-0.281	0.034	0.281	0.043
MF12002	0.685	0.045	-0.394	0.097	0.130	0.035				

Exhibit D.9 **IRT Parameters for TIMSS 2003 Eighth-Grade Mathematics - Algebra**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
MF12017	0.814	0.052	0.506	0.049	0.093	0.018				
MF12025	0.828	0.051	-0.141	0.067	0.135	0.026				
MF12029	0.999	0.050	0.244	0.035	0.054	0.013				
MF12040	1.295	0.080	0.159	0.041	0.250	0.018				
MF12042	1.669	0.085	0.456	0.024	0.100	0.010				
MF22002	0.859	0.078	1.677	0.068	0.087	0.011				
MF22008	0.736	0.032	1.126	0.044	0.000	0.000				
MF22050	0.766	0.074	1.370	0.066	0.154	0.016				
MF22185	1.002	0.065	0.360	0.047	0.172	0.019				
MF22196	1.482	0.067	0.161	0.024	0.058	0.001				
MF22251	0.936	0.091	1.579	0.065	0.134	0.012				
MF22253	1.253	0.043	0.402	0.021	0.000	0.000				
MF22261A	2.072	0.073	0.649	0.016	0.000	0.000				
MF22261B	4.886	0.233	0.982	0.011	0.000	0.000				
MF22261C	2.683	0.102	1.206	0.011	0.000	0.000	-0.219	0.025	0.219	0.026
MF32036	2.662	0.186	0.808	0.023	0.249	0.010				
MF32047	2.049	0.293	1.458	0.042	0.404	0.011				
MF32163	1.484	0.141	1.237	0.040	0.242	0.011				
MF32198	1.145	0.071	0.527	0.038	0.154	0.015				
MF32273	1.448	0.091	0.049	0.040	0.289	0.018				
MF32295	0.947	0.049	-0.362	0.056	0.099	0.024				
MF32419	1.259	0.107	1.105	0.042	0.222	0.013				
MF32424	0.911	0.069	0.946	0.047	0.151	0.016				
MF32477	1.197	0.080	0.764	0.037	0.157	0.013				
MF32538	1.098	0.040	0.540	0.024	0.000	0.000				
MF32540	1.065	0.082	0.440	0.054	0.300	0.020				
MF32637A	1.187	0.040	0.194	0.021	0.000	0.000				
MF32637B	1.277	0.043	0.198	0.020	0.000	0.000				
MF32640	0.521	0.017	1.597	0.049	0.000	0.000	-0.752	0.062	0.752	0.083
MF32673	1.287	0.082	0.638	0.035	0.175	0.014				
MF32683	0.539	0.014	1.074	0.029	0.000	0.000	-1.476	0.069	1.476	0.076
MF32698	1.173	0.070	0.618	0.034	0.110	0.013				
MF32728	2.852	0.260	1.158	0.023	0.220	0.009				
MF32738	0.874	0.049	-0.148	0.060	0.109	0.024				
MF32744	0.702	0.030	1.076	0.045	0.000	0.000				
MF32757	0.576	0.013	0.109	0.022	0.000	0.000	-1.939	0.078	1.939	0.078
MF32760A	1.262	0.033	0.848	0.016	0.000	0.000	-0.747	0.042	0.747	0.045
MF32760B	2.401	0.103	1.140	0.019	0.000	0.000				
MF32760C	2.822	0.148	1.330	0.019	0.000	0.000				
MF32761	1.583	0.057	1.358	0.019	0.000	0.000	-0.201	0.031	0.201	0.039

Exhibit D.10 **IRT Parameters for TIMSS 2003 Eighth-Grade Science - Chemistry**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S012003	0.919	0.044	-0.504	0.062	0.206	0.027				
S012016	0.697	0.054	-0.254	0.121	0.398	0.034				
S012025	1.227	0.130	1.407	0.054	0.405	0.001				
S012040	1.567	0.079	0.496	0.023	0.275	0.010				
S022181	0.927	0.054	0.883	0.033	0.252	0.011				
S022183	1.948	0.098	0.983	0.018	0.262	0.006				
S022187	0.541	0.042	0.813	0.068	0.104	0.024				
S022188	1.185	0.124	1.293	0.052	0.441	0.010				
S022191	0.606	0.012	-0.344	0.016	0.000	0.000	-0.347	0.034	0.347	0.031
S022198	1.180	0.091	1.277	0.038	0.236	0.009				
S022202	1.242	0.080	0.814	0.031	0.302	0.011				
S022206	1.195	0.089	1.082	0.035	0.288	0.010				
S022208	1.650	0.089	1.005	0.021	0.298	0.007				
S022276	0.797	0.046	0.468	0.047	0.271	0.016				
S032056	0.697	0.029	0.541	0.035	0.000	0.000				
S032057	1.158	0.034	1.003	0.022	0.000	0.000				
S032156	0.710	0.060	0.668	0.062	0.112	0.023				
S032502	1.802	0.134	0.824	0.028	0.220	0.011				
S032562	0.631	0.017	0.264	0.021	0.000	0.000	-0.609	0.046	0.609	0.048
S032564	1.482	0.101	1.231	0.030	0.198	0.008				
S032565	0.792	0.035	0.909	0.040	0.000	0.000				
S032570	0.597	0.029	0.978	0.052	0.000	0.000				
S032574	1.708	0.155	0.841	0.035	0.347	0.013				
S032579	0.641	0.095	1.422	0.104	0.256	0.025				
S032672	0.344	0.044	-0.274	0.363	0.211	0.077				
S032679	0.814	0.041	1.444	0.058	0.000	0.000				
S032680	0.625	0.018	-0.356	0.023	0.000	0.000	-0.047	0.045	0.047	0.040
S032683	0.914	0.049	0.851	0.031	0.206	0.011				
S032709	2.707	0.075	0.730	0.001	0.000	0.000				
S032713A	1.799	0.053	1.001	0.016	0.000	0.000				
S032713B	1.026	0.042	1.776	0.050	0.000	0.000				
SF12003	1.477	0.073	-0.130	0.031	0.114	0.016				
SF12016	1.127	0.065	-0.375	0.056	0.178	0.026				
SF12025	0.751	0.083	0.971	0.070	0.226	0.023				
SF12040	0.628	0.054	0.451	0.083	0.119	0.030				
SF22181	1.496	0.110	0.766	0.031	0.209	0.013				
SF22183	1.231	0.087	0.783	0.034	0.140	0.013				
SF22187	0.798	0.064	0.826	0.049	0.096	0.018				
SF22188	1.051	0.098	0.833	0.050	0.273	0.017				
SF22191	0.682	0.018	0.213	0.020	0.000	0.000	-0.372	0.041	0.372	0.042
SF22198	0.489	0.067	1.661	0.119	0.111	0.027				
SF22202	0.520	0.058	0.953	0.099	0.121	0.032				
SF22206	1.264	0.088	0.767	0.033	0.143	0.013				
SF22208	1.248	0.010	0.860	0.038	0.200	0.014				
SF22276	1.036	0.095	0.754	0.051	0.272	0.018				
SF32056	0.951	0.038	0.711	0.030	0.000	0.000				
SF32156	1.184	0.094	0.795	0.038	0.188	0.014				
SF32502	0.881	0.059	0.697	0.040	0.070	0.015				
SF32562	0.717	0.018	0.358	0.019	0.000	0.000	-0.551	0.042	0.551	0.044
SF32565	0.662	0.034	1.311	0.062	0.000	0.000				
SF32570	1.014	0.042	0.904	0.032	0.000	0.000				
SF32574	1.075	0.097	0.741	0.051	0.293	0.018				

Exhibit D.10 **IRT Parameters for TIMSS 2003 Eighth-Grade Science - Chemistry**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
SF32579	1.577	0.155	1.104	0.040	0.268	0.012				
SF32672	0.654	0.082	0.619	0.116	0.347	0.033				
SF32679	0.849	0.041	1.345	0.051	0.000	0.000				
SF32680	0.664	0.018	-0.232	0.021	0.000	0.000	-0.240	0.043	0.240	0.040
SF32683	1.301	0.089	0.783	0.032	0.133	0.012				

Exhibit D.11 **IRT Parameters for TIMSS 2003 Eighth-Grade Science - Physics**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S012002	0.554	0.039	-0.102	0.104	0.212	0.031				
S012004	0.635	0.047	-0.038	0.096	0.324	0.028				
S012015	0.993	0.048	-0.083	0.044	0.265	0.018				
S012029	0.644	0.054	0.426	0.082	0.338	0.024				
S012037	0.654	0.033	-1.707	0.135	0.171	0.049				
S022002	1.041	0.046	0.429	0.028	0.229	0.012				
S022019	0.877	0.039	-0.219	0.049	0.301	0.018				
S022022	0.732	0.018	0.116	0.017	0.000	0.000				
S022035	0.387	0.016	0.242	0.038	0.000	0.000				
S022040	0.689	0.030	-0.286	0.051	0.071	0.018				
S022041	0.740	0.035	-0.732	0.073	0.163	0.027				
S022042	1.301	0.046	0.325	0.020	0.175	0.001				
S022054	1.214	0.052	0.463	0.024	0.251	0.011				
S022058	0.903	0.059	0.251	0.054	0.363	0.018				
S022069	0.996	0.022	0.432	0.014	0.000	0.000				
S022222	1.054	0.056	0.634	0.029	0.173	0.012				
S022225	1.264	0.109	1.527	0.049	0.121	0.007				
S022268	0.658	0.018	0.612	0.022	0.000	0.000				
S022279	0.721	0.022	0.203	0.021	0.000	0.000				
S022281	0.483	0.017	1.256	0.044	0.000	0.000				
S022286	0.786	0.032	1.603	0.052	0.000	0.000				
S022292	0.822	0.019	0.334	0.016	0.000	0.000				
S032024	1.759	0.214	1.232	0.042	0.291	0.011				
S032055	0.717	0.036	-1.483	0.113	0.181	0.042				
S032131	0.935	0.025	-0.094	0.017	0.000	0.000				
S032141	2.207	0.190	0.985	0.027	0.231	0.010				
S032158	0.983	0.101	0.620	0.065	0.385	0.022				
S032184	0.785	0.139	1.551	0.121	0.349	0.020				
S032238	1.188	0.073	0.485	0.033	0.132	0.015				
S032257	0.988	0.101	1.034	0.053	0.208	0.017				
S032272	0.761	0.043	1.555	0.071	0.000	0.000				
S032273	0.742	0.130	1.738	0.139	0.258	0.020				
S032279	0.785	0.085	1.235	0.068	0.150	0.018				
S032281	1.021	0.042	-0.031	0.033	0.145	0.015				
S032369	0.658	0.022	0.705	0.025	0.000	0.000	-0.123	0.039	0.123	0.046
S032375	0.505	0.015	0.834	0.030	0.000	0.000	-1.168	0.062	1.168	0.070
S032392	0.462	0.040	-1.604	0.275	0.206	0.074				
S032394	1.451	0.137	0.769	0.041	0.371	0.015				
S032403	1.281	0.122	0.921	0.042	0.265	0.015				
S032425	1.438	0.120	0.740	0.037	0.286	0.015				
S032625A	1.505	0.037	0.305	0.012	0.000	0.000				
S032625B	1.858	0.047	0.576	0.011	0.000	0.000				
S032626	0.950	0.026	0.306	0.017	0.000	0.000				
S032711	1.003	0.021	0.946	0.014	0.000	0.000	-0.368	0.024	0.368	0.029
S032712A	1.261	0.033	0.584	0.015	0.000	0.000				
S032712B	1.745	0.053	1.056	0.017	0.000	0.000				
SF12002	0.868	0.044	-0.206	0.043	0.057	0.016				
SF12004	1.004	0.049	0.019	0.034	0.057	0.014				

Exhibit D.11 **IRT Parameters for TIMSS 2003 Eighth-Grade Science - Physics**

(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
SF12015	0.803	0.045	-0.239	0.055	0.080	0.021				
SF12029	0.782	0.068	0.570	0.062	0.195	0.023				
SF12037	0.615	0.043	-0.925	0.129	0.140	0.042				
SF22002	1.236	0.066	0.389	0.029	0.089	0.013				
SF22019	1.168	0.056	-0.118	0.034	0.085	0.015				
SF22022	1.078	0.041	0.452	0.023	0.000	0.000				
SF22035	0.574	0.028	0.534	0.041	0.000	0.000				
SF22040	1.087	0.065	0.450	0.033	0.096	0.014				
SF22041	0.753	0.043	0.025	0.049	0.063	0.018				
SF22042	1.421	0.071	0.423	0.024	0.071	0.011				
SF22054	1.438	0.095	0.518	0.031	0.197	0.015				
SF22058	1.054	0.067	0.219	0.045	0.183	0.020				
SF22069	1.532	0.056	0.651	0.019	0.000	0.000				
SF22222	1.024	0.069	0.861	0.036	0.069	0.012				
SF22225	1.137	0.110	1.415	0.058	0.080	0.001				
SF22268	1.068	0.045	0.881	0.030	0.000	0.000				
SF22279	0.831	0.036	0.692	0.032	0.000	0.000				
SF22281	0.802	0.040	1.154	0.048	0.000	0.000				
SF22286	1.182	0.072	1.685	0.067	0.000	0.000				
SF22292	0.920	0.037	0.471	0.026	0.000	0.000				
SF32024	1.479	0.170	1.248	0.046	0.252	0.012				
SF32131	1.404	0.049	0.273	0.018	0.000	0.000				
SF32141	1.279	0.127	1.158	0.044	0.188	0.013				
SF32158	0.936	0.087	0.593	0.059	0.289	0.022				
SF32184	0.523	0.072	1.397	0.109	0.163	0.028				
SF32238	1.014	0.067	0.595	0.037	0.107	0.015				
SF32257	1.873	0.147	0.832	0.028	0.227	0.012				
SF32272	0.938	0.050	1.507	0.059	0.000	0.000				
SF32273	0.872	0.158	1.738	0.136	0.281	0.017				
SF32279	0.765	0.095	1.420	0.084	0.157	0.018				
SF32369	0.619	0.020	0.714	0.026	0.000	0.000	-0.286	0.042	0.286	0.050
SF32375	0.537	0.015	0.889	0.029	0.000	0.000	-1.175	0.060	1.175	0.068
SF32392	0.585	0.052	-0.817	0.180	0.270	0.052				
SF32394	0.880	0.086	0.767	0.060	0.261	0.021				
SF32403	1.324	0.130	0.980	0.042	0.267	0.014				
SF32425	1.015	0.092	0.797	0.049	0.233	0.018				
SF32625A	3.087	0.118	0.636	0.012	0.000	0.000				
SF32625B	4.344	0.189	0.842	0.011	0.000	0.000				

Exhibit D.12 **IRT Parameters for TIMSS 2003 Eighth-Grade Science - Life Science**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S012001	0.587	0.036	-0.138	0.087	0.165	0.028				
S012014	0.994	0.044	-0.502	0.049	0.236	0.020				
S012026	1.023	0.074	0.105	0.066	0.562	0.017				
S012028	0.761	0.040	0.481	0.039	0.113	0.015				
S012038	0.950	0.065	0.521	0.048	0.374	0.016				
S012039	0.904	0.056	-0.161	0.069	0.440	0.021				
S022106	0.799	0.061	1.599	0.053	0.124	0.009				
S022115	0.927	0.041	0.255	0.036	0.262	0.014				
S022117	0.727	0.047	0.693	0.047	0.176	0.017				
S022126	0.492	0.032	0.306	0.083	0.143	0.025				
S022150	0.836	0.044	0.649	0.036	0.225	0.013				
S022152	0.939	0.025	0.256	0.017	0.000	0.000				
S022154	0.634	0.019	-0.197	0.024	0.000	0.000				
S022160	0.635	0.021	0.693	0.028	0.000	0.000				
S022161	0.593	0.020	0.500	0.027	0.000	0.000				
S022235	0.809	0.076	0.842	0.062	0.422	0.017				
S022289	0.724	0.014	0.907	0.015	0.000	0.000	0.752	0.018	-0.752	0.028
S032007	0.851	0.033	0.375	0.027	0.000	0.000				
S032008	0.870	0.048	0.140	0.051	0.273	0.019				
S032015	0.835	0.035	0.734	0.032	0.000	0.000				
S032035	1.283	0.045	0.482	0.019	0.158	0.009				
S032083	0.933	0.066	1.263	0.040	0.121	0.010				
S032087	0.771	0.082	0.973	0.066	0.224	0.022				
S032202	0.607	0.014	-0.002	0.016	0.000	0.000	0.265	0.030	-0.265	0.029
S032206	1.040	0.034	1.175	0.027	0.000	0.000				
S032258	0.799	0.033	-0.069	0.043	0.176	0.016				
S032306	0.496	0.013	0.523	0.026	0.000	0.000	-1.415	0.066	1.415	0.070
S032310D	0.567	0.018	0.024	0.024	0.000	0.000	-0.065	0.046	0.065	0.045
S032315	0.911	0.080	0.609	0.058	0.268	0.021				
S032385	0.741	0.036	-0.043	0.056	0.260	0.019				
S032386	0.891	0.085	1.564	0.061	0.173	0.011				
S032451	0.627	0.015	0.046	0.020	0.000	0.000	-1.183	0.056	1.183	0.056
S032465	0.679	0.057	-0.188	0.110	0.241	0.036				
S032530D	0.506	0.019	0.443	0.030	0.000	0.000	0.715	0.045	-0.715	0.052
S032542	1.472	0.134	0.855	0.038	0.320	0.014				
S032595	1.359	0.122	1.480	0.043	0.183	0.008				
S032606	1.190	0.089	-0.280	0.070	0.452	0.025				
S032607	0.692	0.039	-0.171	0.071	0.192	0.025				
S032611	1.041	0.129	1.367	0.065	0.232	0.015				
S032614	0.788	0.031	0.116	0.028	0.000	0.000				
S032637	1.078	0.079	1.113	0.035	0.230	0.011				
S032640	0.551	0.025	0.015	0.038	0.000	0.000				
S032645	0.716	0.078	0.956	0.072	0.225	0.023				
S032682	1.510	0.095	0.997	0.024	0.239	0.009				
S032693A	1.068	0.038	0.184	0.022	0.000	0.000				
S032693B	0.922	0.031	0.782	0.021	0.000	0.000	0.598	0.026	-0.598	0.037
S032695	0.885	0.028	0.749	0.020	0.000	0.000	-0.021	0.031	0.021	0.038
S032697D	0.916	0.027	0.628	0.018	0.000	0.000	-0.051	0.030	0.051	0.034
S032704	0.945	0.039	0.707	0.029	0.000	0.000				
S032705A	1.383	0.049	0.453	0.019	0.000	0.000				
S032705B	1.400	0.048	0.185	0.018	0.000	0.000				
S032706A	1.144	0.044	0.664	0.024	0.000	0.000				

Exhibit D.12 **IRT Parameters for TIMSS Eighth-Grade Science - Life Science**

(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S032706B	1.340	0.051	0.731	0.022	0.000	0.000				
S032707	1.600	0.078	1.307	0.032	0.000	0.000				
SF12001	0.856	0.052	0.081	0.051	0.116	0.021				
SF12014	1.212	0.063	-0.237	0.041	0.154	0.019				
SF12026	0.974	0.070	-0.313	0.078	0.345	0.029				
SF12028	1.069	0.064	0.511	0.035	0.106	0.014				
SF12038	0.798	0.051	0.062	0.057	0.117	0.022				
SF12039	1.090	0.056	-0.296	0.044	0.126	0.020				
SF22106	1.219	0.124	1.404	0.054	0.108	0.010				
SF22115	1.130	0.072	0.396	0.039	0.179	0.017				
SF22117	0.989	0.079	0.776	0.044	0.179	0.017				
SF22126	0.861	0.058	0.453	0.047	0.119	0.018				
SF22150	1.436	0.097	0.727	0.030	0.173	0.013				
SF22152	1.496	0.052	0.487	0.018	0.000	0.000				
SF22154	0.972	0.036	0.379	0.024	0.000	0.000				
SF22160	0.897	0.038	0.797	0.032	0.000	0.000				
SF22161	0.553	0.029	1.053	0.057	0.000	0.000				
SF22235	0.896	0.087	0.805	0.059	0.264	0.021				
SF22289	1.068	0.036	1.008	0.021	0.000	0.000	0.515	0.023	-0.515	0.038
SF32007	0.942	0.036	0.391	0.025	0.000	0.000				
SF32015	0.991	0.040	0.825	0.030	0.000	0.000				
SF32035	1.193	0.068	0.517	0.030	0.098	0.013				
SF32087	0.781	0.077	0.956	0.060	0.184	0.021				
SF32202	0.878	0.027	0.467	0.018	0.000	0.000	0.074	0.030	-0.074	0.033
SF32258	1.073	0.065	0.310	0.040	0.150	0.017				
SF32306	0.555	0.015	0.564	0.024	0.000	0.000	-1.102	0.057	1.102	0.061
SF32310D	0.504	0.016	0.021	0.025	0.000	0.000	-0.277	0.052	0.277	0.051
SF32315	0.888	0.066	0.446	0.053	0.189	0.021				
SF32385	1.294	0.076	0.229	0.037	0.200	0.017				
SF32451	0.750	0.018	0.217	0.018	0.000	0.000	-0.888	0.047	0.888	0.047
SF32465	0.861	0.062	-0.074	0.072	0.236	0.027				
SF32530D	0.556	0.020	0.479	0.028	0.000	0.000	0.662	0.042	-0.662	0.049
SF32542	1.341	0.099	0.645	0.036	0.249	0.015				
SF32595	2.319	0.223	1.307	0.030	0.136	0.008				
SF32606	0.959	0.057	-0.688	0.072	0.193	0.030				
SF32611	0.984	0.106	1.254	0.057	0.184	0.015				
SF32614	0.858	0.033	0.104	0.026	0.000	0.000				
SF32640	0.715	0.029	0.082	0.030	0.000	0.000				
SF32645	1.513	0.147	1.053	0.037	0.278	0.013				
SF32693A	1.062	0.039	0.438	0.023	0.000	0.000				
SF32693B	0.986	0.034	0.956	0.022	0.000	0.000	0.546	0.025	-0.546	0.039
SF32695	0.884	0.029	0.972	0.024	0.000	0.000	-0.102	0.033	0.102	0.042
SF32697D	0.920	0.028	0.911	0.021	0.000	0.000	-0.169	0.031	0.169	0.039
SF32704	0.975	0.041	0.876	0.031	0.000	0.000				
SF32705A	1.547	0.054	0.523	0.018	0.000	0.000				
SF32705B	1.599	0.054	0.333	0.016	0.000	0.000				
SF32706A	1.127	0.045	0.832	0.027	0.000	0.000				
SF32706B	1.424	0.056	0.890	0.023	0.000	0.000				
SF32707	1.864	0.090	1.295	0.028	0.000	0.000				

Exhibit D.13 **IRT Parameters for TIMSS 2003 Eighth-Grade Science - Earth Science**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S012001	0.587	0.036	-0.138	0.087	0.165	0.028				
S012014	0.994	0.044	-0.502	0.049	0.236	0.020				
S012026	1.023	0.074	0.105	0.066	0.562	0.017				
S012028	0.761	0.040	0.481	0.039	0.113	0.015				
S012038	0.950	0.065	0.521	0.048	0.374	0.016				
S012039	0.904	0.056	-0.161	0.069	0.440	0.021				
S022106	0.799	0.061	1.599	0.053	0.124	0.009				
S022115	0.927	0.041	0.255	0.036	0.262	0.014				
S022117	0.727	0.047	0.693	0.047	0.176	0.017				
S022126	0.492	0.032	0.306	0.083	0.143	0.025				
S022150	0.836	0.044	0.649	0.036	0.225	0.013				
S022152	0.939	0.025	0.256	0.017	0.000	0.000				
S022154	0.634	0.019	-0.197	0.024	0.000	0.000				
S022160	0.635	0.021	0.693	0.028	0.000	0.000				
S022161	0.593	0.020	0.500	0.027	0.000	0.000				
S022235	0.809	0.076	0.842	0.062	0.422	0.017				
S022289	0.724	0.014	0.907	0.015	0.000	0.000	0.752	0.018	-0.752	0.028
S032007	0.851	0.033	0.375	0.027	0.000	0.000				
S032008	0.870	0.048	0.140	0.051	0.273	0.019				
S032015	0.835	0.035	0.734	0.032	0.000	0.000				
S032035	1.283	0.045	0.482	0.019	0.158	0.009				
S032083	0.933	0.066	1.263	0.040	0.121	0.010				
S032087	0.771	0.082	0.973	0.066	0.224	0.022				
S032202	0.607	0.014	-0.002	0.016	0.000	0.000	0.265	0.030	-0.265	0.029
S032206	1.040	0.034	1.175	0.027	0.000	0.000				
S032258	0.799	0.033	-0.069	0.043	0.176	0.016				
S032306	0.496	0.013	0.523	0.026	0.000	0.000	-1.415	0.066	1.415	0.070
S032310D	0.567	0.018	0.024	0.024	0.000	0.000	-0.065	0.046	0.065	0.045
S032315	0.911	0.080	0.609	0.058	0.268	0.021				
S032385	0.741	0.036	-0.043	0.056	0.260	0.019				
S032386	0.891	0.085	1.564	0.061	0.173	0.011				
S032451	0.627	0.015	0.046	0.020	0.000	0.000	-1.183	0.056	1.183	0.056
S032465	0.679	0.057	-0.188	0.110	0.241	0.036				
S032530D	0.506	0.019	0.443	0.030	0.000	0.000	0.715	0.045	-0.715	0.052
S032542	1.472	0.134	0.855	0.038	0.320	0.014				
S032595	1.359	0.122	1.480	0.043	0.183	0.008				
S032606	1.190	0.089	-0.280	0.070	0.452	0.025				
S032607	0.692	0.039	-0.171	0.071	0.192	0.025				
S032611	1.041	0.129	1.367	0.065	0.232	0.015				
S032614	0.788	0.031	0.116	0.028	0.000	0.000				
S032637	1.078	0.079	1.113	0.035	0.230	0.011				
S032640	0.551	0.025	0.015	0.038	0.000	0.000				
S032645	0.716	0.078	0.956	0.072	0.225	0.023				
S032682	1.510	0.095	0.997	0.024	0.239	0.009				
S032693A	1.068	0.038	0.184	0.022	0.000	0.000				
S032693B	0.922	0.031	0.782	0.021	0.000	0.000	0.598	0.026	-0.598	0.037
S032695	0.885	0.028	0.749	0.020	0.000	0.000	-0.021	0.031	0.021	0.038
S032697D	0.916	0.027	0.628	0.018	0.000	0.000	-0.051	0.030	0.051	0.034
S032704	0.945	0.039	0.707	0.029	0.000	0.000				
S032705A	1.383	0.049	0.453	0.019	0.000	0.000				
S032705B	1.400	0.048	0.185	0.018	0.000	0.000				
S032706A	1.144	0.044	0.664	0.024	0.000	0.000				

Exhibit D.13 **IRT Parameters for TIMSS 2003 Eighth-Grade Science - Earth Science**

(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S032706B	1.340	0.051	0.731	0.022	0.000	0.000				
S032707	1.600	0.078	1.307	0.032	0.000	0.000				
SF12001	0.856	0.052	0.081	0.051	0.116	0.021				
SF12014	1.212	0.063	-0.237	0.041	0.154	0.019				
SF12026	0.974	0.070	-0.313	0.078	0.345	0.029				
SF12028	1.069	0.064	0.511	0.035	0.106	0.014				

Exhibit D.14 **IRT Parameters for TIMSS 2003 Eighth-Grade Science - Environmental Science**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S012005	0.684	0.057	0.457	0.069	0.305	0.021				
S012017	1.110	0.059	0.536	0.027	0.197	0.011				
S012042	1.137	0.082	0.779	0.035	0.352	0.012				
S022086	0.760	0.021	0.235	0.021	0.000	0.000				
S022088A	1.016	0.024	-0.416	0.019	0.000	0.000				
S022088B	0.793	0.021	0.168	0.020	0.000	0.000				
S022240	0.959	0.109	1.729	0.084	0.254	0.001				
S022244	1.151	0.029	0.922	0.017	0.000	0.000				
S022249D	0.834	0.023	0.273	0.019	0.000	0.000				
S032063	0.490	0.014	1.555	0.042	0.000	0.000	-0.448	0.041	0.448	0.059
S032120A	0.914	0.026	1.153	0.025	0.000	0.000				
S032120B	1.286	0.040	1.403	0.026	0.000	0.000				
S032122	0.642	0.031	0.941	0.048	0.000	0.000				
S032126	0.593	0.025	-0.086	0.036	0.000	0.000				
S032242	0.725	0.025	1.057	0.033	0.000	0.000				
S032422	2.305	0.160	0.473	0.023	0.309	0.013				
S032446	0.758	0.065	0.658	0.058	0.344	0.018				
S032463	0.976	0.071	0.149	0.055	0.229	0.021				
S032510	0.866	0.058	-0.579	0.089	0.217	0.033				
S032514	0.594	0.067	0.678	0.090	0.190	0.029				
S032516	0.712	0.027	-0.258	0.032	0.000	0.000				
S032519	0.789	0.018	0.407	0.017	0.000	0.000				
S032555	0.932	0.040	0.794	0.031	0.000	0.000				
S032620	0.613	0.098	1.795	0.142	0.183	0.020				
S032665A	0.937	0.037	0.555	0.027	0.000	0.000				
S032665B	4.615	0.209	0.798	0.001	0.000	0.000				
S032665C	2.901	0.119	0.790	0.013	0.000	0.000				
SF12005	2.336	0.170	0.662	0.022	0.271	0.011				
SF12017	2.825	0.178	0.608	0.017	0.189	0.001				
SF12042	1.982	0.135	0.575	0.024	0.256	0.012				
SF22086	1.037	0.040	0.501	0.024	0.000	0.000				
SF22088A	2.739	0.099	0.273	0.011	0.000	0.000				
SF22088B	1.983	0.075	0.562	0.015	0.000	0.000				
SF22240	1.078	0.132	1.476	0.076	0.200	0.012				
SF22244	0.756	0.041	1.470	0.065	0.000	0.000				
SF22249D	1.289	0.053	0.797	0.024	0.000	0.000				
SF32120A	1.703	0.077	1.078	0.026	0.000	0.000				
SF32120B	1.715	0.090	1.344	0.035	0.000	0.000				
SF32122	0.845	0.039	1.018	0.041	0.000	0.000				
SF32126	0.817	0.033	0.412	0.029	0.000	0.000				
SF32422	1.328	0.090	0.405	0.034	0.227	0.015				
SF32463	2.134	0.155	0.687	0.024	0.267	0.012				
SF32510	1.222	0.099	0.236	0.051	0.383	0.019				
SF32514	2.197	0.195	0.987	0.028	0.271	0.011				
SF32516	0.783	0.029	0.052	0.028	0.000	0.000				
SF32519	1.026	0.042	0.680	0.027	0.000	0.000				
SF32555	1.363	0.057	0.862	0.025	0.000	0.000				
SF32620	0.983	0.121	1.618	0.087	0.167	0.012				
SF32665A	1.211	0.050	0.729	0.024	0.000	0.000				
SF32665B	3.142	0.136	0.864	0.013	0.000	0.000				
SF32665C	2.702	0.117	0.888	0.015	0.000	0.000				

Exhibit D.15 **IRT Parameters for TIMSS 2003 Fourth-Grade Mathematics - Number**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M011001	0.885	0.054	-0.890	0.094	0.214	0.041				
M011002	0.905	0.066	0.465	0.058	0.225	0.022				
M011003	0.763	0.046	-0.353	0.079	0.148	0.031				
M011004	0.761	0.044	-0.983	0.098	0.139	0.039				
M011007	1.090	0.062	-1.423	0.083	0.169	0.042				
M011008	0.952	0.049	-0.466	0.058	0.130	0.026				
M011011	1.122	0.059	-0.955	0.062	0.174	0.031				
M011015	0.918	0.055	-0.063	0.061	0.197	0.025				
M011016	0.986	0.061	0.240	0.051	0.195	0.021				
M011018	0.716	0.036	-1.333	0.088	0.078	0.033				
M011019	0.729	0.041	-0.512	0.078	0.106	0.030				
M011020	0.685	0.060	0.781	0.072	0.176	0.024				
M011021	0.748	0.044	-0.597	0.087	0.143	0.034				
M011024	0.826	0.049	-1.994	0.135	0.161	0.060				
M011026	0.623	0.047	-0.426	0.132	0.199	0.045				
M011028	0.711	0.046	-0.442	0.097	0.159	0.037				
M012044	0.908	0.054	0.017	0.057	0.172	0.023				
M012117	0.937	0.067	0.485	0.054	0.218	0.021				
M012119	0.635	0.053	0.069	0.113	0.235	0.036				
M031009	0.760	0.041	0.587	0.044	0.000	0.000				
M031011	0.817	0.029	-0.074	0.026	0.000	0.000				
M031016	0.973	0.052	0.915	0.043	0.000	0.000				
M031029	0.969	0.088	-0.023	0.088	0.270	0.035				
M031030	0.940	0.054	1.202	0.053	0.000	0.000				
M031065	1.004	0.034	0.038	0.022	0.000	0.000				
M031106	0.786	0.023	0.097	0.022	0.000	0.000				
M031108	1.115	0.065	0.362	0.039	0.158	0.017				
M031128	0.419	0.031	-1.561	0.119	0.000	0.000				
M031130	0.943	0.047	-0.505	0.038	0.000	0.000				
M031162	0.654	0.026	-0.691	0.038	0.000	0.000				
M031173	1.364	0.087	-0.279	0.046	0.117	0.024				
M031183	0.598	0.028	0.126	0.033	0.000	0.000	0.454	0.056	-0.454	0.057
M031185	1.429	0.119	0.414	0.047	0.216	0.021				
M031210	0.830	0.113	0.959	0.094	0.289	0.029				
M031216	0.905	0.059	-0.321	0.078	0.263	0.032				
M031218	1.083	0.092	0.204	0.064	0.204	0.028				
M031235	0.672	0.022	0.415	0.027	0.000	0.000				
M031282	0.657	0.013	0.901	0.019	0.000	0.000	-0.997	0.042	0.997	0.047
M031285	0.671	0.023	0.830	0.032	0.000	0.000				
M031286	0.864	0.025	0.244	0.021	0.000	0.000				
M031303	1.378	0.099	-0.484	0.059	0.210	0.032				
M031304	0.904	0.031	-0.557	0.028	0.000	0.000				
M031305	0.680	0.027	-1.043	0.044	0.000	0.000				
M031306	0.710	0.027	-0.596	0.034	0.000	0.000				
M031309	1.243	0.056	-0.299	0.029	0.000	0.000				
M031310	1.269	0.058	-0.631	0.040	0.107	0.021				
M031313	0.563	0.034	-1.335	0.084	0.000	0.000				
M031332	0.838	0.085	0.242	0.098	0.257	0.035				
M031341	0.795	0.046	-0.715	0.085	0.144	0.035				
M031344A	0.712	0.039	0.354	0.043	0.000	0.000				
M031344B	1.222	0.056	0.264	0.027	0.000	0.000				
M031344C	0.710	0.022	0.074	0.025	0.000	0.000	-1.220	0.075	1.220	0.075

Exhibit D.15 **IRT Parameters for TIMSS 2003 Fourth-Grade Mathematics - Number**

(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M031345A	1.084	0.051	-0.061	0.030	0.000	0.000				
M031345B	1.003	0.048	0.016	0.031	0.000	0.000				
M031345C	0.657	0.048	1.659	0.100	0.000	0.000				
M031346A	1.861	0.084	-0.224	0.022	0.000	0.000				
M031346B	1.949	0.090	0.498	0.021	0.000	0.000				
M031346C	1.400	0.056	0.359	0.018	0.000	0.000	0.346	0.027	-0.346	0.030
M031347C	1.017	0.036	0.593	0.025	0.000	0.000				
M031348A	0.785	0.031	0.673	0.032	0.000	0.000				
M031348B	0.726	0.027	1.338	0.033	0.000	0.000	0.543	0.032	-0.543	0.056
M031379	1.207	0.061	0.871	0.034	0.000	0.000				
M031380	1.120	0.064	1.241	0.047	0.000	0.000				
MF11001	1.359	0.077	-0.474	0.040	0.055	0.018				
MF11002	0.971	0.068	0.471	0.046	0.056	0.016				
MF11003	1.253	0.080	0.177	0.039	0.074	0.017				
MF11004	1.484	0.090	-0.075	0.036	0.084	0.018				
MF11007	2.121	0.130	-0.580	0.031	0.090	0.019				
MF11008	1.519	0.086	-0.061	0.032	0.055	0.014				
MF11011	1.199	0.075	-0.790	0.058	0.090	0.027				
MF11015	0.950	0.066	-0.006	0.059	0.093	0.025				
MF11016	1.155	0.085	0.363	0.046	0.110	0.020				
MF11018	1.289	0.078	-0.720	0.050	0.077	0.023				
MF11019	1.082	0.063	-0.349	0.046	0.054	0.018				
MF11020	0.750	0.068	0.612	0.071	0.099	0.025				
MF11021	1.109	0.070	-0.048	0.047	0.079	0.021				
MF11024	1.393	0.089	-1.155	0.056	0.071	0.025				
MF11026	1.151	0.066	0.043	0.037	0.041	0.014				
MF11028	1.322	0.070	-0.003	0.031	0.028	0.010				
MF12044	1.043	0.066	-0.055	0.049	0.073	0.020				
MF12117	0.858	0.065	0.481	0.053	0.062	0.019				
MF12119	0.934	0.064	0.184	0.052	0.071	0.020				
MF31009	0.887	0.046	0.721	0.040	0.000	0.000				
MF31016	1.296	0.066	0.924	0.034	0.000	0.000				
MF31029	0.822	0.091	0.423	0.102	0.285	0.034				
MF31030	0.722	0.046	1.439	0.076	0.000	0.000				
MF31065	1.182	0.053	0.184	0.028	0.000	0.000				
MF31106	1.190	0.054	0.053	0.028	0.000	0.000				
MF31128	0.596	0.035	-0.723	0.059	0.000	0.000				
MF31130	1.028	0.049	0.139	0.031	0.000	0.000				
MF31173	1.263	0.079	0.140	0.039	0.075	0.017				
MF31183	0.912	0.038	0.433	0.025	0.000	0.000	0.366	0.037	-0.366	0.043
MF31185	1.088	0.076	0.149	0.051	0.108	0.021				
MF31210	1.149	0.121	0.843	0.059	0.238	0.021				
MF31218	1.363	0.089	0.319	0.036	0.076	0.016				
MF31235	0.833	0.042	0.311	0.037	0.000	0.000				
MF31282	0.783	0.027	0.847	0.029	0.000	0.000	-0.885	0.064	0.885	0.071
MF31285	0.771	0.042	0.761	0.047	0.000	0.000				
MF31286	1.410	0.062	0.199	0.025	0.000	0.000				
MF31303	1.408	0.103	-0.260	0.055	0.221	0.028				
MF31305	0.817	0.042	-0.786	0.048	0.000	0.000				
MF31309	1.571	0.069	-0.116	0.024	0.000	0.000				
MF31310	1.509	0.010	-0.342	0.047	0.155	0.026				

Exhibit D.15 **IRT Parameters for TIMSS 2003 Fourth-Grade Mathematics - Number**

(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
MF31313	0.625	0.034	-0.759	0.056	0.000	0.000				
MF31332	1.302	0.102	0.375	0.048	0.177	0.021				
MF31344A	0.928	0.050	0.798	0.042	0.000	0.000				
MF31344B	1.905	0.090	0.558	0.022	0.000	0.000				
MF31344C	1.001	0.032	0.443	0.021	0.000	0.000	-0.888	0.059	0.888	0.061
MF31345A	1.379	0.063	0.294	0.025	0.000	0.000				
MF31345B	1.294	0.060	0.369	0.027	0.000	0.000				
MF31345C	0.882	0.064	1.744	0.093	0.000	0.000				
MF31346A	1.702	0.076	-0.253	0.023	0.000	0.000				
MF31346B	1.749	0.083	0.603	0.024	0.000	0.000				
MF31346C	1.250	0.050	0.448	0.020	0.000	0.000	0.408	0.030	-0.408	0.034
MF31379	1.261	0.066	1.035	0.037	0.000	0.000				
MF31380	1.172	0.069	1.351	0.050	0.000	0.000				

Exhibit D.16 **IRT Parameters for TIMSS 2003 Fourth-Grade Mathematics - Measurement**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M011005	0.444	0.033	-1.908	0.271	0.185	0.075				
M011010	0.946	0.053	-0.288	0.055	0.142	0.021				
M011013	1.103	0.092	0.458	0.050	0.351	0.018				
M011017	0.650	0.045	-0.533	0.111	0.152	0.036				
M011023	0.506	0.038	-1.245	0.209	0.174	0.061				
M011025	0.641	0.063	0.493	0.087	0.229	0.028				
M012023	1.183	0.091	0.093	0.058	0.426	0.020				
M012065	1.292	0.094	0.628	0.034	0.230	0.015				
M031004	2.173	0.207	0.815	0.028	0.161	0.013				
M031006	0.690	0.065	-0.823	0.169	0.221	0.055				
M031008	1.173	0.130	1.445	0.062	0.202	0.012				
M031038	0.817	0.089	-0.254	0.130	0.341	0.040				
M031041	0.633	0.021	0.087	0.027	0.000	0.000				
M031043	1.430	0.117	0.382	0.041	0.183	0.019				
M031050	1.255	0.084	0.704	0.031	0.276	0.013				
M031064	1.364	0.130	0.661	0.042	0.189	0.018				
M031068	1.312	0.037	0.310	0.015	0.000	0.000				
M031097	1.290	0.125	0.476	0.051	0.244	0.021				
M031178	0.905	0.096	0.809	0.059	0.130	0.020				
M031219	0.448	0.090	1.064	0.210	0.291	0.049				
M031276	1.691	0.153	0.402	0.042	0.287	0.020				
M031294	0.960	0.074	-0.033	0.063	0.125	0.025				
M031297	0.643	0.038	0.463	0.049	0.000	0.000				
M031298	1.037	0.040	0.664	0.025	0.000	0.000				
M031299	1.302	0.035	0.055	0.016	0.000	0.000				
M031301	1.057	0.028	-0.680	0.023	0.000	0.000				
M031322	0.453	0.021	-1.294	0.064	0.000	0.000				
M031335	1.107	0.054	0.028	0.036	0.198	0.015				
M031338	0.643	0.062	0.155	0.110	0.273	0.033				
M031350A	1.801	0.051	0.482	0.012	0.000	0.000				
M031350B	1.767	0.048	0.155	0.013	0.000	0.000				
M031350C	1.287	0.040	0.687	0.017	0.000	0.000				
MF11005	1.505	0.112	0.035	0.045	0.213	0.021				
MF11010	0.995	0.078	-0.178	0.069	0.160	0.027				
MF11013	0.793	0.074	0.198	0.077	0.141	0.028				
MF11017	0.942	0.064	-0.374	0.065	0.083	0.024				
MF11023	0.954	0.074	-0.579	0.089	0.171	0.034				
MF11025	1.060	0.115	0.804	0.054	0.181	0.020				
MF12023	1.475	0.104	0.066	0.041	0.162	0.019				
MF12065	0.946	0.081	0.529	0.050	0.095	0.018				
MF31004	1.111	0.107	0.864	0.046	0.111	0.016				
MF31006	0.904	0.074	-0.126	0.076	0.152	0.029				
MF31038	0.995	0.071	-0.161	0.059	0.102	0.023				
MF31041	0.554	0.033	0.246	0.052	0.000	0.000				
MF31043	1.407	0.115	0.485	0.039	0.155	0.017				
MF31050	0.815	0.108	0.721	0.085	0.271	0.028				
MF31064	0.863	0.072	0.720	0.049	0.053	0.014				
MF31068	1.661	0.077	0.336	0.022	0.000	0.000				
MF31097	1.781	0.179	0.857	0.034	0.154	0.014				
MF31178	2.900	0.313	0.963	0.025	0.153	0.011				
MF31219	0.998	0.115	0.795	0.060	0.205	0.021				
MF31276	1.081	0.091	0.386	0.050	0.145	0.021				

Exhibit D.16 **IRT Parameters for TIMSS 2003 Fourth-Grade Mathematics - Measurement**

(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
MF31294	2.309	0.191	0.469	0.029	0.213	0.015				
MF31297	2.583	0.123	0.413	0.017	0.000	0.000				
MF31298	1.039	0.059	0.848	0.039	0.000	0.000				
MF31299	1.500	0.068	0.112	0.024	0.000	0.000				
MF31301	1.271	0.056	-0.373	0.030	0.000	0.000				
MF31322	1.043	0.048	-0.392	0.035	0.000	0.000				
MF31335	1.254	0.093	0.055	0.049	0.167	0.021				
MF31350A	2.625	0.127	0.488	0.016	0.000	0.000				
MF31350B	2.950	0.139	0.244	0.016	0.000	0.000				
MF31350C	1.699	0.088	0.702	0.024	0.000	0.000				

Exhibit D.17 **IRT Parameters for TIMSS 2003 Fourth-Grade Mathematics - Data**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M011009	1.032	0.051	-1.409	0.076	0.122	0.035				
M011012	0.819	0.039	-1.224	0.079	0.084	0.030				
M012078	0.908	0.051	-0.796	0.078	0.167	0.030				
M012126	0.825	0.047	-0.576	0.074	0.135	0.027				
M031045	1.386	0.062	-0.181	0.032	0.213	0.015				
M031133	0.740	0.036	-0.965	0.054	0.000	0.000				
M031134	0.582	0.024	1.065	0.043	0.000	0.000				
M031135	1.254	0.091	-0.486	0.064	0.189	0.027				
M031155	1.167	0.091	0.103	0.053	0.171	0.022				
M031172	1.046	0.077	-0.256	0.065	0.143	0.026				
M031240	0.693	0.021	-1.034	0.035	0.000	0.000				
M031242B	1.933	0.090	0.292	0.020	0.000	0.000				
M031242C	2.021	0.178	0.383	0.037	0.303	0.019				
M031264	1.147	0.037	-1.020	0.030	0.000	0.000				
M031265	0.759	0.030	0.340	0.030	0.000	0.000				
M031315	1.606	0.094	0.431	0.027	0.201	0.013				
M031333	1.262	0.116	0.603	0.045	0.174	0.019				
MF11009	1.210	0.089	-1.047	0.087	0.192	0.037				
MF11012	1.854	0.124	-0.605	0.042	0.131	0.020				
MF12078	1.331	0.090	-0.171	0.047	0.134	0.021				
MF12126	1.019	0.087	0.068	0.066	0.199	0.026				
MF31045	2.063	0.128	-0.021	0.028	0.108	0.015				
MF31133	0.760	0.037	-0.329	0.042	0.000	0.000				
MF31134	0.778	0.047	0.979	0.054	0.000	0.000				
MF31135	0.951	0.079	0.228	0.059	0.138	0.023				
MF31155	1.228	0.100	0.346	0.047	0.174	0.020				
MF31172	1.426	0.106	0.379	0.037	0.136	0.017				
MF31240	0.653	0.033	-0.645	0.052	0.000	0.000				
MF31242B	1.697	0.082	0.524	0.022	0.000	0.000				
MF31242C	3.001	0.265	0.552	0.024	0.242	0.015				
MF31264	1.853	0.085	-0.196	0.023	0.000	0.000				
MF31265	1.804	0.089	0.525	0.022	0.000	0.000				
MF31333	2.083	0.192	0.930	0.030	0.129	0.011				

Exhibit D.18 **IRT Parameters for TIMSS 2003 Fourth-Grade Mathematics - Geometry**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M011006	0.518	0.041	-0.037	0.102	0.127	0.031				
M011014	0.708	0.045	-1.746	0.134	0.142	0.042				
M011022	0.582	0.039	-0.920	0.114	0.119	0.034				
M012069	0.380	0.042	0.951	0.143	0.130	0.034				
M031071	1.267	0.115	0.684	0.046	0.142	0.019				
M031083	1.318	0.098	-0.193	0.050	0.156	0.026				
M031085	0.937	0.098	0.611	0.072	0.212	0.027				
M031088	0.574	0.065	-0.440	0.176	0.246	0.051				
M031093	1.159	0.114	0.431	0.058	0.234	0.026				
M031109	0.663	0.070	-0.136	0.125	0.215	0.042				
M031159	1.141	0.100	-0.037	0.065	0.241	0.031				
M031267	0.688	0.030	0.319	0.031	0.000	0.000				
M031269	0.378	0.010	-0.797	0.037	0.000	0.000	-1.839	0.090	1.839	0.082
M031271	0.742	0.025	-1.265	0.042	0.000	0.000				
M031272A	2.364	0.087	-0.617	0.016	0.000	0.000				
M031272B	2.183	0.093	-0.988	0.023	0.000	0.000				
M031272C	2.236	0.075	0.107	0.012	0.000	0.000				
M031274	0.816	0.028	-0.381	0.025	0.000	0.000				
M031325	0.969	0.056	0.703	0.041	0.000	0.000				
M031327	0.440	0.023	-0.191	0.045	0.000	0.000				
M031330	1.009	0.055	-0.854	0.047	0.000	0.000				
M031347A	3.424	0.116	0.137	0.009	0.000	0.000				
M031347B	3.477	0.119	0.177	0.009	0.000	0.000				
M031351	1.542	0.127	0.256	0.039	0.187	0.022				
MF11006	0.854	0.088	0.643	0.076	0.189	0.028				
MF11014	2.072	0.141	-0.505	0.038	0.163	0.025				
MF11022	1.484	0.108	0.057	0.038	0.138	0.021				
MF12069	1.117	0.126	1.203	0.075	0.199	0.019				
MF31071	1.605	0.156	0.879	0.043	0.168	0.016				
MF31083	1.713	0.155	0.250	0.041	0.253	0.024				
MF31085	0.964	0.098	0.875	0.067	0.161	0.022				
MF31088	1.320	0.104	0.051	0.047	0.182	0.025				
MF31093	0.742	0.073	0.602	0.078	0.139	0.027				
MF31109	2.624	0.211	0.268	0.025	0.188	0.017				
MF31159	4.574	0.331	0.168	0.015	0.126	0.013				
MF31269	0.316	0.012	-0.030	0.048	0.000	0.000	-2.214	0.135	2.214	0.134
MF31271	0.776	0.044	-0.768	0.052	0.000	0.000				
MF31274	0.932	0.050	-0.283	0.035	0.000	0.000				
MF31325	1.124	0.064	0.876	0.042	0.000	0.000				
MF31327	0.371	0.030	0.524	0.081	0.000	0.000				
MF31330	1.290	0.064	-0.224	0.027	0.000	0.000				
MF31351	2.438	0.172	0.397	0.021	0.081	0.012				

Exhibit D.19 **IRT Parameters for TIMSS 2003 Fourth-Grade Mathematics - Patterns and Relationships**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
M011027	0.733	0.048	-0.676	0.110	0.173	0.042				
M012048	0.958	0.068	0.096	0.064	0.259	0.025				
M031023	0.626	0.054	-0.043	0.124	0.215	0.040				
M031051	0.767	0.047	-0.591	0.091	0.142	0.036				
M031079B	1.892	0.086	-0.543	0.025	0.000	0.000				
M031079C	1.293	0.062	0.387	0.028	0.000	0.000				
M031098	1.684	0.122	0.240	0.036	0.162	0.017				
M031187	1.035	0.081	-0.513	0.088	0.199	0.038				
M031190	1.329	0.086	0.416	0.036	0.240	0.015				
M031220	0.867	0.050	-0.884	0.091	0.166	0.039				
M031227	1.003	0.037	1.349	0.034	0.000	0.000				
M031242A	0.826	0.039	-0.209	0.038	0.000	0.000				
M031245	1.600	0.149	1.040	0.039	0.122	0.012				
M031247	0.554	0.027	1.273	0.055	0.000	0.000	-0.234	0.065	0.234	0.089
M031249	0.758	0.036	1.499	0.057	0.000	0.000				
M031251	1.046	0.103	0.645	0.058	0.191	0.022				
M031252	0.853	0.074	-0.142	0.093	0.169	0.036				
M031254	0.947	0.091	0.311	0.074	0.217	0.028				
M031255	0.956	0.055	0.092	0.051	0.247	0.019				
M031258	1.011	0.030	0.658	0.021	0.000	0.000				
M031316	0.557	0.037	-2.156	0.121	0.000	0.000				
M031317	0.756	0.067	0.521	0.066	0.081	0.023				
M031334	1.392	0.077	0.683	0.026	0.214	0.011				
MF11027	1.079	0.094	0.262	0.060	0.190	0.025				
MF12048	0.679	0.061	-0.103	0.110	0.125	0.039				
MF31051	1.318	0.094	-0.098	0.051	0.158	0.024				
MF31079B	1.812	0.080	-0.353	0.024	0.000	0.000				
MF31079C	1.542	0.076	0.575	0.025	0.000	0.000				
MF31098	1.751	0.114	0.167	0.031	0.101	0.014				
MF31187	1.619	0.131	0.175	0.044	0.251	0.021				
MF31220	1.649	0.116	-0.199	0.044	0.180	0.022				
MF31227	1.262	0.075	1.253	0.045	0.000	0.000				
MF31242A	1.161	0.053	0.212	0.028	0.000	0.000				
MF31245	1.379	0.130	1.117	0.044	0.101	0.012				
MF31247	0.695	0.034	1.477	0.055	0.000	0.000	-0.227	0.058	0.227	0.086
MF31251	1.893	0.161	0.715	0.032	0.177	0.014				
MF31252	0.974	0.078	-0.030	0.069	0.145	0.029				
MF31254	1.526	0.124	0.437	0.039	0.185	0.018				
MF31255	1.368	0.125	0.219	0.057	0.307	0.024				
MF31258	1.158	0.057	0.609	0.031	0.000	0.000				
MF31316	0.893	0.044	-1.243	0.052	0.000	0.000				
MF31317	0.838	0.062	0.436	0.051	0.050	0.017				
MF31334	1.221	0.113	0.696	0.047	0.180	0.019				

Exhibit D.20 **IRT Parameters for TIMSS 2003 Fourth-Grade Science - Life Science**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S011004	0.590	0.042	-0.656	0.110	0.166	0.032				
S011010	0.516	0.043	-2.916	0.349	0.317	0.086				
S011015	0.521	0.045	-0.971	0.183	0.266	0.045				
S011016	0.550	0.040	-2.553	0.248	0.215	0.067				
S011019	0.544	0.024	-1.287	0.060	0.000	0.000				
S011021	0.437	0.036	-2.159	0.284	0.198	0.065				
S011025	0.735	0.055	-1.068	0.133	0.379	0.037				
S011026	0.285	0.024	-3.703	0.464	0.182	0.081				
S011031	0.750	0.043	-1.834	0.117	0.139	0.035				
S011033	0.382	0.069	1.626	0.183	0.228	0.035				
S012010	1.041	0.071	-0.069	0.057	0.294	0.023				
S012033	0.470	0.047	-0.036	0.141	0.183	0.037				
S031001	0.870	0.067	-0.849	0.099	0.180	0.036				
S031003	0.638	0.048	-0.420	0.103	0.211	0.032				
S031017	0.729	0.059	-0.437	0.112	0.326	0.035				
S031026	0.508	0.012	-0.116	0.020	0.000	0.000	-0.736	0.045	0.736	0.043
S031190	1.004	0.060	0.714	0.038	0.000	0.000				
S031193	0.555	0.061	-0.346	0.152	0.167	0.044				
S031212	0.580	0.046	-0.571	0.127	0.213	0.037				
S031218	0.674	0.029	-0.143	0.032	0.000	0.000				
S031229	1.796	0.108	0.669	0.023	0.287	0.011				
S031230	0.588	0.053	-1.478	0.190	0.164	0.049				
S031233	0.209	0.023	-0.973	0.157	0.000	0.000				
S031235A	1.408	0.042	0.441	0.014	0.000	0.000				
S031235B	1.493	0.045	0.561	0.014	0.000	0.000				
S031236	0.834	0.067	-1.014	0.116	0.205	0.040				
S031239	0.998	0.081	0.285	0.059	0.465	0.019				
S031240D	0.646	0.017	-0.256	0.019	0.000	0.000	0.710	0.034	-0.710	0.028
S031241D	0.631	0.023	0.513	0.024	0.000	0.000	0.563	0.035	-0.563	0.042
S031246	0.798	0.040	1.028	0.044	0.000	0.000				
S031251	0.613	0.033	1.035	0.054	0.000	0.000				
S031252	0.473	0.018	-1.298	0.046	0.000	0.000	0.500	0.072	-0.500	0.046
S031254	1.449	0.292	1.088	0.084	0.503	0.020				
S031255	1.127	0.066	0.036	0.044	0.321	0.018				
S031264	0.846	0.059	-0.444	0.066	0.068	0.022				
S031266	2.966	0.325	0.696	0.030	0.352	0.016				
S031269	0.569	0.059	0.439	0.097	0.226	0.030				
S031270	0.545	0.030	1.701	0.084	0.000	0.000				
S031281	2.934	0.398	0.316	0.041	0.635	0.018				
S031283	0.708	0.072	-0.521	0.138	0.272	0.043				
S031284	0.665	0.133	1.940	0.202	0.228	0.020				
S031287	1.063	0.079	0.136	0.053	0.298	0.022				
S031291	0.994	0.067	-1.033	0.079	0.109	0.027				
S031317	0.768	0.095	0.011	0.126	0.366	0.039				
S031319	1.562	0.108	0.962	0.027	0.211	0.001				
S031325	0.667	0.044	0.417	0.046	0.000	0.000				
S031326D	0.471	0.018	0.346	0.027	0.000	0.000	0.012	0.049	-0.012	0.053
S031330	0.883	0.033	-0.576	0.031	0.000	0.000				
S031338	0.718	0.046	-0.682	0.089	0.183	0.029				
S031340	0.910	0.111	0.833	0.067	0.181	0.024				
S031346	0.811	0.065	1.358	0.086	0.000	0.000				
S031347	0.671	0.065	-0.754	0.145	0.198	0.046				

Exhibit D.20 **IRT Parameters for TIMSS 2003 Fourth-Grade Science - Life Science**

(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S031349	0.684	0.047	-1.324	0.132	0.234	0.040				
S031356	0.635	0.068	-1.624	0.246	0.331	0.060				
S031361	0.808	0.097	0.561	0.079	0.216	0.029				
S031390D	0.546	0.029	0.355	0.035	0.000	0.000	0.123	0.061	-0.123	0.065
S031426	0.796	0.080	-0.160	0.105	0.260	0.037				
S031431	0.718	0.159	1.824	0.212	0.185	0.023				
S031439A	0.898	0.066	1.259	0.071	0.000	0.000				
S031439B	0.693	0.043	0.116	0.042	0.000	0.000				
S031441A	1.559	0.075	-0.089	0.024	0.000	0.000				
S031441B	1.238	0.055	0.580	0.021	0.000	0.000	0.514	0.028	-0.514	0.037
S031442	1.376	0.071	0.194	0.025	0.000	0.000				
S031443	0.946	0.065	0.959	0.051	0.000	0.000				
SF11004	3.624	0.290	0.200	0.019	0.213	0.016				
SF11010	4.840	0.436	-0.023	0.021	0.279	0.019				
SF11015	2.282	0.163	0.213	0.024	0.135	0.016				
SF11016	2.429	0.128	-0.139	0.021	0.017	0.006				
SF11019	1.970	0.092	-0.002	0.019	0.000	0.000				
SF11021	4.966	0.481	0.119	0.019	0.327	0.018				
SF11025	2.778	0.150	-0.112	0.019	0.016	0.005				
SF11026	2.216	0.116	-0.145	0.023	0.019	0.006				
SF11031	2.245	0.124	-0.189	0.025	0.035	0.001				
SF11033	1.134	0.072	0.594	0.033	0.014	0.005				
SF12010	1.610	0.114	0.124	0.033	0.115	0.018				
SF12033	1.072	0.104	0.400	0.054	0.167	0.024				
SF31001	1.066	0.073	-0.532	0.065	0.142	0.028				
SF31017	1.219	0.104	-0.233	0.072	0.256	0.034				
SF31026	0.715	0.026	0.084	0.025	0.000	0.000	-0.611	0.058	0.611	0.057
SF31190	1.308	0.070	0.628	0.029	0.000	0.000				
SF31193	1.062	0.101	0.297	0.060	0.219	0.026				
SF31229	1.281	0.126	0.590	0.045	0.199	0.021				
SF31230	1.155	0.082	-0.548	0.067	0.175	0.030				
SF31233	0.994	0.053	0.261	0.031	0.000	0.000				
SF31235A	2.156	0.107	0.413	0.017	0.000	0.000				
SF31235B	2.243	0.113	0.530	0.018	0.000	0.000				
SF31236	0.819	0.067	-0.971	0.116	0.210	0.040				
SF31239	0.870	0.079	-0.162	0.086	0.216	0.033				
SF31240D	0.724	0.033	-0.080	0.029	0.000	0.000	0.553	0.050	-0.553	0.045
SF31246	1.257	0.074	0.897	0.037	0.000	0.000				
SF31251	0.843	0.057	1.008	0.057	0.000	0.000				
SF31254	1.556	0.154	0.522	0.043	0.279	0.021				
SF31255	1.387	0.116	0.058	0.050	0.226	0.025				
SF31264	1.242	0.100	0.151	0.049	0.174	0.023				
SF31266	1.688	0.128	0.473	0.029	0.116	0.015				
SF31270	0.921	0.066	1.218	0.066	0.000	0.000				
SF31281	1.468	0.127	0.260	0.043	0.224	0.022				
SF31283	1.177	0.089	-0.212	0.061	0.189	0.028				
SF31287	1.080	0.091	0.079	0.058	0.172	0.026				
SF31291	1.650	0.099	-0.464	0.041	0.108	0.021				
SF31317	1.111	0.098	-0.099	0.072	0.264	0.031				
SF31319	1.199	0.124	0.875	0.048	0.129	0.017				
SF31325	0.972	0.058	0.622	0.037	0.000	0.000				

Exhibit D.20 **IRT Parameters for TIMSS 2003 Fourth-Grade Science - Life Science**
(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
SF31340	1.037	0.122	0.789	0.060	0.209	0.023				
SF31346	1.099	0.078	1.310	0.063	0.000	0.000				
SF31347	1.270	0.114	0.139	0.056	0.258	0.026				
SF31356	1.442	0.118	-0.543	0.073	0.337	0.034				
SF31361	0.813	0.093	0.485	0.079	0.216	0.030				
SF31390D	1.122	0.049	0.501	0.020	0.000	0.000	0.125	0.032	-0.125	0.036
SF31426	1.314	0.108	0.081	0.050	0.203	0.025				
SF31431	0.949	0.154	1.479	0.112	0.133	0.018				
SF31439A	1.210	0.077	1.014	0.045	0.000	0.000				
SF31439B	0.906	0.052	0.327	0.034	0.000	0.000				
SF31441A	1.515	0.072	0.046	0.023	0.000	0.000				
SF31441B	1.194	0.055	0.634	0.021	0.000	0.000	0.376	0.028	-0.376	0.038
SF31442	1.401	0.072	0.372	0.024	0.000	0.000				
SF31443	1.484	0.086	0.852	0.032	0.000	0.000				

Exhibit D.21 **IRT Parameters for TIMSS 2003 Fourth-Grade Science - Earth Science**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S011003	0.939	0.065	-0.251	0.072	0.294	0.029				
S011005	0.860	0.047	-0.871	0.069	0.159	0.028				
S011007	0.737	0.046	-0.739	0.088	0.199	0.032				
S011012	0.697	0.047	-1.506	0.131	0.213	0.040				
S011013	0.735	0.059	0.665	0.064	0.189	0.022				
S011018	0.522	0.038	-2.936	0.277	0.192	0.074				
S011022	0.649	0.049	-0.411	0.107	0.251	0.035				
S011023	0.953	0.051	-0.372	0.051	0.138	0.023				
S011027	0.945	0.072	-0.165	0.076	0.348	0.029				
S011032	0.628	0.028	0.186	0.032	0.000	0.000				
S012007	0.587	0.047	-0.886	0.149	0.281	0.042				
S031044	0.565	0.038	0.269	0.051	0.000	0.000				
S031047	0.534	0.036	0.013	0.052	0.000	0.000				
S031060	0.978	0.097	1.328	0.066	0.255	0.016				
S031081	0.446	0.032	-1.046	0.093	0.000	0.000				
S031082	0.467	0.044	-0.663	0.196	0.251	0.048				
S031088D	0.290	0.014	0.673	0.071	0.000	0.000	1.886	0.104	-1.886	0.129
S031275	0.839	0.133	1.616	0.121	0.228	0.022				
S031278	0.670	0.024	-0.479	0.030	0.000	0.000				
S031376	0.940	0.131	1.309	0.098	0.266	0.024				
S031379	0.638	0.044	-0.081	0.080	0.141	0.028				
S031382	0.522	0.026	0.179	0.038	0.000	0.000				
S031383	0.638	0.051	0.955	0.066	0.103	0.020				
S031384A	2.034	0.057	-0.741	0.015	0.000	0.000				
S031384B	1.968	0.052	-0.232	0.013	0.000	0.000				
S031387	0.840	0.126	1.516	0.114	0.215	0.024				
S031389	1.001	0.140	1.337	0.095	0.271	0.022				
S031391D	0.358	0.019	0.407	0.049	0.000	0.000	-0.417	0.092	0.417	0.100
S031393	0.935	0.031	-1.066	0.033	0.000	0.000				
S031396D	0.463	0.022	-1.149	0.061	0.000	0.000	-0.412	0.101	0.412	0.076
S031398	1.040	0.010	0.128	0.070	0.232	0.031				
S031401	1.187	0.076	0.770	0.039	0.296	0.014				
S031440	0.624	0.044	1.138	0.077	0.000	0.000				
SF11003	1.523	0.093	-0.162	0.033	0.052	0.015				
SF11005	2.277	0.136	-0.198	0.026	0.043	0.012				
SF11007	1.377	0.084	-0.217	0.038	0.061	0.017				
SF11012	3.045	0.206	-0.137	0.023	0.073	0.013				
SF11013	1.637	0.113	0.490	0.026	0.036	0.009				
SF11018	1.868	0.120	-0.517	0.045	0.155	0.028				
SF11022	1.350	0.084	-0.130	0.036	0.055	0.016				
SF11023	2.216	0.143	0.004	0.022	0.035	0.001				
SF11027	3.155	0.223	0.125	0.016	0.052	0.009				
SF11032	1.217	0.066	0.575	0.032	0.000	0.000				
SF12007	2.539	0.185	0.083	0.020	0.067	0.012				
SF31044	0.512	0.037	0.858	0.075	0.000	0.000				
SF31047	0.802	0.046	0.247	0.038	0.000	0.000				
SF31081	0.575	0.037	-0.217	0.051	0.000	0.000				
SF31082	0.940	0.081	0.163	0.063	0.149	0.028				
SF31088D	0.727	0.036	1.001	0.038	0.000	0.000	0.659	0.042	-0.659	0.068
SF31275	0.828	0.103	1.430	0.096	0.148	0.021				
SF31278	1.377	0.068	-0.003	0.024	0.000	0.000				
SF31376	1.365	0.150	1.220	0.063	0.208	0.016				

Exhibit D.21 **IRT Parameters for TIMSS 2003 Fourth-Grade Science - Earth Science**

(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
SF31384A	6.236	0.404	-0.195	0.013	0.000	0.000				
SF31384B	5.358	0.317	0.120	0.009	0.000	0.000				
SF31387	0.879	0.104	1.297	0.086	0.138	0.021				
SF31389	1.637	0.143	1.121	0.047	0.119	0.012				
SF31391D	0.624	0.029	0.421	0.031	0.000	0.000	-0.041	0.054	0.041	0.060
SF31393	1.417	0.066	-0.503	0.029	0.000	0.000				
SF31396D	0.474	0.018	-0.307	0.036	0.000	0.000	-1.154	0.089	1.154	0.085
SF31398	0.869	0.081	0.190	0.075	0.170	0.032				
SF31401	1.039	0.097	0.646	0.057	0.163	0.022				
SF31440	0.697	0.048	1.223	0.076	0.000	0.000				

Exhibit D.22 **IRT Parameters for TIMSS 2003 Fourth-Grade Science - Physical Science**

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S011001	0.645	0.043	-1.564	0.162	0.191	0.055				
S011006	0.481	0.047	-0.989	0.251	0.272	0.063				
S011008	0.670	0.046	-0.428	0.094	0.145	0.033				
S011009	0.459	0.038	-1.079	0.212	0.172	0.057				
S011011	0.901	0.093	0.940	0.059	0.279	0.019				
S011014	0.774	0.053	-0.164	0.075	0.194	0.028				
S011017	0.559	0.042	-0.504	0.127	0.153	0.039				
S011029	0.789	0.048	-1.898	0.138	0.194	0.053				
S011030	0.449	0.035	-2.062	0.275	0.194	0.074				
S031005	0.651	0.036	1.492	0.072	0.000	0.000				
S031009	0.712	0.024	-0.086	0.024	0.000	0.000				
S031035	0.689	0.055	-0.687	0.134	0.287	0.043				
S031038	0.511	0.045	-0.993	0.210	0.219	0.059				
S031053	0.532	0.017	-0.013	0.023	0.000	0.000	-0.182	0.047	0.182	0.046
S031061	0.546	0.067	-0.387	0.212	0.249	0.059				
S031068	1.135	0.135	0.818	0.060	0.238	0.023				
S031072	0.672	0.030	0.073	0.032	0.000	0.000	0.767	0.052	-0.767	0.052
S031075	0.348	0.081	1.211	0.351	0.329	0.062				
S031076	0.558	0.040	0.908	0.072	0.000	0.000				
S031077	0.574	0.057	-0.921	0.194	0.187	0.058				
S031078	0.690	0.060	0.258	0.094	0.318	0.030				
S031197D	0.483	0.019	-0.688	0.044	0.000	0.000	-0.799	0.091	0.799	0.078
S031204	0.431	0.035	0.845	0.087	0.000	0.000				
S031205	0.622	0.044	0.051	0.085	0.194	0.028				
S031273	0.528	0.062	0.255	0.132	0.136	0.041				
S031298	0.602	0.120	1.728	0.182	0.204	0.029				
S031299	0.576	0.040	0.703	0.061	0.000	0.000				
S031306	0.974	0.090	0.961	0.048	0.194	0.017				
S031311	0.756	0.062	-0.159	0.081	0.106	0.030				
S031313	0.854	0.096	1.131	0.065	0.231	0.020				
S031370	0.954	0.036	0.199	0.023	0.000	0.000				
S031371	0.586	0.087	0.953	0.118	0.176	0.036				
S031372A	1.004	0.035	-0.284	0.024	0.000	0.000				
S031372B	0.805	0.025	0.910	0.024	0.000	0.000	-0.238	0.034	0.238	0.043
S031399A	1.405	0.039	0.285	0.014	0.000	0.000				
S031399B	1.453	0.039	0.087	0.013	0.000	0.000				
S031406A	0.862	0.033	-0.570	0.031	0.000	0.000				
S031406B	0.994	0.046	1.169	0.040	0.000	0.000				
S031409	0.945	0.081	-0.159	0.080	0.199	0.033				
S031410	0.385	0.047	-0.691	0.283	0.166	0.065				
S031414A	1.874	0.050	-0.165	0.012	0.000	0.000				
S031414B	1.699	0.046	-0.203	0.013	0.000	0.000				
S031418	0.867	0.109	0.905	0.075	0.188	0.026				
S031420	0.706	0.072	1.031	0.067	0.161	0.022				
S031421	0.428	0.031	-0.600	0.076	0.000	0.000				
S031422	0.601	0.052	-1.654	0.202	0.164	0.062				
S031427	0.460	0.055	-0.404	0.219	0.171	0.058				
S031445A	1.690	0.080	0.359	0.021	0.000	0.000				
S031445B	1.348	0.065	-0.466	0.030	0.000	0.000				
S031446A	1.148	0.060	0.577	0.031	0.000	0.000				
S031446B	0.937	0.056	0.826	0.045	0.000	0.000				
S031446C	0.846	0.046	0.015	0.035	0.000	0.000				

Exhibit D.22 **IRT Parameters for TIMSS 2003 Fourth-Grade Science - Physical Science**

(...Continued)

Item	Slope (a_j)	S.E. (a_j)	Location (b_j)	S.E. (b_j)	Guessing (c_j)	S.E. (c_j)	Step 1 (d_{j1})	S.E. (d_{j1})	Step 2 (d_{j2})	S.E. (d_{j2})
S031447	0.463	0.029	1.026	0.061	0.000	0.000	0.324	0.067	-0.324	0.092
SF11001	1.697	0.107	-0.401	0.039	0.113	0.021				
SF11006	1.065	0.074	-0.213	0.055	0.103	0.024				
SF11008	1.610	0.099	-0.044	0.032	0.076	0.016				
SF11009	1.606	0.109	0.025	0.036	0.138	0.020				
SF11011	1.663	0.139	0.694	0.033	0.122	0.014				
SF11014	2.053	0.129	0.332	0.024	0.073	0.012				
SF11017	1.101	0.084	0.096	0.051	0.132	0.024				
SF11029	5.168	0.469	0.135	0.017	0.235	0.016				
SF11030	4.958	0.455	0.140	0.018	0.255	0.016				
SF31005	0.913	0.067	1.479	0.080	0.000	0.000				
SF31009	0.899	0.048	0.193	0.033	0.000	0.000				
SF31053	0.808	0.032	0.204	0.023	0.000	0.000	-0.116	0.045	0.116	0.046
SF31061	0.707	0.077	-0.081	0.128	0.238	0.044				
SF31068	1.259	0.126	0.758	0.047	0.165	0.019				
SF31072	0.797	0.037	0.235	0.027	0.000	0.000	0.562	0.043	-0.562	0.045
SF31075	1.064	0.133	0.699	0.070	0.310	0.025				
SF31076	0.965	0.054	0.773	0.041	0.000	0.000				
SF31077	3.196	0.292	0.544	0.025	0.278	0.015				
SF31078	0.810	0.067	0.120	0.066	0.087	0.026				
SF31197D	0.568	0.021	-0.160	0.031	0.000	0.000	-0.688	0.071	0.688	0.067
SF31204	0.821	0.050	0.847	0.050	0.000	0.000				
SF31205	0.570	0.064	0.378	0.110	0.124	0.036				
SF31273	2.164	0.173	0.617	0.027	0.152	0.014				
SF31298	0.917	0.156	1.464	0.109	0.226	0.021				
SF31299	1.211	0.067	0.857	0.036	0.000	0.000				
SF31306	1.003	0.086	0.798	0.047	0.065	0.016				
SF31311	2.322	0.192	0.597	0.027	0.185	0.014				
SF31371	1.199	0.114	0.789	0.046	0.121	0.018				
SF31372A	2.285	0.103	0.246	0.017	0.000	0.000				
SF31372B	1.704	0.074	0.998	0.020	0.000	0.000	-0.053	0.028	0.053	0.037
SF31399A	3.308	0.161	0.335	0.014	0.000	0.000				
SF31399B	3.151	0.150	0.222	0.014	0.000	0.000				
SF31409	1.369	0.108	0.061	0.050	0.214	0.027				
SF31410	0.738	0.072	-0.017	0.102	0.192	0.037				
SF31414A	5.099	0.273	0.077	0.011	0.000	0.000				
SF31414B	3.961	0.197	0.070	0.012	0.000	0.000				
SF31418	0.891	0.100	0.810	0.066	0.147	0.024				
SF31421	0.515	0.033	-0.284	0.057	0.000	0.000				
SF31422	1.362	0.129	-0.185	0.073	0.410	0.031				
SF31427	1.022	0.101	0.323	0.068	0.246	0.028				
SF31445A	1.784	0.085	0.528	0.021	0.000	0.000				
SF31445B	1.392	0.065	-0.161	0.026	0.000	0.000				
SF31446A	1.455	0.073	0.649	0.026	0.000	0.000				
SF31446B	1.062	0.061	0.849	0.039	0.000	0.000				
SF31446C	0.906	0.048	0.204	0.033	0.000	0.000				
SF31447	0.574	0.033	1.156	0.054	0.000	0.000	0.332	0.054	-0.332	0.082

E

Appendix E

Summary Statistics and Standard Errors for Proficiency in Mathematics and Science Content Areas in the Eighth and Fourth Grades

Exhibit E.1 Summary Statistics and Standard Errors for Proficiency in Number in the Eighth Grade

Country	Sample Size	Number		Jackknife Sampling Error	Overall Standard Error
		Mean Proficiency	Standard Deviation		
Armenia	5726	473.267	79.505	3.109	3.126
Australia	4791	498.444	85.356	4.543	4.591
Bahrain	4199	380.319	80.907	1.628	1.887
Belgium (Flemish)	4970	539.216	68.211	2.462	2.666
Botswana	5150	382.318	68.160	1.939	2.183
Bulgaria	4117	476.678	83.568	3.979	4.108
Chile	6377	389.569	83.970	2.922	3.101
Chinese Taipei	5379	585.280	101.405	4.491	4.561
Cyprus	4002	463.727	83.227	1.330	1.492
Egypt	7095	420.540	85.114	2.986	3.011
England	2830	484.823	79.581	4.794	5.017
Estonia	4040	522.729	71.906	3.035	3.120
Ghana	5100	289.258	100.874	4.675	5.065
Hong Kong, SAR	4972	585.716	71.429	3.209	3.233
Hungary	3302	528.668	83.381	3.517	3.617
Indonesia	5762	421.155	85.931	4.498	4.581
Iran, Islamic Rep. of	4942	416.253	76.089	2.146	2.348
Israel	4318	503.578	84.842	3.239	3.317
Italy	4278	479.593	76.516	3.042	3.221
Japan	4856	556.710	89.474	2.180	2.343
Jordan	4489	413.374	94.335	4.276	4.409
Korea, Rep. of	5309	585.844	86.000	1.916	2.117
Latvia	3630	506.774	74.256	3.050	3.210
Lebanon	3814	429.979	68.779	3.112	3.260
Lithuania	4964	499.762	80.485	2.512	2.676
Macedonia, Rep. of	3893	437.596	80.003	3.408	3.485
Malaysia	5314	524.137	72.511	3.937	4.019
Moldova, Rep. of	4033	462.580	76.786	3.706	3.844
Morocco	2943	384.420	67.522	2.249	2.668
Netherlands	3065	538.549	68.309	3.519	3.562
New Zealand	3801	481.346	83.145	5.864	5.999
Norway	4133	455.986	73.523	2.038	2.266
Palestinian Nat'l Auth.	5357	385.293	95.520	3.336	3.597
Philippines	6917	393.469	87.108	5.048	5.101
Romania	4104	474.491	88.736	4.724	4.915
Russian Federation	4667	505.121	81.205	3.864	3.989
Saudi Arabia	4295	307.052	87.039	4.867	5.331
Scotland	3516	483.860	78.723	3.871	4.164
Serbia	4296	477.223	84.391	2.487	2.827
Singapore	6018	617.541	77.753	3.440	3.462
Slovak Republic	4215	514.232	82.868	3.266	3.338
Slovenia	3578	498.419	70.933	1.968	2.011
South Africa	8952	273.938	106.715	5.209	5.433
Sweden	4256	495.812	71.351	2.498	2.605
Tunisia	4931	419.390	61.945	2.044	2.272
United States	8912	507.636	80.968	3.258	3.352

Exhibit E.2 Summary Statistics and Standard Errors for Proficiency in Number in the Fourth Grade

Country	Sample Size	Number		Jackknife Sampling Error	Overall Standard Error
		Mean Proficiency	Standard Deviation		
Armenia	5674	473.328	84.044	2.779	2.958
Australia	4321	478.579	89.937	4.174	4.333
Belgium (Flemish)	4712	548.604	65.837	1.853	1.912
Chinese Taipei	4661	567.638	70.116	1.696	1.821
Cyprus	4328	513.697	90.674	2.391	2.652
England	3585	519.046	96.405	3.928	4.064
Hong Kong, SAR	4608	573.781	69.400	3.212	3.320
Hungary	3319	523.559	75.099	2.726	2.909
Iran, Islamic Rep. of	4352	410.107	78.901	3.682	3.742
Italy	4282	502.491	85.291	3.567	3.590
Japan	4535	555.833	83.143	1.711	2.024
Latvia	3687	530.908	76.392	2.527	2.630
Lithuania	4422	535.081	76.314	2.644	2.901
Moldova, Rep. of	3981	506.567	86.178	4.536	4.654
Morocco	4264	359.182	92.172	4.640	4.651
Netherlands	2937	536.190	63.546	2.113	2.246
New Zealand	4308	474.811	94.099	2.257	2.330
Norway	4342	440.423	86.541	1.996	2.203
Philippines	4572	380.199	102.046	7.360	7.413
Russian Federation	3963	531.660	79.149	4.512	4.589
Scotland	3936	475.008	86.036	3.122	3.303
Singapore	6668	612.257	95.779	5.951	6.002
Slovenia	3126	461.128	84.004	2.593	2.682
Tunisia	4334	360.350	98.732	4.084	4.133
United States	9829	516.383	84.557	2.511	2.639

Exhibit E.3 **Summary Statistics and Standard Errors for Proficiency in Algebra in the Eighth Grade**

Country	Sample Size	Algebra		Jackknife Sampling Error	Overall Standard Error
		Mean Proficiency	Standard Deviation		
Armenia	5726	489.164	90.308	2.562	2.626
Australia	4791	498.656	83.238	4.239	4.375
Bahrain	4199	410.653	84.809	1.492	2.523
Belgium (Flemish)	4970	523.137	75.603	2.699	2.757
Botswana	5150	376.764	78.683	2.102	2.730
Bulgaria	4117	480.740	84.273	3.932	3.952
Chile	6377	384.442	85.956	2.678	3.090
Chinese Taipei	5379	585.377	108.125	4.883	4.905
Cyprus	4002	455.256	85.869	1.488	1.663
Egypt	7095	407.710	102.031	3.768	3.904
England	2830	491.928	79.014	4.365	4.532
Estonia	4040	528.407	65.280	2.490	2.610
Ghana	5100	287.733	104.397	4.095	4.820
Hong Kong, SAR	4972	579.891	72.063	3.102	3.167
Hungary	3302	533.892	76.312	2.956	3.107
Indonesia	5762	418.125	87.359	4.139	4.481
Iran, Islamic Rep. of	4942	411.505	77.259	2.642	3.149
Israel	4318	497.562	86.094	3.140	3.170
Italy	4278	476.644	78.305	3.329	3.429
Japan	4856	567.844	80.205	1.925	2.015
Jordan	4489	434.032	92.925	4.133	4.442
Korea, Rep. of	5309	596.957	92.569	2.067	2.153
Latvia	3630	508.076	73.597	2.957	3.160
Lebanon	3814	447.559	66.871	2.991	3.113
Lithuania	4964	501.454	74.163	2.333	2.357
Macedonia, Rep. of	3893	441.903	96.056	3.631	3.649
Malaysia	5314	494.560	75.253	3.780	3.855
Moldova, Rep. of	4033	464.133	88.002	4.049	4.187
Morocco	2943	400.393	75.569	2.320	2.753
Netherlands	3065	513.750	75.453	3.914	3.993
New Zealand	3801	489.955	76.780	5.109	5.217
Norway	4133	428.212	80.696	2.477	2.713
Palestinian Nat'l Auth.	5357	392.103	99.279	3.418	3.508
Philippines	6917	400.315	94.143	5.093	5.237
Romania	4104	480.319	94.920	4.596	4.665
Russian Federation	4667	516.000	72.413	3.013	3.194
Saudi Arabia	4295	330.644	89.845	3.851	4.679
Scotland	3516	488.061	79.774	3.777	3.936
Serbia	4296	487.769	87.120	2.222	2.549
Singapore	6018	589.546	85.023	3.395	3.515
Slovak Republic	4215	504.689	79.361	3.174	3.273
Slovenia	3578	486.549	70.845	2.198	2.284
South Africa	8952	274.571	113.462	4.968	5.055
Sweden	4256	480.280	75.548	2.558	2.975
Tunisia	4931	404.628	66.670	2.057	2.446
United States	8912	509.947	77.481	3.067	3.107

Exhibit E.4 **Summary Statistics and Standard Errors for Proficiency in Patterns and Relationships in the Fourth Grade**

Country	Sample Size	Patterns and Relationships			
		Mean Proficiency	Standard Deviation	Jackknife Sampling Error	Overall Standard Error
Armenia	5674	460.537	106.827	3.912	4.095
Australia	4321	495.442	77.372	3.508	3.715
Belgium (Flemish)	4712	542.229	58.887	1.647	1.894
Chinese Taipei	4661	554.689	67.869	1.822	2.413
Cyprus	4328	518.944	82.079	2.301	2.420
England	3585	523.413	90.395	3.382	3.891
Hong Kong, SAR	4608	568.013	67.647	3.164	3.463
Hungary	3319	544.519	85.845	3.247	3.658
Iran, Islamic Rep. of	4352	394.170	90.984	3.801	3.912
Italy	4282	496.055	87.927	4.113	4.274
Japan	4535	553.964	68.983	1.299	1.448
Latvia	3687	531.718	78.409	2.583	3.359
Lithuania	4422	531.105	77.017	2.754	2.951
Moldova, Rep. of	3981	520.961	95.179	5.024	5.107
Morocco	4264	360.317	96.697	4.385	4.694
Netherlands	2937	527.422	53.077	2.000	2.399
New Zealand	4308	494.507	83.412	2.065	2.879
Norway	4342	438.981	86.763	2.207	2.671
Philippines	4572	382.114	109.018	6.917	7.031
Russian Federation	3963	530.806	77.418	4.458	4.962
Scotland	3936	494.813	71.953	2.715	2.892
Singapore	6668	578.702	85.143	5.295	5.414
Slovenia	3126	489.697	75.300	2.170	2.706
Tunisia	4334	330.042	117.258	4.551	4.745
United States	9829	523.722	73.942	2.290	2.658

Exhibit E.5 **Summary Statistics and Standard Errors for Proficiency in Measurement in the Eighth Grade**

Country	Sample Size	Measurement			Overall Standard Error
		Mean Proficiency	Standard Deviation	Jackknife Sampling Error	
Armenia	5726	488.173	94.024	2.988	3.277
Australia	4791	510.720	79.063	4.275	4.313
Bahrain	4199	388.454	88.081	1.579	2.109
Belgium (Flemish)	4970	534.762	68.471	2.454	2.525
Botswana	5150	377.298	70.558	1.932	2.029
Bulgaria	4117	472.564	87.808	4.494	4.623
Chile	6377	403.690	74.536	2.687	2.893
Chinese Taipei	5379	574.173	94.175	4.041	4.354
Cyprus	4002	459.015	86.395	1.496	2.204
Egypt	7095	400.810	91.796	3.075	3.335
England	2830	505.279	72.035	4.237	4.261
Estonia	4040	528.055	73.989	2.904	2.961
Ghana	5100	261.840	98.886	3.468	3.669
Hong Kong, SAR	4972	584.053	68.238	2.883	3.337
Hungary	3302	524.564	80.169	2.936	3.079
Indonesia	5762	394.008	98.268	4.759	4.940
Iran, Islamic Rep. of	4942	398.594	78.978	2.275	2.636
Israel	4318	480.449	82.980	3.222	3.440
Italy	4278	499.855	79.525	3.070	3.210
Japan	4856	559.021	73.550	1.900	1.989
Jordan	4489	417.787	88.606	3.733	4.371
Korea, Rep. of	5309	577.293	82.804	1.855	2.044
Latvia	3630	500.480	70.496	2.960	2.987
Lebanon	3814	429.708	72.622	2.783	3.674
Lithuania	4964	492.150	85.546	2.842	3.042
Macedonia, Rep. of	3893	434.070	89.149	3.523	3.637
Malaysia	5314	504.245	82.996	4.429	4.487
Moldova, Rep. of	4033	467.999	81.228	3.774	3.953
Morocco	2943	376.151	71.757	2.001	3.401
Netherlands	3065	548.558	69.632	3.569	3.719
New Zealand	3801	500.086	77.058	4.732	4.847
Norway	4133	480.990	68.057	2.228	2.882
Palestinian Nat'l Auth.	5357	385.511	91.722	2.721	2.762
Philippines	6917	371.507	83.053	4.702	4.811
Romania	4104	485.319	86.982	4.656	4.720
Russian Federation	4667	507.275	77.883	3.733	3.908
Saudi Arabia	4295	337.991	79.357	2.891	3.400
Scotland	3516	507.722	69.697	3.423	3.646
Serbia	4296	475.120	94.688	2.480	2.525
Singapore	6018	610.527	79.744	3.403	3.581
Slovak Republic	4215	507.633	86.279	3.607	3.737
Slovenia	3578	495.902	75.423	2.121	2.299
South Africa	8952	298.433	92.146	4.210	4.679
Sweden	4256	512.055	69.888	2.554	2.584
Tunisia	4931	406.762	72.902	2.080	2.234
United States	8912	495.230	78.655	3.099	3.175

Exhibit E.6 **Summary Statistics and Standard Errors for Proficiency in Measurement in the Fourth Grade**

Country	Sample Size	Measurement		Jackknife Sampling Error	Overall Standard Error
		Mean Proficiency	Standard Deviation		
Armenia	5674	465.283	78.798	3.093	3.138
Australia	4321	513.701	74.151	3.543	3.677
Belgium (Flemish)	4712	549.773	50.562	1.391	1.419
Chinese Taipei	4661	556.920	55.440	1.493	1.644
Cyprus	4328	505.571	79.548	2.218	2.343
England	3585	534.851	77.764	3.134	3.302
Hong Kong, SAR	4608	562.531	56.375	2.593	2.665
Hungary	3319	532.438	69.004	2.600	2.717
Iran, Islamic Rep. of	4352	398.380	84.321	3.053	3.181
Italy	4282	504.135	77.090	3.240	3.382
Japan	4535	567.905	62.440	1.223	1.576
Latvia	3687	544.758	65.287	2.337	2.628
Lithuania	4422	539.999	68.405	2.382	2.700
Moldova, Rep. of	3981	504.922	76.133	3.912	4.010
Morocco	4264	344.551	102.512	5.398	5.470
Netherlands	2937	544.562	53.222	1.701	2.181
New Zealand	4308	502.977	75.207	1.865	2.037
Norway	4342	474.896	75.000	1.827	2.166
Philippines	4572	330.135	119.368	7.672	7.778
Russian Federation	3963	538.153	71.515	3.762	3.844
Scotland	3936	499.440	70.080	2.769	3.083
Singapore	6668	566.461	73.011	4.575	4.639
Slovenia	3126	496.828	77.133	2.294	2.818
Tunisia	4334	308.145	128.241	5.282	5.479
United States	9829	499.696	69.350	2.104	2.133

Exhibit E.7 Summary Statistics and Standard Errors for Proficiency in Geometry in the Eighth Grade

Country	Sample Size	Geometry		Jackknife Sampling Error	Overall Standard Error
		Mean Proficiency	Standard Deviation		
Armenia	5726	480.773	72.312	2.890	3.100
Australia	4791	491.313	80.929	4.450	4.771
Bahrain	4199	437.985	72.013	1.356	2.085
Belgium (Flemish)	4970	527.484	79.194	2.770	3.067
Botswana	5150	334.878	88.400	2.800	3.864
Bulgaria	4117	484.279	84.512	4.219	4.513
Chile	6377	377.911	87.684	3.226	3.334
Chinese Taipei	5379	587.647	110.178	4.839	5.121
Cyprus	4002	457.068	76.195	1.393	2.418
Egypt	7095	407.955	103.483	3.581	3.616
England	2830	491.588	82.289	4.166	4.519
Estonia	4040	539.538	65.627	2.383	2.648
Ghana	5100	277.838	104.213	3.963	4.303
Hong Kong, SAR	4972	588.178	79.767	3.470	3.625
Hungary	3302	515.251	80.854	3.115	3.131
Indonesia	5762	413.173	88.823	4.183	4.606
Iran, Islamic Rep. of	4942	437.415	75.313	2.426	3.092
Israel	4318	487.692	85.826	3.247	3.651
Italy	4278	468.911	80.825	3.308	3.470
Japan	4856	586.640	80.239	1.940	2.112
Jordan	4489	446.245	80.542	3.649	3.955
Korea, Rep. of	5309	597.568	86.864	1.818	2.574
Latvia	3630	514.815	74.044	2.889	3.274
Lebanon	3814	459.003	66.327	2.563	3.016
Lithuania	4964	506.352	77.593	2.264	2.472
Macedonia, Rep. of	3893	441.632	87.512	3.113	3.731
Malaysia	5314	494.773	82.008	4.639	4.776
Moldova, Rep. of	4033	462.709	93.192	4.605	4.744
Morocco	2943	414.860	65.980	1.958	2.271
Netherlands	3065	512.837	74.252	3.868	4.129
New Zealand	3801	488.184	76.044	4.576	4.626
Norway	4133	460.915	69.545	2.138	2.757
Palestinian Nat'l Auth.	5357	422.976	88.744	2.838	3.104
Philippines	6917	344.473	87.804	4.844	5.326
Romania	4104	476.338	87.583	4.662	4.899
Russian Federation	4667	514.765	81.219	3.982	4.213
Saudi Arabia	4295	381.626	79.169	3.614	4.272
Scotland	3516	490.587	70.267	3.290	3.314
Serbia	4296	471.105	87.159	2.555	2.989
Singapore	6018	579.530	80.657	3.512	3.714
Slovak Republic	4215	501.210	85.464	3.522	3.626
Slovenia	3578	482.974	73.153	2.149	2.530
South Africa	8952	246.743	117.960	5.237	5.398
Sweden	4256	467.001	77.223	2.725	3.421
Tunisia	4931	427.475	61.331	1.915	2.044
United States	8912	471.992	74.399	2.749	3.120

Exhibit E.8 **Summary Statistics and Standard Errors for Proficiency in Geometry in the Fourth Grade**

Country	Sample Size	Geometry		Jackknife Sampling Error	Overall Standard Error
		Mean Proficiency	Standard Deviation		
Armenia	5674	430.756	100.807	3.495	3.791
Australia	4321	523.976	76.348	3.076	3.718
Belgium (Flemish)	4712	532.615	60.052	1.299	1.799
Chinese Taipei	4661	553.044	69.193	1.720	2.459
Cyprus	4328	504.950	73.497	2.053	2.298
England	3585	541.547	89.162	3.487	3.659
Hong Kong, SAR	4608	556.585	64.033	2.684	2.913
Hungary	3319	514.014	73.733	2.642	3.266
Iran, Islamic Rep. of	4352	415.769	85.733	3.745	3.884
Italy	4282	522.118	80.075	3.313	3.494
Japan	4535	559.078	76.230	1.478	1.882
Latvia	3687	522.819	51.885	1.806	2.193
Lithuania	4422	524.239	66.586	1.941	2.164
Moldova, Rep. of	3981	500.576	93.567	4.606	4.887
Morocco	4264	362.169	107.953	4.650	4.901
Netherlands	2937	520.631	63.614	2.430	3.154
New Zealand	4308	517.422	72.375	1.606	1.841
Norway	4342	477.819	77.603	1.711	2.164
Philippines	4572	335.112	142.055	8.551	8.816
Russian Federation	3963	528.268	82.528	4.626	4.791
Scotland	3936	511.091	68.371	2.291	2.464
Singapore	6668	569.790	103.649	5.419	5.454
Slovenia	3126	498.405	70.334	1.814	2.159
Tunisia	4334	346.363	121.433	4.394	5.109
United States	9829	517.962	73.306	1.902	2.157

Exhibit E.9 Summary Statistics and Standard Errors for Proficiency in Data in the Eighth Grade

Country	Sample Size	Data: Mean Proficiency	Data: Standard Deviation	Jackknife Sampling Error	Overall Standard Error
Armenia	5726	418.771	92.733	2.568	2.704
Australia	4791	531.324	77.134	3.681	3.784
Bahrain	4199	413.913	72.553	1.140	2.132
Belgium (Flemish)	4970	545.879	74.497	2.537	2.889
Botswana	5150	374.820	77.507	2.392	2.728
Bulgaria	4117	458.419	89.768	3.708	3.936
98145.451	6377	412.213	90.325	2.847	3.371
Chinese Taipei	5379	567.786	83.146	3.292	3.352
Cyprus	4002	458.009	73.606	1.465	1.713
Egypt	7095	393.420	82.696	2.887	3.170
England	2830	534.888	77.203	3.930	4.067
Estonia	4040	535.207	73.957	2.785	2.839
Ghana	5100	292.952	99.584	4.028	4.061
Hong Kong, SAR	4972	566.137	72.044	2.925	2.963
Hungary	3302	525.808	79.419	2.731	2.932
Indonesia	5762	418.498	84.763	3.682	4.045
Iran, Islamic Rep. of	4942	404.296	80.772	2.480	2.590
Israel	4318	491.537	93.043	3.194	3.320
Italy	4278	490.058	77.600	2.647	2.974
Japan	4856	572.656	71.782	1.757	1.877
Jordan	4489	430.241	81.647	3.020	3.459
Korea, Rep. of	5309	568.995	72.148	1.437	1.975
Latvia	3630	506.401	80.263	3.258	3.827
Lebanon	3814	393.836	82.505	3.408	4.004
Lithuania	4964	501.897	81.108	2.179	2.512
Macedonia, Rep. of	3893	418.819	98.175	3.542	3.588
Malaysia	5314	505.007	64.203	3.118	3.235
Moldova, Rep. of	4033	428.005	78.777	3.123	3.433
Morocco	2943	373.654	79.352	2.142	2.457
Netherlands	3065	560.019	69.502	3.097	3.119
New Zealand	3801	525.944	79.020	5.019	5.143
Norway	4133	498.263	80.811	2.378	2.471
Palestinian Nat'lAuth.	5357	390.369	83.415	2.484	2.834
Philippines	6917	390.435	78.848	4.113	4.453
Romania	4104	445.307	92.067	4.289	4.551
Russian Federation	4667	484.062	76.984	3.086	3.207
Saudi Arabia	4295	338.687	84.091	3.544	3.809
Scotland	3516	531.169	77.077	3.334	3.665
Serbia	4296	456.210	91.982	2.388	2.576
Singapore	6018	579.491	78.307	3.152	3.198
Slovak Republic	4215	495.278	85.922	2.782	2.899
Slovenia	3578	493.812	76.455	2.212	2.301
South Africa	8952	296.246	110.593	4.944	5.319
Sweden	4256	539.423	79.890	2.866	2.950
Tunisia	4931	386.850	71.829	1.713	2.172
United States	8912	526.691	79.301	3.012	3.164

Exhibit E.10 **Summary Statistics and Standard Errors for Proficiency in Data in the Fourth Grade**

Country	Sample Size	Data		Jackknife Sampling Error	Overall Standard Error
		Mean Proficiency	Standard Deviation		
Armenia	5674	416.987	81.117	3.161	3.609
Australia	4321	524.754	71.216	3.418	3.614
Belgium (Flemish)	4712	548.013	59.057	1.599	2.167
Chinese Taipei	4661	563.826	61.406	1.558	2.309
Cyprus	4328	509.193	74.667	2.097	2.295
England	3585	551.513	77.268	2.943	3.413
Hong Kong, SAR	4608	561.875	46.191	1.977	2.253
Hungary	3319	513.197	71.992	3.003	3.160
Iran, Islamic Rep. of	4352	356.402	105.500	4.311	4.371
Italy	4282	496.763	68.422	2.753	2.991
Japan	4535	592.823	81.730	1.578	1.637
Latvia	3687	525.532	71.531	2.483	2.701
Lithuania	4422	517.399	69.538	2.485	2.545
Moldova, Rep. of	3981	476.500	76.839	4.088	4.262
Morocco	4264	355.276	89.038	4.614	5.012
Netherlands	2937	552.893	52.956	1.812	2.434
New Zealand	4308	521.631	82.056	1.851	1.972
Norway	4342	479.085	86.295	2.123	2.279
Philippines	4572	383.997	110.383	7.330	7.463
Russian Federation	3963	505.067	64.951	3.892	4.050
Scotland	3936	515.899	68.288	2.375	2.702
Singapore	6668	575.070	60.984	3.652	3.889
Slovenia	3126	486.244	65.844	2.372	2.739
Tunisia	4334	308.240	107.923	4.514	4.650
United States	9829	548.671	68.144	1.885	2.039

Exhibit E.11 **Summary Statistics and Standard Errors for Proficiency in Life Science in the Eighth Grade**

Country	Sample Size	Life Science		Jackknife Sampling Error	Overall Standard Error
		Mean Proficiency	Standard Deviation		
Armenia	5726	452.987	84.319	3.179	3.303
Australia	4791	532.408	75.249	3.614	3.768
Bahrain	4199	444.590	74.040	1.523	1.865
Belgium (Flemish)	4970	525.827	69.487	2.242	2.369
Botswana	5150	369.942	91.757	2.635	2.731
Bulgaria	4117	474.154	95.259	5.045	5.160
Chile	6377	426.770	84.360	2.646	2.729
Chinese Taipei	5379	562.553	73.036	3.044	3.120
Cyprus	4002	436.785	85.134	1.812	2.221
Egypt	7095	425.244	99.625	3.615	3.739
England	2830	543.037	74.459	3.752	3.902
Estonia	4040	546.508	63.890	2.269	2.427
Ghana	5100	256.392	124.017	5.382	5.604
Hong Kong, SAR	4972	551.244	62.267	2.805	2.940
Hungary	3302	536.455	70.137	2.633	2.695
Indonesia	5762	423.610	74.770	3.593	3.883
Iran, Islamic Rep. of	4942	446.729	68.980	2.213	2.599
Israel	4318	490.941	86.316	3.017	3.042
Italy	4278	497.596	80.728	2.916	3.247
Japan	4856	549.342	70.356	1.663	2.017
Jordan	4489	474.942	86.706	3.629	3.991
Korea, Rep. of	5309	558.424	66.645	1.440	1.560
Latvia	3630	511.283	65.869	2.394	2.534
Lebanon	3814	359.968	107.157	4.784	4.972
Lithuania	4964	516.944	71.688	2.321	2.396
Macedonia, Rep. of	3893	448.126	94.952	3.737	3.832
Malaysia	5314	504.265	66.440	3.638	3.693
Moldova, Rep. of	4033	465.955	72.721	3.229	3.679
Morocco	2943	389.621	80.109	2.446	2.650
Netherlands	3065	536.420	61.812	3.075	3.290
New Zealand	3801	523.192	77.440	5.059	5.146
Norway	4133	495.516	75.616	2.200	2.460
Palestinian Nat'l Auth.	5357	434.984	84.075	3.092	3.606
Philippines	6917	386.977	108.993	5.753	5.819
Romania	4104	471.215	90.979	4.751	4.791
Russian Federation	4667	514.092	76.750	3.255	3.284
Saudi Arabia	4295	411.604	69.577	3.413	3.915
Scotland	3516	512.326	76.714	3.185	3.312
Serbia	4296	468.196	83.442	2.472	2.580
Singapore	6018	568.671	87.670	3.946	4.020
Slovak Republic	4215	513.592	72.358	2.758	2.946
Slovenia	3578	520.802	69.395	1.981	2.223
South Africa	8952	250.120	130.677	5.871	5.958
Sweden	4256	527.809	77.000	2.655	2.709
Tunisia	4931	417.391	60.779	1.873	1.969
United States	8912	536.992	82.372	2.952	3.007

Exhibit E.12 **Summary Statistics and Standard Errors for Proficiency in Life Science in the Fourth Grade**

Country	Sample Size	Life Science		Jackknife Sampling Error	Overall Standard Error
		Mean Proficiency	Standard Deviation		
Armenia	5674	435.430	103.424	4.299	4.365
Australia	4321	523.337	75.827	3.834	3.845
Belgium (Flemish)	4712	523.708	55.555	1.567	1.733
Chinese Taipei	4661	540.346	59.459	1.355	1.560
Cyprus	4328	482.330	70.652	2.039	2.104
England	3585	531.827	76.461	2.983	3.106
Hong Kong, SAR	4608	534.657	58.283	2.532	2.557
Hungary	3319	536.362	74.227	2.440	2.521
Iran, Islamic Rep. of	4352	423.809	90.406	4.213	4.560
Italy	4282	521.048	79.174	3.485	3.494
Japan	4535	529.763	65.393	1.076	1.306
Latvia	3687	530.716	62.182	2.198	2.276
Lithuania	4422	516.335	58.103	1.893	1.951
Moldova, Rep. of	3981	503.702	81.576	3.854	3.925
Morocco	4264	299.842	132.612	5.753	6.122
Netherlands	2937	547.105	54.141	1.717	1.819
New Zealand	4308	520.033	81.206	2.193	2.272
Norway	4342	479.867	77.406	2.183	2.234
Philippines	4572	330.343	138.693	8.802	8.983
Russian Federation	3963	526.146	72.622	4.666	4.726
Scotland	3936	505.629	73.858	2.783	3.131
Singapore	6668	557.581	79.970	5.002	5.045
Slovenia	3126	488.638	75.461	2.525	2.904
Tunisia	4334	289.669	141.489	5.891	5.941
United States	9829	536.996	76.386	2.161	2.171

Exhibit E.13 **Summary Statistics and Standard Errors for Proficiency in Chemistry in the Eighth Grade**

Country	Sample Size	Chemistry		Jackknife Sampling Error	Overall Standard Error
		Mean Proficiency	Standard Deviation		
Armenia	5726	465.612	102.835	4.070	4.201
Australia	4791	506.347	71.621	3.779	3.806
Bahrain	4199	441.296	77.552	1.548	2.649
Belgium (Flemish)	4970	502.576	56.551	1.940	2.014
Botswana	5150	348.198	100.597	2.908	3.126
Bulgaria	4117	482.440	96.228	5.334	5.709
Chile	6377	404.909	91.194	2.919	3.267
Chinese Taipei	5379	583.713	92.897	3.896	3.994
Cyprus	4002	442.726	79.782	1.616	2.647
Egypt	7095	441.566	106.786	3.644	3.814
England	2830	527.158	81.031	4.156	4.198
Estonia	4040	551.590	59.554	2.014	2.127
Ghana	5100	275.668	134.164	5.578	6.552
Hong Kong, SAR	4972	541.873	53.921	2.444	2.555
Hungary	3302	559.987	77.925	2.796	3.072
Indonesia	5762	391.239	80.490	3.508	3.848
Iran, Islamic Rep. of	4942	445.277	83.331	2.470	2.715
Israel	4318	499.488	82.575	2.731	3.407
Italy	4278	486.928	75.820	2.943	3.303
Japan	4856	552.241	63.019	1.574	2.107
Jordan	4489	477.592	96.889	3.643	4.392
Korea, Rep. of	5309	528.840	67.060	1.542	2.506
Latvia	3630	513.700	72.904	3.078	3.204
Lebanon	3814	433.490	91.751	4.012	4.877
Lithuania	4964	533.996	71.536	2.144	2.335
Macedonia, Rep. of	3893	466.684	100.045	3.664	3.861
Malaysia	5314	513.611	63.774	3.390	3.764
Moldova, Rep. of	4033	478.615	83.602	3.511	3.943
Morocco	2943	401.934	72.549	2.328	2.748
Netherlands	3065	514.421	49.978	2.361	2.623
New Zealand	3801	500.668	75.073	5.015	5.605
Norway	4133	484.554	59.164	1.834	3.008
Palestinian Nat'l Auth.	5357	444.464	101.417	3.402	3.924
Philippines	6917	342.004	113.497	5.877	6.073
Romania	4104	474.135	97.035	4.852	4.926
Russian Federation	4667	527.165	80.476	3.846	3.987
Saudi Arabia	4295	381.756	83.625	4.243	4.774
Scotland	3516	498.871	73.060	3.128	3.214
Serbia	4296	473.951	90.876	2.582	3.193
Singapore	6018	582.455	93.609	4.147	4.198
Slovak Republic	4215	519.321	76.986	3.147	3.645
Slovenia	3578	531.858	70.713	1.916	2.570
South Africa	8952	285.156	115.565	5.190	5.944
Sweden	4256	526.082	62.466	2.347	2.556
Tunisia	4931	413.383	63.943	1.698	2.526
United States	8912	512.699	78.001	2.955	3.183

Exhibit E.14 **Summary Statistics and Standard Errors for Proficiency in Physics in the Eighth Grade**

Country	Sample Size	Physics		Jackknife Sampling Error	Overall Standard Error
		Mean Proficiency	Standard Deviation		
Armenia	5726	479.279	73.659	3.141	3.230
Australia	4791	521.169	72.150	3.594	3.746
Bahrain	4199	443.265	82.295	1.737	1.968
Belgium (Flemish)	4970	513.506	62.509	2.162	2.504
Botswana	5150	371.459	89.168	2.631	3.196
Bulgaria	4117	485.110	88.538	4.757	4.951
Chile	6377	400.703	81.344	2.727	3.076
Chinese Taipei	5379	569.215	73.925	3.092	3.284
Cyprus	4002	449.522	76.224	1.545	1.743
Egypt	7095	413.711	108.129	3.926	4.127
England	2830	544.673	69.098	3.424	3.484
Estonia	4040	544.496	62.543	2.080	2.353
Ghana	5100	239.154	125.028	5.287	5.371
Hong Kong, SAR	4972	554.974	65.468	2.687	2.779
Hungary	3302	536.145	75.387	2.588	2.671
Indonesia	5762	430.013	83.677	3.873	4.039
Iran, Islamic Rep. of	4942	445.161	82.336	2.437	2.990
Israel	4318	483.882	83.372	2.842	2.927
Italy	4278	470.370	81.045	3.051	3.163
Japan	4856	563.776	70.331	1.637	1.875
Jordan	4489	465.099	93.001	3.586	3.847
Korea, Rep. of	5309	578.661	67.520	1.407	1.563
Latvia	3630	511.950	64.046	2.133	2.377
Lebanon	3814	418.801	81.557	3.260	3.972
Lithuania	4964	519.343	61.788	2.042	2.662
Macedonia, Rep. of	3893	457.850	82.112	2.833	3.057
Malaysia	5314	519.323	65.686	3.429	3.637
Moldova, Rep. of	4033	478.759	75.438	3.330	3.743
Morocco	2943	409.840	73.189	2.410	2.655
Netherlands	3065	538.346	61.140	2.998	3.406
New Zealand	3801	515.377	65.431	4.509	4.706
Norway	4133	487.740	68.022	2.115	2.563
Palestinian Nat'l Auth.	5357	432.184	100.381	3.359	3.579
Philippines	6917	380.461	92.748	4.629	4.728
Romania	4104	472.830	83.044	3.983	4.052
Russian Federation	4667	511.444	73.356	3.395	3.450
Saudi Arabia	4295	394.032	80.422	3.730	3.871
Scotland	3516	515.026	67.740	2.915	3.029
Serbia	4296	470.908	86.545	2.284	2.630
Singapore	6018	578.558	77.566	3.339	3.403
Slovak Republic	4215	519.022	72.230	2.570	2.930
Slovenia	3578	508.840	58.455	1.596	1.802
South Africa	8952	244.189	135.112	6.103	6.207
Sweden	4256	524.573	74.067	2.667	2.869
Tunisia	4931	385.681	76.073	2.193	2.526
United States	8912	515.324	74.698	2.800	2.930

Exhibit E.15 **Summary Statistics and Standard Errors for Proficiency in Physical Science in the Fourth Grade**

Country	Sample Size	Physical Science		Jackknife Sampling Error	Overall Standard Error
		Mean Proficiency	Standard Deviation		
Armenia	5674	429.183	100.500	4.181	4.274
Australia	4321	517.951	80.712	3.854	3.885
Belgium (Flemish)	4712	507.102	56.861	1.641	2.273
Chinese Taipei	4661	553.988	78.602	1.954	2.034
Cyprus	4328	479.128	80.981	2.225	2.287
England	3585	546.379	80.070	3.204	3.248
Hong Kong, SAR	4608	547.513	60.248	2.664	2.711
Hungary	3319	526.355	75.392	2.519	2.678
Iran, Islamic Rep. of	4352	418.515	101.626	4.321	4.468
Italy	4282	511.860	83.371	3.442	3.543
Japan	4535	557.287	78.186	1.543	1.721
Latvia	3687	531.607	69.904	2.519	2.611
Lithuania	4422	512.251	67.170	2.120	2.475
Moldova, Rep. of	3981	488.758	82.616	3.879	3.925
Morocco	4264	307.951	127.849	6.819	7.001
Netherlands	2937	505.069	52.522	1.829	1.879
New Zealand	4308	516.083	84.557	2.185	2.337
Norway	4342	455.895	77.696	1.919	2.295
Philippines	4572	343.008	144.669	9.156	9.588
Russian Federation	3963	526.551	82.559	5.053	5.151
Scotland	3936	502.507	74.195	2.543	2.647
Singapore	6668	577.257	94.871	5.881	5.891
Slovenia	3126	496.784	71.577	2.285	2.345
Tunisia	4334	324.191	130.841	5.207	5.255
United States	9829	531.091	77.369	2.227	2.267

Exhibit E.16 **Summary Statistics and Standard Errors for Proficiency in Earth Science in the Eighth Grade**

Country	Sample Size	Earth Science		Jackknife Sampling Error	Overall Standard Error
		Mean Proficiency	Standard Deviation		
Armenia	5726	459.668	90.703	3.564	3.692
Australia	4791	531.122	74.100	3.728	4.233
Bahrain	4199	440.493	69.331	1.236	2.419
Belgium (Flemish)	4970	508.081	68.467	2.410	2.517
Botswana	5150	360.761	91.307	2.410	3.147
Bulgaria	4117	490.590	93.483	4.752	4.876
Chile	6377	435.172	76.091	2.500	3.104
Chinese Taipei	5379	548.021	73.234	2.656	3.080
Cyprus	4002	446.831	77.578	1.548	2.088
Egypt	7095	403.353	113.591	3.830	4.391
England	2830	544.196	79.252	3.836	4.110
Estonia	4040	558.191	69.759	2.380	2.931
Ghana	5100	254.473	119.906	5.420	5.592
Hong Kong, SAR	4972	548.516	64.045	2.569	2.855
Hungary	3302	537.374	76.283	2.531	3.082
Indonesia	5762	430.747	79.439	3.672	3.821
Iran, Islamic Rep. of	4942	467.697	80.090	2.374	2.889
Israel	4318	484.991	80.039	2.692	3.025
Italy	4278	513.257	75.911	2.932	3.174
Japan	4856	530.325	67.245	1.721	2.100
Jordan	4489	472.090	74.362	3.131	3.961
Korea, Rep. of	5309	539.964	69.187	1.590	1.903
Latvia	3630	514.324	73.650	2.529	2.846
Lebanon	3814	394.578	89.658	3.771	4.041
Lithuania	4964	512.313	77.688	2.443	2.665
Macedonia, Rep. of	3893	440.452	99.928	3.834	4.316
Malaysia	5314	501.767	61.602	3.172	3.811
Moldova, Rep. of	4033	474.673	79.560	3.651	4.023
Morocco	2943	396.923	80.858	2.554	3.438
Netherlands	3065	533.877	59.591	3.102	3.164
New Zealand	3801	524.739	73.061	4.612	4.835
Norway	4133	516.537	71.811	2.491	2.693
Palestinian Nat'l Auth.	5357	438.805	80.530	2.553	3.028
Philippines	6917	376.533	110.759	5.644	5.705
Romania	4104	468.655	96.658	4.883	5.157
Russian Federation	4667	517.725	80.893	3.229	3.297
Saudi Arabia	4295	393.985	84.503	3.725	3.969
Scotland	3516	515.171	77.586	3.466	3.780
Serbia	4296	471.254	90.929	2.700	2.998
Singapore	6018	548.955	88.368	3.746	3.859
Slovak Republic	4215	523.228	83.599	3.169	3.321
Slovenia	3578	523.490	74.118	1.661	2.228
South Africa	8952	247.256	131.254	6.219	6.251
Sweden	4256	532.018	68.798	2.475	3.337
Tunisia	4931	407.803	61.401	1.535	2.022
United States	8912	531.958	78.526	2.858	2.910

Exhibit E.17 **Summary Statistics and Standard Errors for Proficiency in Earth Science in the Fourth Grade**

Country	Sample Size	Earth Science Mean Proficiency	Earth Science Standard Deviation	Jackknife Sampling Error	Overall Standard Error
Armenia	5674	449.879	84.355	3.579	3.634
Australia	4321	518.244	79.475	3.962	4.133
Belgium (Flemish)	4712	522.247	51.449	1.451	1.662
Chinese Taipei	4661	559.130	85.741	1.963	2.555
Cyprus	4328	487.048	73.140	2.283	2.544
England	3585	535.364	85.401	3.396	3.518
Hong Kong, SAR	4608	536.091	62.818	2.611	2.692
Hungary	3319	525.716	94.961	3.484	3.694
Iran, Islamic Rep. of	4352	428.096	87.072	2.900	3.007
Italy	4282	518.685	88.153	3.514	3.696
Japan	4535	534.579	78.134	1.683	1.891
Latvia	3687	534.154	70.726	2.495	2.922
Lithuania	4422	503.326	79.021	2.670	3.187
Moldova, Rep. of	3981	505.218	93.678	4.667	4.876
Morocco	4264	310.522	127.167	6.023	6.076
Netherlands	2937	502.865	75.109	2.193	2.272
New Zealand	4308	522.448	79.352	1.943	2.345
Norway	4342	472.531	96.010	2.377	2.751
Philippines	4572	323.999	152.023	9.131	9.207
Russian Federation	3963	527.221	93.505	5.875	5.975
Scotland	3936	498.056	74.558	2.582	2.592
Singapore	6668	537.778	87.719	5.012	5.212
Slovenia	3126	490.207	77.972	2.507	2.699
Tunisia	4334	336.195	124.365	4.680	4.796
United States	9829	534.850	83.083	2.340	2.505

Exhibit E.18 **Summary Statistics and Standard Errors for Proficiency in Environmental Science in the Eighth Grade**

Country	Sample Size	Environmental Science		Jackknife Sampling Error	Overall Standard Error
		Mean Proficiency	Standard Deviation		
Armenia	5726	416.753	105.967	4.024	4.365
Australia	4791	535.686	69.987	3.242	3.391
Bahrain	4199	438.832	88.113	1.559	3.145
Belgium (Flemish)	4970	522.824	76.228	2.518	2.694
Botswana	5150	380.758	103.741	2.676	3.298
Bulgaria	4117	463.507	96.901	4.509	5.005
Chile	6377	435.551	80.628	2.384	2.934
Chinese Taipei	5379	559.681	63.187	2.525	3.109
Cyprus	4002	440.632	90.760	1.857	2.302
Egypt	7095	429.995	106.365	3.579	4.030
England	2830	539.639	77.584	3.336	4.236
Estonia	4040	539.576	67.702	2.033	2.242
Ghana	5100	267.357	136.114	5.812	6.209
Hong Kong, SAR	4972	555.476	64.323	2.460	2.567
Hungary	3302	527.643	79.651	2.747	2.936
Indonesia	5762	453.887	78.971	3.318	3.414
Iran, Islamic Rep. of	4942	486.574	66.515	1.777	2.123
Israel	4318	486.210	84.331	2.484	2.905
Italy	4278	496.895	78.638	2.800	3.024
Japan	4856	536.744	70.759	1.709	1.972
Jordan	4489	492.397	91.950	3.138	3.238
Korea, Rep. of	5309	543.517	63.430	1.252	1.448
Latvia	3630	507.776	70.917	2.340	3.305
Lebanon	3814	374.494	116.453	4.723	5.056
Lithuania	4964	506.784	70.706	1.964	2.004
Macedonia, Rep. of	3893	442.411	99.375	3.465	3.652
Malaysia	5314	512.898	63.332	3.126	3.157
Moldova, Rep. of	4033	454.075	97.123	3.454	3.782
Morocco	2943	395.911	97.067	3.099	3.315
Netherlands	3065	538.726	61.812	2.694	2.845
New Zealand	3801	525.490	67.388	3.827	3.925
Norway	4133	495.662	68.549	2.062	2.235
Palestinian Nat'l Auth.	5357	444.190	101.761	3.362	3.703
Philippines	6917	403.486	107.450	5.196	5.374
Romania	4104	472.002	88.353	4.365	4.719
Russian Federation	4667	491.028	76.883	2.775	3.185
Saudi Arabia	4295	410.218	79.930	3.342	3.766
Scotland	3516	510.847	80.984	3.398	3.486
Serbia	4296	457.358	83.984	2.298	2.423
Singapore	6018	567.541	85.614	3.676	3.828
Slovak Republic	4215	508.600	71.164	2.630	2.758
Slovenia	3578	515.395	65.368	1.676	2.242
South Africa	8952	260.934	141.225	6.418	6.600
Sweden	4256	499.405	73.884	2.332	2.575
Tunisia	4931	435.711	69.777	1.772	2.181
United States	8912	533.048	76.347	2.690	2.926