TIMSS advanced 2008 technical report

Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, 2009

TRENDS IN INTERNATIONAL MATHEMATICS AND SCIENCE STUDY

TIMSS

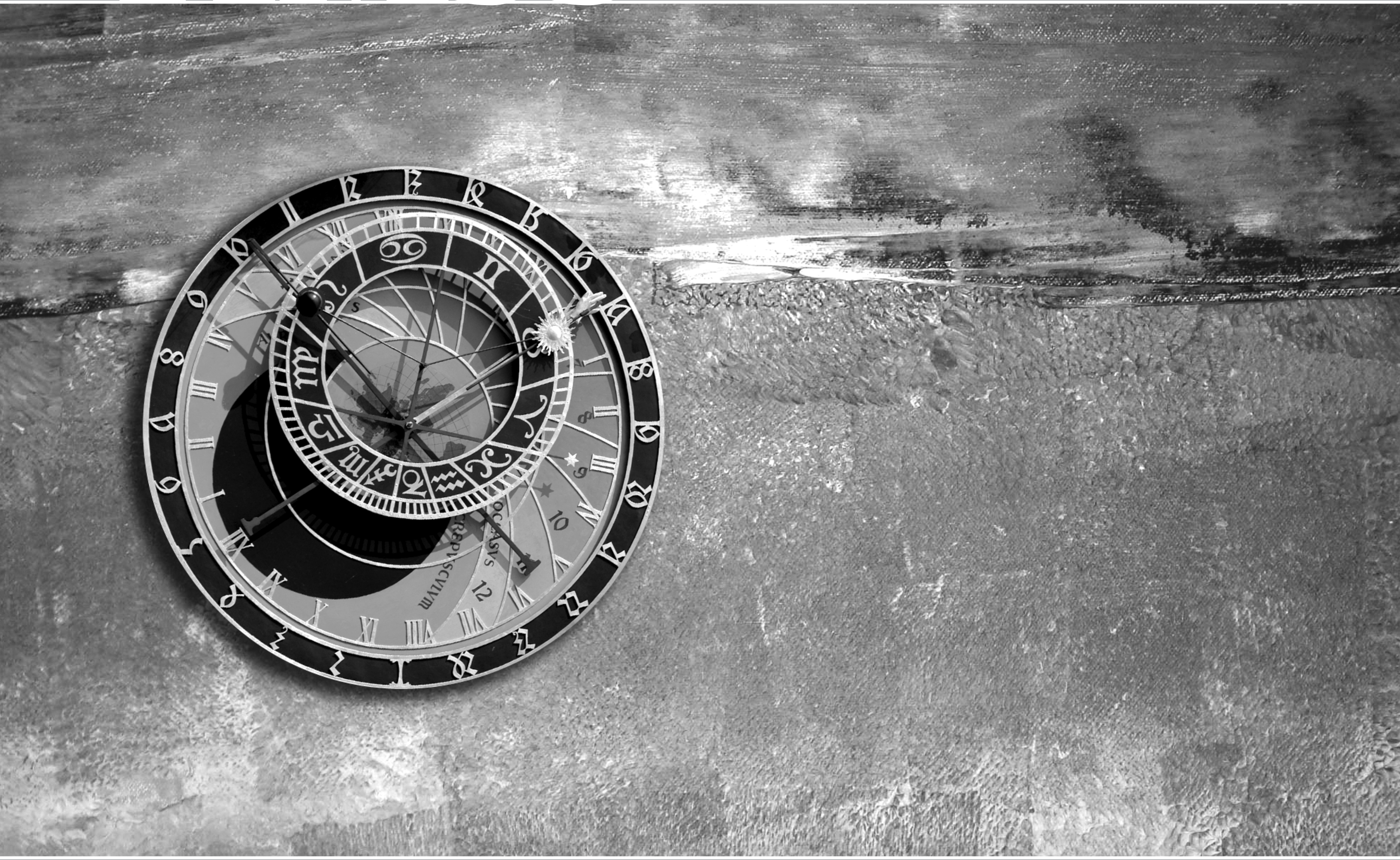
TIMSS Advanced 2008 Technical Report

Edited by Alka Arora, Pierre Foy, Michael O. Martin, and Ina V.S. Mullis

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

TRENDS IN INTERNATIONAL MATHEMATICS AND SCIENCE STUDY

TIMSS



TIMSS Advanced 2008 Technical Report

Edited by Alka Arora, Pierre Foy, Michael O. Martin, and Ina V.S. Mullis

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

Published December 2009

TIMSS Advanced 2008 Technical Report

Edited by Alka Arora, Pierre Foy, Michael O. Martin, and Ina V.S. Mullis

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

ISBN: 1-889938-56-4

For more information about TIMSS contact:

TIMSS & PIRLS International Study Center

Lynch School of Education

Boston College

Chestnut Hill, MA 02467

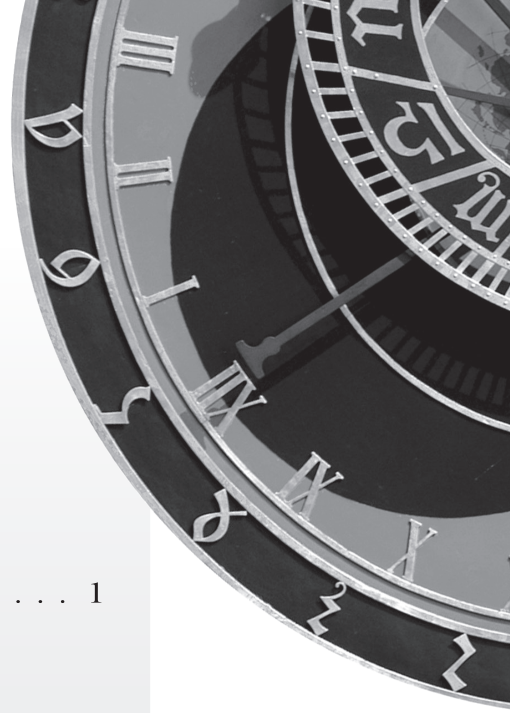
United States

tel: +1-617-552-1600

fax: +1-617-552-1203

e-mail: timss@bc.edu

http://timssandpirls.bc.edu



Contents

Chapter 1

Overview of TIMSS Advanced 2008 Procedures for Ensuring Comparative Validity

Michael O. Martin and Ina V.S. Mullis

1.1 Introduction

The design, development, and implementation of TIMSS Advanced 2008 are documented in a series of publications produced at various stages of the project. The *TIMSS Advanced 2008 Assessment Frameworks* (Garden, Lie, Robitaille, Angell, Martin, Mullis, Foy, & Arora, 2006) contains the advanced mathematics and physics frameworks underlying the assessments and describes the assessment design. The findings of the study are presented in the *TIMSS Advanced 2008 International Report* (Mullis, Martin, Robitaille, & Foy, 2009) for the 10 participating countries. The complete TIMSS Advanced 2008 international database is available on DVD accompanied by the *TIMSS Advanced 2008 User Guide* (Foy & Arora, 2009). The DVD also contains the *TIMSS Advanced 2008 Assessment Frameworks*, the *TIMSS Advanced 2008 International Report*, and the present publication, the *TIMSS Advanced 2008 Technical Report*.

As a high-profile international assessment of trends in student achievement in mathematics and science, TIMSS seeks to apply the highest standards of educational measurement throughout. Underpinning this endeavor are the fundamental measurement concerns of reliability and validity, but as an international study dealing with international comparisons of student achievement, TIMSS Advanced also must have comparative validity. For comparative validity, the essential concerns of reliability and validity still apply, but the concepts are extended to encompass the idea that the data should be internationally comparable. That is, that inferences made about achievement differences between countries can be substantiated.

Demonstrating its commitment to comparative validity, the various chapters of the *TIMSS Advanced 2008 Technical Report* document the steps taken by TIMSS Advanced 2008 to ensure high quality comparative data by providing details of the processes underlying the development of the TIMSS Advanced 2008 instruments and the methods used in sampling, data collection, scaling, and data analysis. In particular, the report documents the numerous steps and procedures that comprise the rigorous quality assurance program conducted by all those involved, including the TIMSS & PIRLS International Study Center, the IEA Secretariat, the IEA Data Processing and Research Center, Statistics Canada, Educational Testing Service, and the National Research Coordinators and their teams in the participating countries.

1.2 TIMSS Advanced 2008 Instruments

A valid international assessment of advanced mathematics and physics requires unified agreement in conceptualizing and articulating the constructs of advanced mathematics and physics as they apply to the programs and tracks that constitute the target populations of the study, and unified agreement that the items included in the assessments measure these articulations of advanced mathematics and physics, respectively. Chapter 2 describes how the *TIMSS Advanced 2008 Frameworks* were developed through a process of collaboration among the participating countries, including iterative reviews by the National Research Coordinators and experts. Chapter 2 also describes how the items and scoring guides were developed in accordance with the frameworks to assess specified topics, and according to a careful plan for measuring trends. The items were reviewed extensively by experts and the participating countries. Developing the instruments and operational procedures for TIMSS Advanced 2008 involved a full-scale field test that was essential for confirming the appropriateness and comparability of the items.

The TIMSS Advanced 2008 assessment contained 72 items in advanced mathematics and 71 in physics. For the advanced mathematics assessment, the items were assembled into seven blocks of items, and then the blocks were combined into four booklets, each one consisting of three blocks of advanced mathematics items assembled according to a rotated design. Each student was administered a single booklet. The physics assessment followed a similar plan, with seven blocks of physics items assembled into four student booklets, following the same rotational design as in advanced mathematics. Details about the development process and types of items can be found in Chapter 2.

Chapter 2 also contains information about developing the four different types of contextual questionnaires used in TIMSS Advanced 2008. In both the advanced mathematics and physics assessments, students completed a student questionnaire with questions pertaining to their home and school environments, educational aspirations, and motivation for studying advanced mathematics or physics. The advanced mathematics and physics teachers of the sampled students responded to questions about the school environment, characteristics of the class tested, instructional activities for teaching advanced mathematics or physics, the topics covered in students' lessons, calculator and computer use, homework and assessment, and their education, training, and opportunities for professional development. The principals of schools responded to questions about enrollment and school characteristics, school climate for learning, and school staffing and resources. National Research Coordinators were responsible for completing a curriculum questionnaire for each assessment, providing data about the country's curriculum for advanced mathematics and physics.

To increase reliability in reporting background data, the questions in the background questionnaires formed a number of scales. These scales and other sets of background questions were used to create background indices for reporting.

1.3 Translation Verification

Chapter 3 describes the steps involved in translating the test instruments and background questionnaires from English into the languages of the participating countries for the field test and for the main assessment. To ensure comparability among translated instruments, participants were given detailed specifications of the process to use in translating the materials, the IEA Secretariat managed rigorous translation verification procedures using external verifiers, and the TIMSS & PIRLS International Study Center conducted a verification of final instrument layout before instruments were printed. Every effort was made to ensure that the translations were comparable across countries.

1.4 Sample Design and Implementation

As explained in Chapter 4, the TIMSS Advanced 2008 assessment was administered to carefully drawn probability samples of students from the advanced mathematics and physics target populations in each country. Countries chose their target populations in terms of the programs or tracks that provided instruction in the mathematics and physics content described in the *TIMSS Advanced 2008 Assessment Frameworks*. These target populations overlapped to some extent in all countries, with many students belonging to both populations.[1] To assist in interpreting achievement differences among countries, TIMSS Advanced 2008 developed a coverage index for advanced mathematics and a coverage index for physics to quantify the proportion of the school-leaving age cohort taking these courses and included in the target populations in each country. Presented in conjunction with achievement results, these coverage indices remind the reader that the advanced mathematics or physics students assessed by TIMSS Advanced in each country represent relatively small and select proportions of the age cohort corresponding to the final year of secondary school.

TIMSS Advanced 2008 employed a uniform sample design that could be adapted to the specific sampling requirements of individual countries. The basic design was a two-stage stratified cluster sample, with the first stage consisting of schools, and the second stage consisting of one or more intact classrooms from the list of eligible classes in the sampled schools. Typically for each population (advanced mathematics and physics), countries sampled 120 schools and one classroom, although how this was implemented depended on school organization for teaching advanced mathematics and physics, as described in the sampling chapter. Each country worked closely with Statistics Canada to tailor the basic design to its particular situation so

1 The target populations in Armenia, Iran, and Lebanon overlapped completely, with all students in the target program/track taking both advanced mathematics and physics.

as to ensure the most effective coverage of the target populations while maximizing the comparability across countries of national samples.

Information about the sampling weights and documentation of the participation rates also is presented in Chapter 4. Countries were very successful in assessing the appropriate programs or tracks, including all students in their definition of the target population, keeping exclusions to a minimum (lower than 5%), and implementing accurate classroom sampling using the WinW3S software developed by IEA and Statistics Canada for this purpose. Almost all countries achieved the minimum acceptable participation rates—85 percent of both the schools and students, or a combined rate (the product of schools' and students' participation) of 75 percent.

1.5 Survey Operations and Quality Assurance in Data Collection

Chapter 5 describes the steps taken to ensure that the TIMSS Advanced assessments were conducted under standardized conditions in all participating countries. Each participating country was responsible for carrying out all aspects of data collection and scoring, following carefully documented standardized procedures and using customized software for all aspects of sampling students within schools. The IEA Data Processing and Research Center provided data entry software and variable codebooks to standardize data preparation and conducted extensive training seminars in the use of both sampling and data entry software.

In addition, to document that the TIMSS Advanced 2008 data collection was conducted according to the same standardized conditions in all countries, the TIMSS & PIRLS International Study Center together with the IEA Secretariat conducted an independent quality control program. The reports from the Quality Control

Monitors provided in Chapter 6 indicated that, in general, the national centers were able to conduct the data collection efficiently, professionally, and in compliance with international procedures.

1.6 The TIMSS Advanced 2008 International Database

To ensure comparable, high-quality data for analysis, the IEA Data Processing and Research Center took great care in creating the international database. Once the data files had been created and checked by national centers, the files were forwarded to Hamburg where the data underwent an exhaustive cleaning process. As described in Chapter 7, the data were checked and double-checked for consistency within and across countries. The national centers were contacted regularly and given multiple opportunities to review the data for their countries.

1.7 Scaling the Achievement Data and the International Benchmarks

Chapter 8 provides details of the process implemented by the TIMSS & PIRLS International Study Center to create achievement scales that would provide reliable measures of student achievement, including changes in average achievement since 1995. Subsequent to the field test, and then again, prior to scaling, a thorough review of the psychometric properties of the achievement items was conducted. This process began with an extensive review of item statistics for each achievement item in each country, including scoring reliability data for the constructed-response items—within country, across countries, and for trends. Also, the data were reviewed for item-by-country interactions. In general, the items exhibited very good measurement properties in all countries, and the scoring reliability was satisfactory (above 90% agreement in most cases).

As described in Chapter 8, student achievement in advanced mathematics and physics was summarized in TIMSS Advanced using item response theory (IRT) scaling methods. For accurate estimation of results for subpopulations of students, the scaling made use of plausible-value technology. The chapter describes this scaling methodology, and how the fitted model for each item was checked against the observed data. For trend items, the fit was plotted separately to ensure that the item was a good fit to both sets of assessment data—1995 and 2008. Scale scores (plausible values) were generated separately for each country and all of the results were plotted and checked. The achievement score distributions were very satisfactory and provided an excellent basis for further analysis and reporting the results.

In addition to describing student achievement in terms of average performance on the advanced mathematics and physics achievement scales, TIMSS Advanced 2008 reported student achievement at specific points on the scales, known as TIMSS Advanced 2008 International Benchmarks. Chapter 8 also documents the scale anchoring analysis conducted by TIMSS Advanced 2008 to describe and interpret student achievement at the Advanced (625), High (550), and Intermediate (475) International Benchmarks.

1.8 Conclusion

In conclusion, a major purpose of the *TIMSS Advanced 2008 Technical Report* is to provide detailed documentation about the procedures and methods used by TIMSS Advanced to provide internationally comparative data of high quality. This report explains the multi-faceted attention to quality and the many quality assurance steps that were implemented from the development of the assessment frameworks for TIMSS Advanced 2008 through release of the international database and User's Guide.

References

Garden, R.A., Lie, S., Robitaille, D.F., Angell, C., Martin, M.O., Mullis, I.V.S., Foy, P., & Arora, A. (2006). *TIMSS Advanced 2008 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Foy, P. & Arora, A. (2009). *TIMSS Advanced 2008 user guide for the international database.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I.V.S., Martin, M.O., Robitaille, D.F., & Foy, P. (2009). *TIMSS Advanced 2008 international report: Findings from IEA's study of achievement in advanced mathematics and physics in the final year of secondary school.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Chapter 2

Developing the TIMSS Advanced 2008 Instruments

Alka Arora and Ina V.S. Mullis

2.1 Introduction

Developing the TIMSS Advanced 2008 assessment began with work on the assessment framework in January 2006 and continued until July 2007, when the international version of the assessment was finalized for data collection. The development was a collaborative process involving the National Research Coordinators (NRCs) and item developers from participating countries. The process was managed by the TIMSS & PIRLS International Study Center staff with expert advice and guidance provided by the international coordinators Robert Garden for advanced mathematics and Svein Lie for physics, as well as the TIMSS Advanced Task Force members. The task force included both subject coordinators; staff from the TIMSS & PIRLS International Study Center; Wolfgang Dietrich from the National Agency of Education in Sweden; Torgier Onstad, Carl Angell, and Liv Sissel Gronmo from University of Oslo, Norway; and Helen Lye from Australian Council of Educational Research, Australia.

TIMSS Advanced 2008 was the second cycle of this assessment for advanced mathematics and physics students in their final year of secondary school and built on the first time this population was tested in TIMSS 1995. Since this is a trend study, TIMSS Advanced 2008 is structured so it includes new material, as well as the material from the 1995 assessment. Following the release of the 1995 results, half of the advanced mathematics and physics items were released. The replacement items were developed to include a wide distribution of items, as specified in the frameworks (see next section).

Experts from various participating countries contributed to developing the items. Items were reviewed and, if needed, revised by task force members. NRCs were responsible for final approval of the field-test items, and a field test was conducted in February–March 2007. The field test provided important information about the measurement properties of the items across the countries. Based on that information, items were selected and finalized for the TIMSS Advanced 2008 data collection. This chapter describes in detail the instrument development process. An overview of the process is shown in Exhibit 2.1.

Exhibit 2.1 Overview of the TIMSS Advanced 2008 Frameworks and Instrument Development Process

Date(s)		Group and Activity
January–February	2006	TIMSS & PIRLS International Study Center began work on TIMSS Advanced 2008 Assessment Frameworks
March	2006	TIMSS & PIRLS International Study Center sent draft frameworks to National Research Coordinators for their review and recommendations
March–April	2006	Experts from the participating countries began developing field test items
May	2006	1st National Research Coordinators Meeting (Amsterdam) finalized frameworks and reviewed field test item pool
August	2006	TIMSS & PIRLS International Study Center published the TIMSS Advanced 2008 Assessment Frameworks
September	2006	2nd National Research Coordinators Meeting (Oslo) reviewed and finalized field test instruments- items and questionnaires
October	2006	TIMSS & PIRLS International Study Center distributed final field test instruments to the National Research Coordinators
October	2006	TIMSS & PIRLS International Study Center conducted a Pilot test of constructed-response items in several countries to collect sample responses for the constructed-response items.
November	2006	Task Force met (Boston) to finalized scoring guides for constructed-response items and develop scoring training materials for the 3rd NRC Meeting
February	2007	3rd National Research Coordinators Meeting (Rome) conducted field test scoring training
February–March	2007	TIMSS Advanced 2008 field test administered
May	2007	TIMSS & PIRLS International Study Center reviewed the field test item statistics to propose the assessment items for review at the 4th NRC Meeting
June	2007	4th National Research Coordinators Meeting (Lubeck) reviewed and approved items and questionnaires for the TIMSS Advanced 2008 assessment
July	2007	TIMSS & PIRLS International Study Center distributed the TIMSS Advanced 2008 Data Collection instruments to National Research Coordinators
September	2007	Task Force met (Oslo) to review and finalize scoring guides and scoring training materials for the 4th NRC Meeting Reviewed and refined the proposed curriculum questionnaires
January	2008	5th National Research Coordinators Meeting (Portoroz) conducted scoring training for constructed-response items
February–May	2008	TIMSS Advanced 2008 data collection

2.2 Assessment Frameworks

The *TIMSS Advanced 2008 Assessment Frameworks* (Garden, R.A., Lie, S., Robitaille, D.F., Angell, C, Martin, M.O., Mullis, I.V. S., Foy, P., & Arora, A., 2006) contains a detailed description of the TIMSS Advanced 2008 assessment in advanced mathematics and physics. The basic structure of each of the frameworks has two dimensions: content and cognitive domain. The content domains specify the subject matter to be assessed within each subject, and the cognitive domains describe the thinking processes to be assessed.

Advanced mathematics has three content domains or areas: algebra, calculus, and geometry. Physics has four content domains: mechanics, electricity, heat and temperature, and atomic and nuclear physics. For both subjects, the cognitive domains are the same. There are three cognitive domains: knowing, applying, and reasoning. Exhibits 2.2 and 2.3 show the target percentages for advanced mathematics and physics devoted to the content and the cognitive domains, as described in the framework.

Exhibit 2.2 Target Percentages for the TIMSS Advanced Mathematics Assessment Devoted to Content and Cognitive Domains

Content Domains	Percentages
Algebra	35%
Calculus	35%
Geometry	30%

Cognitive Domains	Percentages
Knowing	35%
Applying	35%
Reasoning	30%

Exhibit 2.3 Target Percentages for the TIMSS Advanced 2008 Physics Assessment Devoted to Content and Cognitive Domains

Content Domains	Percentages
Mechanics	30%
Electricity and Magnetism	30%
Heat and Temperature	20%
Atomic and Nuclear Physics	20%

Cognitive Domains	Percentages
Knowing	30%
Applying	40%
Reasoning	30%

2.3 Number of Items

The TIMSS Advanced 2008 assessment design is described in the *TIMSS Advanced 2008 Assessment Frameworks.* In brief, the assessment consists of 14 item blocks. Out of the total of 14 blocks, 6 blocks consist of trend items (items that were used in 1995 assessment), and 8 blocks consist of items newly developed for the 2008 assessment. These 14 blocks were distributed across 8 booklets. The design was chosen to ensure that each student responded to a sufficient number of items to provide a reliable measure, as well as to ensure that the trends across content and cognitive domains were reliably measured. Based on the design, a total of 72 advanced mathematics and 71 physics items were included in the assessment.

2.4 Developing Advanced Mathematics and Physics Items and Scoring Guides

Developing the replacement items for the TIMSS Advanced 2008 assessment was a collaborative effort of participating institutions. The development work on the items began in March 2006, immediately after the draft framework was posted for the NRCs review. In May 2006, the first NRC meeting to review the item pool was held. There were 181 advanced mathematics items and 80 physics items to review at this meeting, developed mostly by the international subject coordinators. Norway, Russia, and Slovenia also contributed items to the item pool. During the meeting, participants gave suggestions for revising some of the items, and a few items were rejected. From June 2006 to August 2006, the item development work continued. More new items were developed by Sweden and Norway, and items also were developed by physics experts in Australia and mathematics experts in Bulgaria. In September 2006, new items were developed to cover those areas in the framework for which there were few items. Approximately 125 advanced mathematics and 110 physics items were presented for the discussion at the second NRC meeting in September. As a result of the review during the NRC meeting, 90 items for each subject were selected for the field test.

Each constructed-response question was developed with a scoring guide. Constructed-response items generally were worth 1 or 2 score points, depending on the nature of the task or skills required to complete it. Constructed-response items worth 1 score point typically require a numerical response or a brief descriptive response, while those worth 2 score points require students to show their work or provide explanations using words and/or diagrams to demonstrate their conceptual understanding.

2.5 Conducting the TIMSS Advanced 2008 Field Test

Newly developed items for the TIMSS Advanced 2008 assessment were field tested in February–March 2007. Eight countries participated in the field test. Approximately twice the number of items were field tested as were needed for the data collection. For each subject, 90 items were assembled into nine blocks, and placed into three booklets. Typically countries sampled between 18–28 schools and approximately 500 students.

The field test provided information for evaluating the measurement properties of the new items developed for this assessment.

2.6 Piloting Items for the Scoring Guides

The TIMSS Advanced 2008 constructed-response items elicit a wide range of responses from students. It is very important to score these responses consistently across countries and languages. This requires extensive training in applying the scoring guides. For training purposes, a pilot test was conducted to obtain students' responses for a selected set of constructed response items. The TIMSS Advanced 2008 pilot test contained 15 items for each subject. It was conducted in September–October 2006. At that time, Australia, Serbia, Armenia, the Netherlands, and Norway conducted the pilot test. Responses from non-English-speaking countries were translated into English before they were used as example responses. Participating countries conducted the pilot test in either one or two classes.

2.7 Field Test Scoring Training for Constructed-response Items

In preparation for the field test scoring training meeting, the task force met to prepare the NRC training materials. Task force members first reviewed each scoring guide to determine whether all types of responses were covered in the response categories mentioned in the guide and also whether all the categories were mutually exclusive. Then, they reviewed the responses collected during the pilot test, and selected 8–12 examples and 8–12 practice responses for the example items that members considered elicited varied range of responses. These responses were included in the training binder that was prepared for the scoring training meeting.

The TIMSS & PIRLS International Study Center conducted the constructed-response scoring training meeting in February 2007 for NRCs and their scoring managers who implemented the constructed-response scoring in their respective countries.

2.8 Item Selection for Data Collection

The selection of items for data collection was based on results from the field test. After the field test, the countries sent the data files to the IEA Data Processing and Research Center (DPC) for cleaning and verification. After verifying and transforming data into the international format, the IEA DPC sent the data to the TIMSS & PIRLS International Study Center. The TIMSS & PIRLS International Study Center then prepared data almanacs for review and presented the results of all items that were field tested. Two sets of almanacs were produced for each subject. The first set gave an overall picture of the item. For each item, the difficulty, discrimination, and reliability indices were displayed. Additionally, for multiple-choice items, the almanacs also included information on how many students chose the particular response option. Also, for constructed response items what percentage of student received 0, 1 or 2 scores was displayed The second set of almanacs, showed for each participating country, the percent of students who chose a specific response option. These almanacs were the bases of evaluating the performance and quality of the achievement items and making suggestions for revisions for the data collection.

For each item, the results were reviewed in the light of the difficulty of the item, how well the item discriminated between high- and lowachieving students, the effectiveness of the alternatives, and the scoring reliability for the constructed-response items.

First, in May 2007, the TIMSS & PIRLS International Study Center staff and the mathematics and physics coordinators reviewed data from the field test. The items were categorized into "proposed" and "alternate" items. The proposed items were then reviewed by the NRCs at the fourth NRC meeting held in June 2007. During the review process, some proposed items were replaced by alternate items, and some minor changes were made to a few of the proposed items.

Finally, 90 items were selected, 45 for each subject from the pool of 180 items that had been field tested. These newly developed items, together with the trend items from 1995, form the TIMSS Advanced 2008 assessment. The trend items were also mapped into content and cognitive categories described in the TIMSS Advanced 2008 frameworks.[1]

Mathematics Assessment

Exhibit 2.4 shows the distribution of new and trend items in the TIMSS Advanced 2008 mathematics test by content and cognitive domains. Additionally, this exhibit also includes information about item formats.

1 Four items in mathematics and two items in physics could not be classified according to the new categories. These six items were not included in Exhibits 2.4 through 2.9

Exhibit 2.4 Advanced Mathematics Items by Content and Cognitive Domains and Item Format

Content Domain	Trend Items in 2008 Assessment	New Items in 2008 Assessment	All Items in 2008 Assessment	Multiple-Choice Items in 2008 Assessment	Constructed-Response Items in 2008 Assessment
Algebra	10	16	26	17	9
Calculus	7	18	25	13	12
Geometry	10	11	21	16	5
Total	**27**	**45**	**72**	**46**	**26**

Cognitive Domain	Trend Items in 2008 Assessment	New Items in 2008 Assessment	All Items in 2008 Assessment	Multiple-Choice Items in 2008 Assessment	Constructed-Response Items in 2008 Assessment
Knowing	14	14	28	21	7
Applying	8	19	27	14	13
Reasoning	5	12	27	11	6
Total	**27**	**45**	**72**	**46**	**26**

Exhibit 2.5 shows the score point distribution for the mathematics assessment by content and cognitive domains. The target percentages for content domains, described in the framework, were met within an acceptable difference (2%). For the cognitive domains, the percentage of items assessing reasoning was a little less than desired(4%) and consequently the percentage of knowing and applying were somewhat higher.

Exhibit 2.5 Distribution of Score Points in the Advanced Mathematics Assessment by Content and Cognitive Domains

Content Domain	Cognitive Domain			Total Score Points	Percentage of Score Points
	Knowing	Applying	Reasoning		
Algebra	12	11	7	30	37%
Calculus	13	8	8	29	35%
Geometry	5	12	6	23	28%
Total Score Points	**30**	**31**	**21**	**82**	
Percentage of Score Points	**37%**	**38%**	**26%**		

The number of score points across the content domains for each booklet is shown in Exhibit 2.6. The number of score points per booklet varied from 36 to 38 points, except in booklet 1. Booklet 1 composed only of trend blocks had 30 points.

Exhibit 2.6 Number of Score Points in the Advanced 2008 Mathematics Booklets by Content Domain

Content Domain	Booklet			
	1	2	3	4
Algebra	12	15	12	14
Calculus	7	13	15	12
Geometry	11	9	11	10
Total in Mathematics	**30**	**37**	**38**	**36**

Physics Assessment

Exhibit 2.7 shows the distribution of new and trend items in the TIMSS Advanced 2008 physics test by content and cognitive domains. Additionally, this exhibit also includes information about item formats.

Exhibit 2.7 Physics Items by Content and Cognitive Domains and Item Format

Content Domain	Trend Items In 2008 Assessment	New Items in 2008 Assessment	All Items in 2008 Assessment	Multiple-Choice Items in 2008 Assessment	Constructed-Response Items in 2008 Assessment
Mechanics	9	11	20	11	9
Electricity and Magnetism	8	13	21	13	8
Heat and Temperature	2	13	15	7	8
Atomic and Nuclear Physics	7	8	15	11	4
Total	**26**	**45**	**71**	**42**	**29**

Cognitive Domain	Trend Items In 2008 Assessment	New Items in 2008 Assessment	All Items in 2008 Assessment	Multiple-Choice Items in 2008 Assessment	Constructed-Response Items in 2008 Assessment
Knowing	3	15	18	12	6
Applying	15	21	36	25	11
Reasoning	8	9	17	5	12
Total	**26**	**45**	**71**	**42**	**29**

Exhibit 2.8 shows the score point distribution for the physics assessment by content and cognitive domains. Mostly, the target percentages described in the framework were met within the acceptable difference. The percentage of items assessing knowing was less and applying was more than desired.

Exhibit 2.8 Distribution of Score Points in the Physics Assessment by Content and Cognitive Domains

Content Domain	Cognitive Domain			Total Score Points	Percentage of Score Points
	Knowing	Applying	Reasoning		
Mechanics	4	11	9	24	29%
Electricity and Magnetism	5	12	7	24	29%
Heat and Temperature	4	10	6	20	24%
Atomic and Nuclear Physics	5	8	3	16	19%
Total Score Points	**18**	**41**	**25**	**84**	
Percentage of Score Points	**21%**	**49%**	**30%**		

Exhibit 2.9 shows the number of score points in each of the four physics booklets by content domain. The total number of items per booklet ranged from 31–38 across the four booklets.

Exhibit 2.9 Number of Score Points in the Physics Booklets by Content Domain

Content Domain	Booklet			
	5	6	7	8
Mechanics	10	9	10	10
Electricity and Magnetism	10	12	11	8
Heat and Temperature	3	12	8	10
Atomic and Nuclear Physics	8	5	6	8
Total in Physics	**31**	**38**	**35**	**36**

2.9 Finalizing the Scoring Guides for Constructed-response Items

In September 2007, the TIMSS Advanced Task Force met to review and revise the constructed-response scoring guides and training materials. Based on the field test, some response categories were deleted and some categories were revised. There also were adjustments in the trend scoring guides to align them with the scoring guides for the new items. Also, some response categories that were used in 1995 were collapsed to decrease the scoring burden of the scorers and increase scoring reliability.

The training materials for the 55 constructed-response items were arranged by assessment block, unlike the field test which was arranged by the booklets. Training materials included 6–15 example responses and 6–15 practice responses for each of the constructed–response items. During the 5th NRC meeting, NRCs and their scoring managers were given extensive training on how to use these materials in their countries. Discussions in the training session led to further refinements of some categories in the scoring guides. After the meeting, those revisions were made, and the final versions of the scoring guides were made available to the NRCs in February 2008.

2.10 Developing the TIMSS Advanced 2008 Background Questionnaires

TIMSS Advanced 2008 collected information about key factors related to student's home and school environments. In order to collect this information, TIMSS Advanced 2008 administered questionnaires to NRCs, school principals, teachers, and students. These questionnaires are described in detail in the next section (2.11). However, in brief:

- The *Curriculum Questionnaires* for advanced mathematics and physics were completed by NRCs. Newly developed for 2008, these questionnaires collected information about the organization of the advanced mathematics curriculum and the physics curriculum.

- The *School Questionnaire* asked school principals to provide information about the school contexts and the resources available for advanced mathematics and physics instruction.
- The *Teacher Questionnaires*, one for mathematics teachers and the other one for physics teachers, gathered information about teachers' background as well as the structure and content of instruction in the classroom.
- The *Student Questionnaires*, one for advanced mathematics and one for physics, collected information about students' background, their home, experience in and out of school, their attitudes, and the resources available at home and in school for learning.

Developing the TIMSS Advanced 2008 background questionnaires was a collaborative effort between the TIMSS & PIRLS International Study Center and the NRCs of the participating countries.

The development work began in August of 2006 with staff at the TIMSS & PIRLS International Study Center reviewing TIMSS 1995 questionnaires for the advanced populations of students in the final years of secondary school and TIMSS 2007 questionnaires. Based on these two different sets of questionnaires, a set of TIMSS Advanced 2008 questionnaires was drafted with items collecting important information on the contexts of teaching and learning for this particular population of students.

These draft questionnaires were presented at the 2nd NRC meeting in September 2006. NRCs reviewed the draft questionnaires and gave suggestions for improvements. There were some questions that were revised, and some new questions were added. The revised questionnaires were formatted for the field test and distributed to

NRCs, along with the filed test achievement booklets in October 2006.[2] The questionnaires were field tested in February–March of 2007.

After the field test, the countries sent the data files to the IEA Data Processing and Research Center (DPC) for cleaning and verification. After verifying and transforming data into the international format, the IEA DPC forwarded the data to the TIMSS & PIRLS International Study Center. The TIMSS & PIRLS International Study Center then prepared data almanacs of the field test results of all questionnaires. For every participating country, each almanac displayed student-weighted distributions of responses for 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 advanced mathematics or physics. For questions with numeric responses, mean, mode, and selected percentiles were given. These almanacs were used to evaluate the performance and quality of the field test questionnaire items.

The review of the field test data almanacs was completed at the 4th NRC meeting in June 2007. The group examined the item statistics and some items were deleted. In a few instances, the language was clarified. Following the meeting, the TIMSS & PIRLS International Study Center updated the questionnaires. The final questionnaires were made available to the NRCs in July 2007 so countries could begin the translation and verification process.[3]

Because the NRCs from the 10 countries were responsible for completing the curriculum questionnaires, they did not need to be field tested. Work begun in August 2007, with TIMSS & PIRLS International Study Center staff drafting curriculum questionnaires based on *TIMSS 2007 Curriculum Questionnaires*. These drafts were first discussed with task force members during the September 2007 meeting. During this meeting, some of the existing questions were modified or rejected, and

2 The curriculum questionnaires were not distributed at this stage.

3 The translation and verification process is described in detail in Chapter 3.

new questions were added. In October, the revisions were made to the draft questionnaires based on feedback from the task force meeting. These revised questionnaires were discussed during the fifth NRC meeting in January 2008. During this meeting, more revisions were made. The final curriculum questionnaires were distributed to NRCs in February 2008.

2.11 Content of the Background Questionnaires

The content of each TIMSS Advanced 2008 questionnaire is summarized below. Exhibits 2.10 through 2.13 provide descriptions of the variables within each questionnaire. The variables in each questionnaire are grouped according to their contextual factors.

Curriculum Questionnaires

The NRCs were responsible for completing the curriculum questionnaires. The curriculum questionnaires were designed to collect basic information about the organization, content, and implementation of the intended mathematics and physics curriculum in each country. The questionnaires also contained questions about requirements for teachers.

The two versions of the curriculum questionnaires for advanced mathematics and physics, respectively, were parallel in structure and very similar in content, with slight modifications made to accommodate the subject-specific content.

School Questionnaire

The principal of each sampled school for TIMSS Advanced 2008 completed a school questionnaire. The questionnaire was designed to collect information about the school's demographic characteristics, resources for teaching, and the school environment. Principals also answered questions about their role as an administrator.

Teacher Questionnaires

Teachers of the assessed mathematics and physics classes responded to the corresponding teacher questionnaire for advanced mathematics or physics. The questionnaires were designed to gather information about the classroom contexts of teaching and learning. Teachers also answered questions about their professional preparation and experience in teaching.

The general structure of the two questionnaires was the same. However, questions pertaining to instructional and assessment practices and content coverage were tailored to the specific subject.

Student Questionnaire

Each student participating in the study completed the appropriate advanced mathematics or physics student questionnaire. The student questionnaires were designed to gather information on some of the major factors that influence student achievement in the areas of advanced mathematics and physics. The questionnaire included questions about the home background and resources for learning, attitudes about advanced mathematics and physics, and experiences in learning these subjects. Once again, the two questionnaires—for advanced mathematics and physics students were similar. However, when necessary, the subject-specific content was tailored to the specific subject.

Exhibit 2.10 Content of TIMSS Advanced Curriculum Questionnaire

Items		Context	Variables
Mathematics	Physics		
1	1	Curriculum charactistics	Year the curriculum was implemented
			Whether the curriculum is being revised
6	6		Forms in which the curriculum is made available
8	8		Total amount of class time prescribed by the curriculum for students in the track assessed
7	7	Governance of education system	Whether textbooks used in the track or course assessed were certified by an education authority
			Who is responsible for cost of textbooks
8	8		Whether the country has national requirement on the number of school days per year for the track or course being assessed
13	13		Whether the national education authority administers any examinations that have consequences for individual students
2	2	Curriculum Policy	Whether the curriculum has prerequisite courses or tracks for students
			Percentage of students fulfilling the prerequisites
			Whether taking the mathematics/physics track or course is a prerequisite for further study
4	4		Whether the national curriculum addresses the use of computers in the track or course being assessed
5	5		Whether TIMSS Advanced mathematics/physics topics are included in the curriculum over the course of the year
9	9		Whether there is an official policy to encourage students to choose advanced courses
12	12		Methods used to evaluate the implementation of the curriculum
3	3	Emphasis on calculator use	Whether the curriculum for students being assessed addresses the use of calculators
			Whether the curriculum specifies the type of calculators that may be used
			Whether the curriculum permits use of calculators in national examinations
			Who pays for the cost of calculators
10	10	Teacher preparations	National requirements for a teacher of the track or course being assessed
11	11		Methods used to communicate changes about the curriculum to teachers

Exhibit 2.11 Content of TIMSS Advanced School Questionnaire

Context	Item	Variables
School Characteristics	1	Number of students enrolled in the school and grade tested
	2	Size of community in which school is located
	3	Percentage of students in the school from economically affluent and disadvantaged homes
	4	Percentage of students whose native language is the language of the test
	5	Percentage of students taking the TIMSS Advanced 2008 tests
School Policy	6	Whether the school had a policy encouraging students to take courses in mathematics and physics
	9/10-	Whether mathematics/physics teachers' practices were evaluated by the principal or senior staff, level of student achievement, etc.
	12	Whether the school uses incentives to recruit or retain teachers
School Climate	7	Principal's time allocation for different tasks and functions
	8	Principal's perception of different aspects of school climate
	11	Difficulties of filling teaching vacancies
	13	Principal's perception of the frequency and severity of different problems within the school
Resources and Technology	14	Material factors affecting school's capacity to provide instruction
	15	Whether a physics laboratory and assistance was provided for students' experiments
	16	Whether school had support in helping teachers use information and communication technology for teaching and learning
	17	Number of computers and internet available for educational purposes

Exhibit 2.12 Content of the TIMSS Advanced Teacher Questionnaire

Item number		Context	Variables
Mathematics	Physics		
1	1	Teacher Demographics	Age
2	2		Gender
3	3		Total number of years teaching and number of years teaching advanced mathematics/physics
4	4		Expected time to continue teaching advanced mathematics/physics
5	5		Teacher's highest level of formal education
6	6		Teacher's major areas of study during post-secondary education
7	7	Teacher Training and Preparation	Whether the teacher has a license or certificate
8	8		How ready the teacher feels to teach the topics included in the TIMSS Advanced mathematics/physics test
9	9		Frequency of various types of interactions the teacher has with colleagues
10	10		Whether the teacher is a member of professional organization
			Whether the teacher has regularly participated in professional organization activities over the past two years
11	11		Whether the teacher has participated in various professional development activities over the past two years
12	12		Whether the teacher has participated in various activities in mathematics/physics fields
13	13	School Environment and Structure	Teacher's perception about the school's safety
14	14		Teacher's perception about the school's facilities
15	15		Teacher's perception of teachers' job satisfaction, understanding of and success in school's curriculum, and expectations for student achievement; of parental support and involvement; and of students' regard for school property, and desire to do well in school
16	16	Class Charactistics and Climate	Number of students in TIMSS class
22	22		Extent to which the teacher perceives various student and resource factors as limiting teaching
17	17		Minutes per week the teacher teaches advanced mathematics/physics to the TIMSS class
18	18		Minutes per week the teacher spends on preparation for teaching the TIMSS class
20	20		Percentage of time in a week spent on various teaching activities in advanced mathematics/physics lessons
21	21		Frequency with which the teacher asks students to do various learning activities in the TIMSS class
23	23		Percentage of time spent on advanced mathematics/physics content areas over the course of the year
19	19	Instructional Materials and Technology	Whether a textbook is used as the basis of instruction
			Whether each student has his or her own textbook
			Frequency with which the teacher asks students to do various textbook-related activities in advanced mathematics/physics
24	24		Coverage of topics in the advanced mathematics/physics content areas while teaching over the course of the year
25	25		Frequency with which the teacher uses a computer to demonstrate advanced mathematics/physics

Exhibit 2.12 Content of the TIMSS Advanced Teacher Questionnaire (Continued)

Item number		Context	Variables
Mathematics	Physics		
26	26	Instructional Materials and Technology	Whether students have access to calculators, computers or other computing technology in class
			Type of calculators majority students have access to in class
			Whether computers have access to internet
27	27		Frequency with which the students use calculators or computers for various learning activities
28	28	Homework and Assessment	Whether the teacher assigns advanced mathematics/physics homework to the TIMSS class
29	29		Frequency with which the teacher assigns mathematics/physics homework to the TIMSS class
30	30		Number of minutes taken by an average student to complete an advanced mathematics/physics homework assignment
31	31		Frequency with which the teacher assigns various types of homework
32	32		Emphasis the teacher places on various sources to monitor students' progress
33	33		Frequency with which the teacher gives a test or examination
34	34		Item formats the teacher typically uses in tests or examinations
35	35		Types of questions the teacher includes in tests or examinations

Exhibit 2.13 Content of the TIMSS Advanced Student Questionnaire

Item number		Context	Variables
Mathematics	Physics		
1	1	Student Characteristics	Year and month of student's birth
2	2		Gender
3	3		Student's frequency of use of the language of test at home
7	7		Whether the student's mother and father were born in country
8	8		Whether the student was born in country, and, if not, age at which the student emigrated
4	4	Economic and Educational Resources in the Home	Number of books in the student's home
5	5		Educational resources and general possessions in the student's home
6	6		Highest education level completed by the student's mother and father
20	20		Frequency of tutoring in advanced mathematics/physics
9	9	Student Attitudes	Whether the student intends to continue his or her education after secondary school
10	10		Subject that the student intends to study if he or she plans to continue education
13	13		Reasons why the student is taking advanced mathematics/physics
11	11	Computer/Calculator Activities	Average time in a day spent on a computer by the student
			Frequency with which the student uses a computer in various places
			Whether the student uses a computer for various learning activities
17	17		Frequency with which the student uses a calculator, computer or other computing technology in advanced mathematics/physics lessons
			Type of calculator the student uses in advanced mathematics/physics lessons
19	19		Frequency with which the student uses a computer for work on advanced mathematics/physics outside of class
12	12	Activities in School	Average time in a normal school day the student spends on various activities before or after school
14	14		Minutes per week the student spends in advanced mathematics/physics class
			Whether the advanced mathematics student is taking or has taken the physics track/course and Whether the physics student is taking or has taken the advanced mathematics track/course
15	15		Frequency with which the student uses various learning methods in advanced mathematics/physics lessons
16	16		Frequency with which the student engages in various learning activities in advanced mathematics/physics lessons
18	18	Homework and Assessment	Minutes per week the student spends doing advanced mathematics/physics homework
			Frequency with which the student does various activities for doing homework
21	21		Frequency with which the student prepares for a test or examination in advanced mathematics/physics

References

Garden, R.A., Lie, S., Robitaille, D.F., Angell, C., Martin, M.O., Mullis, I.V.S., Foy, P., & Arora, A. (2006). *TIMSS Advanced 2008 assessment frameworks.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Chapter 3

Translation and National Adaptations of the TIMSS Advanced 2008 Instruments

Barbara Malak and Suzanne Morony

3.1 Introduction

The international versions of the TIMSS Advanced 2008 assessment items, background questionnaires, and procedural manuals were developed in English, the working language of the International Association for the Evaluation of Education Achievement (IEA). Using the international versions prepared by the TIMSS & PIRLS International Study Center, participants translated the materials into their target language(s) and adapted them as required for their national context.

Throughout this translation and adaptation process, the primary purpose was to ensure that the set of instruments in each country was internationally comparable, while still allowing each country to adapt the materials to their national needs. Guidelines for translating the materials were described in the *Survey Operations Procedures Unit 3: Preparing Materials for the TIMSS Advanced 2008 Data Collection and Administering the Assessment* (TIMSS & PIRLS International Study

Center, 2007), and were discussed at meetings of National Research Coordinators (NRCs). The translated texts were subjected to a stringent international translation verification process that was managed by the IEA secretariat in Amsterdam. The IEA secretariat scheduled activites and trained personnel at Lionbridge, an independent translation company (based in Brussels, Belgium).

Each participating country was asked to submit materials for verification prior to both the field test and the main data collection. National Research Coordinators received a Translation Verification Report in which all verifier comments were recorded either directly into the submitted documents or in separate tables. NRCs made changes to their instruments as appropriate, and submitted the final version to the TIMSS & PIRLS International Study Center for verification of the layout of the assembled instruments.

Finally, international Quality Control Monitors (QCMs) reviewed the final printed instruments against the Translation Verification Report and recorded instances where the verifier's suggestions were not implemented. These reports were forwarded to the TIMSS & PIRLS International Study Center and were supplemented with Translation Verification Summary reports, in which NRCs explained the justification for not implementing verifier suggestions concerning any serious error. NRCs were not obliged to implement suggestions from the verifier; however, they did take responsibility for any errors or mistranslations in the instruments

3.2 TIMSS Advanced materials to be translated

For TIMSS Advanced 2008, the following instruments and related materials required translation:

- Seven blocks of advanced mathematics achievement items and 7 blocks of physics achievement items;

- Two sets of booklet covers and introductory information (including directions)—one each for advanced mathematics and physics;
- Background questionnaires for students, teachers, and schools;
- School coordinator, test administrator, and national quality control monitor manuals;
- Scoring guides for constructed-response items, where necessary.

The TIMSS & PIRLS International Study Center provided each country with the electronic files necessary to facilitate the translation of the blocks and the subsequent creation of the booklets. There were 8 test booklets for the TIMSS Advanced 2008 data collection: 4 for the advanced mathematics assessment, and 4 for the physics assessment. The booklets comprised blocks of items that were assigned in a systematic fashion. To create the booklets, each item block and an introduction for each subject were translated and then these "components" of the booklets were later distributed throughout the booklets.

Three types of background questionnaires were prepared and administered as part of the TIMSS Advanced 2008 data collection: a school questionnaire, two teacher questionnaires (one for advanced mathematics, and one for physics), and two student questionnaires (one for students assessed in advanced mathematics, and one for students assessed in physics). Each student questionnaire contained a general block and one of the subject-specific blocks (advanced mathematics or physics). For the student questionnaire, the general block was translated once and later included in both student questionnaires.

3.3 Number of Languages Used for Translation

In total, the TIMSS Advanced 2008 data collection instruments were translated into 11 languages. Two countries (Lebanon and Norway) prepared the test materials in two languages, and two countries collected data in English. Participants who tested in English were also expected to go through the verification of their national adaptations and the layout of the instruments. Exhibit 3.1 shows the languages used by each participant for the various instruments.

Exhibit 3.1 Participants and Languages in TIMSS Advanced 2008

Country	Language
Armenia	Armenian
Iran, Islamic Rep. of	Farsi
Italy	Italian
Lebanon	English French
Netherlands	Dutch
Norway	Bokmål Nynorsk
Philippines	English
Russian Federation	Russian
Slovenia	Slovenian
Sweden	Swedish

3.4 Translators and Reviewers

Participating countries were strongly advised to hire an experienced translator who was qualified to translate the TIMSS Advanced 2008 items and questionnaires, and a reviewer to review the translations. It was important for the translator to have had experience translating texts in mathematics and physics, preferably at the level of the target grade; and it was desirable for the translator to be familiar with test development.

Guidelines suggested that both the translator and the reviewer of the TIMSS Advanced instruments should have:

- an excellent knowledge of English;
- an excellent knowledge of the target language;
- experience in the national cultural context; and
- experience with students in the target grade.

The ideal translation reviewer would be a teacher teaching students in final year of secondary school in the country with the qualities listed above, and an understanding of the subject matter.

Countries could employ more than one translator and/or reviewer (per subject). In some cases it was not possible to engage translators and reviewers with the required language skills and subject matter knowledge in both advanced mathematics and physics. Another reason for dividing the work of translators was that subject matter experts were not needed for the questionnaire materials. In such cases, National Research Coordinators were reminded to ensure the consistency of the translations within and across instruments. Countries preparing translations in more than one language were encouraged to involve professionals familiar with the various languages in order to make sure that the translations were equivalent across languages.

3.5 Translation and Adaptation Guidelines

To ensure that appropriate translations and adaptations were made when the TIMSS Advanced instruments were produced, the TIMSS & PIRLS International Study Center provided basic guidelines for the translating and adapting process. These guidelines are summarized in the list below.

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

3.5.1 Adaptations to Test Items

In order to ensure international equivalence of the achievement items across countries, NRCs were instructed to minimize adaptations to them. Unfamiliar vocabulary and expressions could and should be adapted to ensure that the terminology was equally familiar to students in all countries, so long as this did not change the meaning or the difficulty level of the item. The major concern was to convey the same meaning and style as the text of the international version. Guidelines for specific terms, including units of measurement, were listed in the *Survey Operations Procedures Unit 3*.

3.5.2 Adaptations to Questionnaires

Guidelines for adapting terminology in the questionnaires were similar to that for the achievement items; however, unlike the achievement items, there were places in the questionnaires where adaptations were required. These places were marked with carets (< >), indicating information that must be replaced with the country-appropriate term. For example, <country> would be replaced with the country name and <language of test> would be replaced with the language of the test in that country. The NRC received adaptation notes for the questionnaires.

3.6 Documenting National Adaptations

All deviations from the international versions of the advanced mathematics and physics assessment booklets or questionnaires were documented on National Adaptations Forms (NAFs). The National Adaptations Forms consisted of a set of 8 forms: the first 6 were for the background questionnaires, the seventh was for the achievement items, and the last one consisted of general adaptations that applied to all instruments. The forms were supplied as one electronic document to be treated as a set, and each version was submitted as a single document upon completion. The forms listed any changes made and, in the rare cases of not-administered questions, the rationale behind these decisions. These forms were updated after each stage of the verification process. NRCs completed Version I of the forms during the internal translation and review process and sent it along with the rest of the materials for translation verification. After translation verification, NRCs updated the forms (Version II) to reflect any changes resulting from the verification process and sent them along with the national instruments for TIMSS & PIRLS International Study Center review. After finalizing the national instruments, NRCs uploaded the forms again (Version III) for data-collection processing at the IEA Data Processing and Research Center (DPC) and as a final documentation of their national adaptations. The NAFs were completed in English only, so that the staff reviewing them at different stages could review the changes and ensure that they were acceptable and did not affect the international comparability of the instruments.

3.7 International Translation Verification

Each translation went through a rigorous verification process that included internal verification of the translations at the national centers, independent verification of the translations at the national centers, independent verification by an international translation company, and a check by international QCMs to determine whether or not the verifier's suggestions had been adapted. As the last step, the TIMSS & PIRLS International Study Center reviewed the assembled test instruments from all participating countries. For more information on the checking process used by international QCMs and the review process used by TIMSS & PIRLS International Study Center staff, please refer to Chapter 5.

Once the instruments had been translated and internally reviewed, the text of the booklet cover pages, introductions, assessment blocks (including trend items from 1995 where relevant), and questionnaires were submitted for international translation verification. This process was managed by the IEA Secretariat in Amsterdam, that enlisted Lionbridge, an independent translation company (based in Brussels, Belgium), to verify translations for each of the countries. Of the 10 participants in TIMSS Advanced 2008, all except Lebanon and Philippines (and Norway for Nynorsk language) participated in and submitted materials for verification for the field test. All participants submitted instruments for verification before the main data collection;

however, in the Philippines the results of the verification were not used to make final changes to the assessment materials;[1] and, in Lebanon the verification was done only for the French version, and without National Adaptations Forms.

3.7.1 Verification of Translations at National Centers after the Field Test

International translation verification procedures for the field test were equally as rigorous as for the main data collection. The results of item analyses from the TIMSS Advanced 2008 field test, conducted during February and March of 2007, 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 mis-translated. If needed, they corrected the translation for the final TIMSS Advanced 2008 test instruments.

3.7.2 International Translation Verifiers

The international translation verifiers for TIMSS Advanced 2008 were required to be educated at university level: preferably to have postgraduate qualifications in mathematics or science; to have the target language as their first language; to have formal credentials as translators working in English; and, if possible, to live and work in the country for which the verification was being carried out (or in close contact with this country). Experience translating technical texts was preferred.

Verifiers received general information about the study and the design of the instruments, together with a description of the translation procedures used by the national centers. They also received detailed instructions for reviewing the instruments and registering deviations from the international version. Wherever possible, verifiers were asked to make their comments and changes directly on the PDF documents.

1 The Philippines NRC and the international translation verifier for the Philippines reported that there were no changes made to the international achievement items in the national version of the instruments

3.7.3 Process of Translation Verification

National Research Coordinators were instructed to send PDF files of the following documents to the IEA Secretariat for translation verification:

- 7 blocks of advanced mathematics achievement items;
- 7 blocks of physics achievement items;
- 1 set of advanced mathematics booklet covers;
- 1 set of physics booklet covers;
- the calculator use survey;
- the advanced mathematics booklet introduction;
- the physics booklet introduction;
- the school questionnaire;
- the advanced mathematics teacher questionnaire;
- the physics teacher questionnaire;
- the general, advanced mathematics, and physics item blocks of the student questionnaires;
- covers for the student questionnaires; and
  a Microsoft®-Word document of:
- the completed National Adaptations Forms.

Verifiers received these materials together with the international English versions of instruments and the *Guidelines for Translation Verification of the TIMSS Advanced 2008 Instruments*. For the 5 countries (Italy, Norway, Russia, Slovenia, and Sweden) that also participated in the TIMSS 1995 study at this level, verifiers were responsible for ensuring that the 2008 national versions of the trend items were identical to those administered in 1995. Accordingly, verifiers reviewing instruments for the trend countries also received the translated trend items used in that country in 1995. Verifiers of trend items were instructed to check that the 1995 and 2008 items matched exactly, and to detail the nature of any change identified.

Verifiers were given the option of registering their comments directly on the submitted PDF documents of the translation (using the eXPert PDF 4 Professional application), or in a specially created report (using Microsoft® Word) especially for languages written right-to-left. The instruments were returned to the NRC of each country with the verifier's suggestions (Translation Verification Report). The NRCs were responsible for reviewing the translation verifier's suggestions and revising the instruments. NRCs also had the opportunity to comment on any aspect of translation verification, on the Translation Verification Summary report.

Verifiers were instructed to compare the translated version of each document with the international version—sentence by sentence, rather than word by word—ensuring that all the information in the international version was also in the translated version, and that the latter was fluent. If the translated text differed in any way from the international version, verifiers documented the deviations. Where, in the judgment of the verifier, the translated version of an achievement or questionnaire item deviated from the international version, the translation verifier indicated the severity of the deviation (using a severity code as defined below), a description of the change, and a suggested alternative translation.

Instructions to verifiers included a list of "severity codes," which were used to help identify the nature and severity of any deviation. The severity codes ranged from 1 (serious error) to 4 (acceptable adaptation). The severity codes were defined as follows:

1. **Major Change or Error**: Examples include incorrect order of response options in a multiple-choice item; omission of an item or a graphic; incorrect order of item or question; incorrect translation resulting in the answer being suggested by the item; and an incorrect translation which changes the meaning or difficulty of the item or question.

2. **Minor Change or Error**: Examples include spelling errors that do not affect comprehension or purely linguistic errors that do not affect content or equivalence.

3. **Suggestion for Alternative**: The translation may be adequate, but the verifier has suggested a different wording.

4. **Acceptable Change:** (also known as "appropriate but undocumented adaptation") the change was acceptable and appropriate.

3.7.4 Translation Verification Summary

Following translation verification, National Research Coordinators were asked to document all the verifier comments marked with a code 1 (or 1?)[2] and indicate

- whether they adopted the suggestion or not during the revision process; and
- why they disagreed with the verifier (in cases where suggestions were not adopted in full).

This summary report served two purposes: first, it alerted NRCs to pay special attention to any verifier comments or interventions marked with "code 1," thus providing an extra check of the more significant suggestions; second, it provided a structured format for NRCs to deliver feedback to verifiers. This feedback was particularly useful

2 When in doubt, verifiers were asked to use code 1 with a question mark (1?)

following the field trial, and feedback was delivered to the verifier who was working on the instruments for the main data collection.

The Translation Verification Summary forms were forwarded to the TIMSS & PIRLS International Study Center, where they were reviewed for any misunderstandings or mistranslations. Furthermore, the forms could be reviewed following data collection and analysis, if any item characteristics suggested an error. NRCs were not obliged to implement suggestions from the verifier; however, they did take responsibility for any errors or mistranslations in the instruments. For the field trial, all participants returned completed Translation Verification Summary forms (or indicated that there were no code 1 errors, or that they had accepted all code 1 interventions). For the data collection consisting of the field tested and trend items, only 5 of 10 participants returned the completed forms.

3.8 Summary

The rigorous procedures for translation, national adaptations, translation verification, and review of the instruments implemented for TIMSS Advanced 2008 provided for comparable translations of the instruments across participating countries. The verification process—consisting of internal review, external translation verification by bilingual judges, review by the TIMSS & PIRLS International Study Center, careful item analysis of the data from the field test and the assessment, and checking by the international QCMs—proved to be a comprehensive program for translation verification, helping to ensure the validity of the data analyzed and reported for TIMSS Advanced 2008.

References

IEA Secretariat. (2007). *Guidelines for Translation Verification of the TIMSS Advanced 2008 Instruments*. Amsterdam: IEA

TIMSS & PIRLS International Study Center. (2007). *TIMSS Advanced 2008 survey operations procedure unit 3: Preparing materials for the TIMSS Advanced 2008 data collection and administering the assessment.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.


OCCASVS
CREPVSCVLVM

Chapter 4

TIMSS Advanced 2008 Sampling

Sylvie LaRoche, Olaf Zuehlke, and Marc Joncas

4.1 Overview

This chapter describes the sample design developed and implemented for TIMSS Advanced 2008 and the derivation of sampling weights for the study. It explains how the target populations were defined in the participating countries and how the national sample designs were developed and implemented. It also explains how the sampling weights and participation rates were calculated. To complement the general information in this chapter, Appendix B provides specific details about the national sample designs and their implementation, including details of population coverage and exclusions, stratification variables, and sample size allocation.

TIMSS Advanced 2008 proposed a uniform sample design to all participants to ensure that differences in survey results were not attributable to the use of different sampling methodologies. This uniform sample design was flexible enough to accommodate the distinctiveness of national school systems at the upper secondary level and how the target populations were defined across participating countries. The TIMSS Advanced 2008 National Research Coordinators (NRCs) were responsible for developing their national sample designs and implementing them in their own countries, with the support of the sampling consultants.[1]

4.2 The TIMSS Advanced 2008 Target Populations

TIMSS Advanced 2008 measured student achievement in two student populations at the end of secondary schooling: advanced mathematics students and physics students. To allow for meaningful interpretation of the TIMSS Advanced 2008 data, and to ensure the comparability of the results across the participating countries, it was important that both target populations be accurately and consistently defined.

The TIMSS Advanced 2008 target population for advanced mathematics was defined as *the students in the final year of secondary schooling who have taken courses in advanced mathematics*. For physics, the TIMSS Advanced 2008 target population was defined as *the students in the final year of secondary schooling who have taken courses in physics.*

4.2.1 Courses in Advanced Mathematics and Physics

The courses that would define the target populations had to cover most, if not all, of the advanced mathematics and physics topics that were outlined in the *TIMSS Advanced 2008 Assessment Frameworks* (Garden, et al., 2006). The students attending these courses were likely to be the most advanced mathematics or physics students in the final year of upper secondary school in the participating countries. It was the responsibility of the NRCs to identify these advanced mathematics courses and physics courses. In many cases, the courses were found in specific academic programs, or tracks, in upper secondary schools.

1 The team of statisticians from Statistics Canada and the Sampling Unit of the IEA Data Processing and Research Center (DPC)–under the responsibility of Statistics Canada–served as sampling consultants for TIMSS Advanced 2008.

Exhibit 4.1 and Exhibit 4.2 give an overview of the national target population definitions for advanced mathematics and physics, respectively, in terms of the courses or programs in which the eligible students would be found. In all instances, these students were in their final year of secondary schooling; although this meant the students had varied numbers of years of schooling across the participating countries and were of different average age. Chapter 1 and Chapter 7 of the *TIMSS Advanced 2008 International Report* (Mullis, Martin, Robitaille, & Foy, 2009) provide descriptions of the upper secondary educational systems in the participating countries and how the national target populations were ultimately defined.

EXHIBIT 4.1 – TIMSS Advanced 2008 Advanced Mathematics Populations

Country	Advanced Mathematics Population
Armenia*	Students in the 11th grade in "Physmat" schools
Iran, Islamic Rep. of	Students in the 12th grade in the advanced mathematics and physics track in the pre-university stage
Italy	Students in the 13th grade in an advanced mathematics program or an advanced mathematics and physics program, found in Liceo Scientifico (general schools with scientific focus), Liceo Scientifico Tecnologico (general schools with focus on technology) and Instituti Technici (vocational full time training)
Lebanon	Students in the 12th grade in the general science program
Netherlands	Students in the 12th grade who had taken the advanced mathematics course Math B2 in the pre-university track (VWO)
Norway**	Students in the 13th grade who had taken the 3MX advanced mathematics course in the natural science program of the academic track
Philippines	Students in the 10th grade who had taken advanced mathematics courses and attended either a "science and technology oriented" high school, a regional science high school, a university rural high school and laboratory school, or other public science high school
Russian Federation	Students in the 11th grade who had taken 6 hours or more per week of instruction in mathematics
Slovenia	Students in the 12th grade in general gymnasia programs
Sweden	Students in the 12th grade in the natural science program and the technology program who had taken the mathematics D course and may have taken the mathematics E course

* As a result of recent reforms to increase the number of years of school, Armenian students were assessed in what is now called the 11th grade. However, since the assessed students skipped a grade as part of implementing the reforms, they have had 10 years of formal schooling.

** After implementing a curriculum reform, the Norwegian school system consists of 13 years of schooling. However, having started school before the reform was implemented, the students in the TIMSS Advanced 2008 target population, although in the 13th grade, had just 12 years of schooling.

Exhibit 4.2 TIMSS Advanced 2008 Physics Populations

Country	Physics Population
Armenia*	Students in the 11th grade in "Physmat" schools
Iran, Islamic Rep. of	Students in the 12th grade in the advanced mathematics and physics track in the pre-university stage
Italy	Students in the 13th grade in an advanced mathematics and physics program, found in Liceo Scientifico (general schools with scientific focus), Liceo Scientifico Tecnologico (general schools with focus on technology) and Instituti Technici (vocational full time training)
Lebanon	Students in the 12th grade in the general science program
Netherlands	Students in the 12th grade who had taken the Physics 2 course in the pre-university track (VWO)
Norway**	Students in the 13th grade who had taken the 3FY physics course in the natural science program of the academic track
Russian Federation	Students in the 11th grade who had taken 3 hours or more per week of instruction in physics
Slovenia	Students in the 12th grade in general gymnasia programs who chose to take an additional physics course in their final year
Sweden	Students in the 12th grade in the natural science program and the technology program who had taken the physics B course

* As a result of recent reforms to increase the number of years of school, Armenian students were assessed in what is now called the 11th grade. However, since the assessed students skipped a grade as part of implementing the reforms, they have had 10 years of formal schooling.

** After implementing a curriculum reform, the Norwegian school system consists of 13 years of schooling. However, having started school before the reform was implemented, the students in the TIMSS Advanced 2008 target population, although in the 13th grade, had just 12 years of schooling.

4.2.2 TIMSS Advanced Coverage Indices

In order to quantify the proportion of the school-leaving age cohort taking advanced mathematics and physics courses, TIMSS Advanced computed a TIMSS Advanced Mathematics Coverage Index (TAMCI) and a TIMSS Advanced Physics Coverage Index (TAPCI) for each participating country. In part, these indices reflect the overall sampling coverage of the respective populations in each country; but, more importantly, they show that only a very select group of final-year students were considered eligible for TIMSS Advanced 2008, and that the percentages of these students varied across countries.

The TIMSS Advanced coverage indices for advanced mathematics and physics were defined as follows:

$$TAMCI = \frac{\text{Estimated total number of students in the advanced mathematics population}}{\text{Total national population in the corresponding age cohort}} \times 100\%$$

$$TAPCI = \frac{\text{Estimated total number of students in the physics population}}{\text{Total national population in the corresponding age cohort}} \times 100\%$$

The numerator in each index is the total number of students eligible for TIMSS Advanced 2008, either for advanced mathematics or physics, as estimated from the weighted samples. The denominator is an estimate of the size of the eligible age cohort size in 2008 corresponding to the mean age of the target population. The TIMSS Advanced coverage indices are presented in Exhibit 4.3. The final-year age cohort for each country was defined to be the age corresponding to its average age at the time of testing, as estimated from the weighted samples, and its size was estimated from national census figures. The estimated target populations were estimated from the weighted samples.

Exhibit 4.3 TIMSS Advanced 2008 Coverage Indices

Country	Years of Formal Schooling	Final-year Age Cohort	Estimated Size of Final-year Age Cohort	Estimated Target Population		TIMSS Advanced Coverage Indices	
				Advanced Mathematics	Physics	Advanced Mathematics	Physics
Armenia	10	18	62,758	2,684	2,684	4.3%	4.3%
Iran, Islamic Rep. of	12	18	1,705,000	111,298	111,908	6.5%	6.6%
Italy	13	19	605,507	119,162	23,176	19.7%	3.8%
Lebanon	12	18	79,784	4,702	4,724	5.9%	5.9%
Netherlands	12	18	205,200	7,091	6,889	3.5%	3.4%
Norway	12	19	61,093	6,668	4,181	10.9%	6.8%
Philippines	10	16	1,900,656	14,007	—	0.7%	—
Russian Federation	10/11	17	2,073,041	29,672	52,934	1.4%	2.6%
Slovenia	12	19	21,815	8,836	1,635	40.5%	7.5%
Sweden	12	19	125,923	16,116	13,873	12.8%	11.0%

4.3 The Sample Design

The basic TIMSS Advanced 2008 sample design consisted of two sampling stages: schools were sampled at the first stage, and one or more intact classes of students were sampled from a list of eligible classes within a selected school at the second stage. Two methods were used to sample schools in TIMSS Advanced 2008. In countries where the number of schools in the population greatly exceeded the number required in the sample, a systematic probability-proportional-to-size (PPS) sampling method was used. This method, followed by the selection of classes within the selected schools in a second sampling stage, is often referred to as systematic two-stage *PPS* sampling and is described in most sampling textbooks (*e.g.*, Cochran 1977, Lohr 1999). The PPS sampling approach was used in Iran, Italy and the Russian Federation. In other countries where the number of schools to sample from was relatively small, schools were sampled with equal probabilities. This was the case in Lebanon, the Netherlands, Norway, the Philippines and Sweden. In Armenia and Slovenia, a census of schools was taken. In all countries, classes were sampled within selected schools using a random systematic sampling method.

National sample designs had to take into account the expected overlap across the advanced mathematics and physics populations. In some countries, students in a specific program belonged to both advanced mathematics and physics populations. In other countries, eligible students were found in two programs: students in one program belonged to both populations, while students from the other program belonged only to the advanced mathematics population. Finally, in a third category of countries students were free to choose the courses they were to attend and thus the degree of overlap between the two populations could not be predicted. Thus, two principal sample designs—a single school sample and separate school samples—were developed. While countries that participated in TIMSS Advanced 2008 adopted one of these two sample designs, some opted for slight modifications to account for particular national circumstances. Since the Philippines participated only in advanced mathematics, a single school sample was sufficient.

Countries were encouraged to use explicit and implicit stratification to ensure good representation of specific population groups in the national samples and to increase the efficiency of the national sample designs. Explicit stratification was often required to separate schools according to the populations found in schools—schools with advanced mathematics only, physics only, or with both populations. For example, the sampling frame for Norway was divided into a total of six explicit strata based on the populations present and the size of the schools. Appendix B describes the use of explicit and implicit stratification in the participating countries.

4.3.1 Sample Design for Completely Overlapping Populations

This sample design was implemented in countries where there was complete overlap of both the advanced mathematics and physics populations and consisted of selecting a single sample of schools and one or more classes for both populations. Students in each sampled class were randomly assigned an advanced mathematics booklet or a physics booklet. Consequently, about half of the students received an advanced mathematics test booklet while the other half received a physics test booklet. Armenia, Iran and Lebanon implemented this design.

4.3.2 Sample Design for Partially Overlapping Populations

This sample design was implemented in countries where students belonged to either, or even both, populations with no discernible pattern as students were free to choose which courses they would attend. In order to streamline the within-school operations and avoid testing students twice, this sample design consisted of selecting two separate school samples. Both samples of schools were selected simultaneously to prevent overlap. In one school sample, only the advanced mathematics classes were listed for class sampling, and students in the sampled classes were assigned one of the four advanced mathematics test booklets. In the other school sample, only physics classes were listed for class sampling, and students in the sampled classes were assigned one of the four physics test booklets. The Netherlands, Norway and Sweden used this sample design.

In Sweden, an additional sampling step was introduced after the selection of the school samples. In schools where the two programs of interest—natural science and technology—were present, classes from only one program were sampled to keep response burden to a minimum and simplify the within-school operations. Therefore, the sample of two-program schools was randomly divided into natural science and technology explicit strata and each school's sample of classes was drawn from the classes of the relevant program.

4.3.3 Special Adaptation for the Russian Federation

In the Russian Federation, a sample of regions was selected prior to the sampling of schools. Approximately half of the regions were sampled. Regions were selected with probability proportional to size, the largest regions being sampled with certainty. Thus, the sample of regions consisted of a group of certainty regions and a group of sampled regions. In a second stage, school samples were selected from the participating regions, with each school being assigned to only one population—advanced mathematics or physics.

From the group of certainty regions, all schools were grouped into three explicit strata, regardless of region, according to the TIMSS

Advanced populations found in the schools: schools with advanced mathematics classes, schools with advanced mathematics classes and physics classes, and schools with physics classes. Regions were used as implicit strata within each explicit stratum. The sample of schools for advanced mathematics was selected among the first two strata while the sample of schools for the physics sample was selected among the second and third strata. No overlap was allowed in the stratum of schools with both study populations; hence schools could be selected for only one population.

For the group of sampled regions, the sampling of schools was done within each sampled region individually—regions being the primary sample units—and schools within each sampled region were split into two groups. The sample of schools for advanced mathematics was selected from the first group of schools, comprising all schools with only the advanced mathematics population and half of the schools, randomly selected, where both populations were found. Conversely, the sample of schools for physics was selected from the second group of schools: all schools with only the physics population and the remaining half of the schools where both populations were found.

4.3.4 Special Adaptation for Italy

In Italy, the complex structure of advanced mathematics and physics education in the schools and classes required a combination of the two established sample designs, since the courses of interest were found in a program with advanced mathematics classes and another program with advanced mathematics classes and physics classes. Schools were stratified according to the three possible combinations: schools with the advanced mathematics program, schools with both programs, and schools with the advanced mathematics and physics program.

The sample design adopted for Italy mostly followed the sample design for partially overlapping populations, as described in section 4.3.2, whereby only one of the subjects would be assessed in most sampled schools. In order to meet the sample size requirements, however, both subjects were tested in some sampled schools where both populations could be found.

First, a sample of schools was selected from each of the three strata. Schools from the first stratum were eligible only for the advanced mathematics sample. Of the schools sampled from the second stratum, approximately half were randomly assigned to both the advanced mathematics and the physics samples, while the remaining schools were assigned only to the physics sample. In schools sampled for both populations, one class was sampled from a list of classes from the advanced mathematics program and one class was sampled from a list of classes from the advanced mathematics and physics program. While students from the sampled class of the advanced mathematics program received only advanced mathematics booklets, half of the students from the sampled class of the advanced mathematics and physics program received an advanced mathematics booklet and the other half received a physics booklet. Of the schools sampled from the third stratum, one sixth were randomly assigned to the advanced mathematics sample, while the remaining schools were assigned to the physics sample.

4.3.5 Special Adaptation for Slovenia

In Slovenia, the relatively small student populations made it impossible to meet the TIMSS Advanced 2008 student sample size requirements with either of the two standard sample designs. In particular, all physics students in the country had to be selected. Moreover, all schools were selected for both subjects given the small number of schools in the country.

In each school, the advanced mathematics classes and the physics classes were listed separately. A sample of classes was drawn from the list of advanced mathematics classes while all classes from the list of physics classes were selected. Since some students in the selected physics classes could have been sampled for advanced mathematics as well, some students were assessed for both subjects. The order in which the two assessments was administered was determined randomly in each school.

4.3.6 Replacement Schools

Although all participating countries strove to achieve participation rates of 100 percent, this was not always possible. To avoid sample losses, the TIMSS Advanced 2008 sample design identified replacement schools for each sampled school whenever possible. In general, the school immediately preceding and the school immediately following a sampled school on the ordered school sampling frame were designated as replacement schools, and always within the same explicit stratum. Since schools were grouped into strata and ordered by size, it was expected that the characteristics of the replacement schools would be similar to those of the originally sampled schools they were intended to replace. This strategy was implemented in Iran, Italy, Lebanon, the Philippines and the Russian Federation.

The Netherlands did not have designated replacement schools as there were not enough schools left after sampling. An alternative procedure was implemented, whereby a list of replacement schools identified by the sampling consultants was provided to the NRCs. These replacement schools were used in their order of appearance on the list, as necessary.

In Armenia and Slovenia, there were no replacement schools, as all eligible schools were in the sample for both populations. In Norway and

Sweden, since all schools were selected for the advanced mathematics sample or for the physics sample, there were no replacement schools available either.

4.3.7 Sampling for the Field Test

Prior to the TIMSS Advanced 2008 data collection, an extensive field test was conducted during March and April of 2007 in 8 of the 10 participating countries. The goal of this field test was to check all instruments—particularly the achievement tests—and operational procedures under conditions similar to those of the data collection.

The goal was to select approximately 25 schools for the field test, which would yield 600 tested students in advanced mathematics, and 600 tested students in physics. Appendix B provides details on the sampling carried out for the field test in each participating country.

4.4 Sampling Precision and Sample Sizes

TIMSS Advanced 2008 set high standards for sampling precision to guarantee that the survey estimates would be accurate, thereby making comparisons between and within countries meaningful. The goal was to achieve a 95 percent confidence interval for the estimate of a national student-level mean to be within 10 percent of its standard deviation. Because the TIMSS Advanced achievement scales have a standard deviation of 100 points, this would translate into standard errors of approximately 5 points.

With this in mind, and taking into account the clustered nature of the samples and the added uncertainty stemming from the imputation used in scaling the achievement data (see Chapter 8), the minimum sample sizes required were set at 2,000 tested students for advanced mathematics and 2,000 for physics, selected from a minimum of 120 schools. These minima were fixed after looking at the sample sizes and precision achieved with the TIMSS Advanced 1995 results. As these were minima, most countries increased their sample sizes to account for non-response. For the Russian Federation, the sample size was increased further because of the additional clustering effect due to sampling regions before sampling schools. The selected and achieved national school sample sizes are presented Appendix B.

4.5 Sampling Implementation

Sound and rigorous sampling of schools and students was a key quality component of TIMSS Advanced 2008. The TIMSS Advanced sampling consultants selected the school samples for all countries but one[2] and trained NRCs in selecting probability samples of classes and students using IEA's within-school sampling software (IEA, 2007) provided by the IEA DPC. As an essential part of their sampling activities, NRCs were responsible for providing detailed documentation describing their national sampling procedures: target populations, school sampling frames, school sample selection if conducted by the NRC and within-school sampling procedures. The documentation submitted by each TIMSS Advanced participant was reviewed by the sampling consultants, who then provided additional information, including coverage and exclusion levels, stratification variables, sampling methods, participation rates, and preliminary variance estimates. The TIMSS & PIRLS International Study Center and the TIMSS Advanced 2008 Sampling Referee, Dr. Keith Rust of Westat, Inc., used this information to evaluate the quality of the samples. All participating countries produced acceptable implementations of the TIMSS Advanced 2008 sample design.

4.5.1 Population Coverage and Exclusions

All participating countries were able to provide full coverage of their defined target populations of advanced mathematics students and

2 Italy selected its own school samples. The procedures used and the samples drawn were verified and approved by the sampling consultants.

physics students. However, countries were allowed specific types of exclusions of schools and students that would have been either too difficult or too costly to assess. For example, very small or remote schools were sometimes excluded. Within some selected schools, students with special needs or students not fluent in the language of the test were sometimes excluded. Exhibit 4.4 summarizes population coverage and exclusions for the TIMSS Advanced 2008 advanced mathematics and physics populations. For every participant, the overall percentage of excluded students (combining school-level and within-sample exclusions) was less than 5 percent. Some TIMSS Advanced 2008 participants had no within-school exclusions. Details on national exclusion categories are presented in Appendix B.

Exhibit 4.4 Coverage and Exclusions of TIMSS Advanced 2008 Populations

Advanced Mathematics

Country	Coverage	School-level Exclusions	Within-sample Exclusions	Overall Exclusions
Armenia	100%	0.0%	0.0%	0.0%
Iran, Islamic Rep. of	100%	0.0%	0.0%	0.0%
Italy	100%	0.0%	0.5%	0.5%
Lebanon	100%	1.3%	0.0%	1.3%
Netherlands	100%	2.5%	0.0%	2.5%
Norway	100%	0.9%	0.1%	1.0%
Philippines	100%	0.0%	0.0%	0.0%
Russian Federation	100%	0.0%	0.0%	0.0%
Slovenia	100%	0.0%	1.3%	1.3%
Sweden	100%	1.5%	0.2%	1.7%

Physics

Country	Coverage	School-level Exclusions	Within-sample Exclusions	Overall Exclusions
Armenia	100%	0.0%	0.0%	0.0%
Iran, Islamic Rep. of	100%	0.0%	0.0%	0.0%
Italy	100%	0.0%	0.9%	0.9%
Lebanon	100%	1.3%	0.0%	1.3%
Netherlands	100%	2.5%	0.2%	2.7%
Norway	100%	0.4%	0.0%	0.5%
Russian Federation	100%	0.0%	0.0%	0.0%
Slovenia	100%	0.0%	0.5%	0.5%
Sweden	100%	2.1%	0.1%	2.3%

4.5.2 Population and Sample Sizes

The minimum school sample size required to meet the TIMSS Advanced sampling standards was 120 schools for each study population. All but three countries complied with this rule. In Armenia, the number of schools with eligible advanced mathematics and physics students was 38 and all of them were selected. Italy was given permission to select 100 schools for advanced mathematics and 112 for physics. In Slovenia, there were only 87 schools with advanced mathematics students, of which 66 also had physics students; all 87 schools were sampled. Most countries sampled one or two classes per sampled school. Details on the national samples of schools and classes are provided in Appendix B.

Exhibit 4.5 displays the number of eligible schools and students in each country's target populations, based on the sampling frame used to select the TIMSS Advanced 2008 school samples and after removing any excluded schools. This exhibit also includes the number of sampled schools and students that participated in the assessments and provides an estimate of the student population size based on the student samples. The estimate of the student population size (the sixth column of Exhibit 4.5) was derived using sampling weights, while the population of students (the third column) was taken from the national sampling frames. The estimated populations should be fairly close to the student populations taken from the sampling frames. Differences are attributable to within-sample exclusions and to changes in populations from the time the sampling frames were created to the time the TIMSS Advanced 2008 assessments were conducted. The observed differences are largest for Iran and Italy. In Iran, many sampled schools were closed, which explains the lower estimated population from the sample. In Italy, many sampled schools were found to be ineligible as they did not have any eligible students. Also, many sampled schools had fewer eligible students than expected in both populations, especially so in physics.

Exhibit 4.5 TIMSS Advanced 2008 Population and Sample Sizes

Advanced Mathematics

Country	Population		Sample			Average Age at Time of Testing
	Schools	Students	Schools	Students	Estimated Population	
Armenia	38	2,755	38	858	2,684	17.7
Iran, Islamic Rep. of	3,187	123,802	119	2,425	111,298	18.1
Italy	1,318	149,558	91	2,143	119,162	19.0
Lebanon	345	5,037	212	1,615	4,702	17.9
Netherlands	493	6,906	112	1,537	7,091	18.0
Norway	253	7,424	107	1,932	6,668	18.8
Philippines	165	15,105	118	4,091	14,007	16.4
Russian Federation	1,031	29,285	143	3,185	29,672	17.0
Slovenia	87	9,945	79	2,156	8,836	18.8
Sweden	328	11,934	116	2,303	16,116	18.8

Physics

Country	Population		Sample			Average Age at Time of Testing
	Schools	Students	Schools	Students	Estimated Population	
Armenia	38	2,755	38	894	2,684	17.7
Iran, Islamic Rep. of	3,187	123,802	119	2,434	111,908	18.0
Italy	642	31,163	91	1,861	23,176	18.9
Lebanon	345	5,037	210	1,600	4,724	17.9
Netherlands	493	6,906	116	1,511	6,889	18.1
Norway	236	4,404	101	1,642	4,181	18.8
Russian Federation	1,875	54,782	149	3,166	52,934	17.1
Slovenia	66	1,752	54	1,120	1,635	18.7
Sweden	317	10,134	121	2,291	13,873	18.8

4.6 Calculating Sampling Weights

The estimation method used to produce estimates of totals from TIMSS Advanced 2008 data was a simple weighted sum of all responding students for any variable of interest. Estimates of percentages or means were taken as ratios of these estimated totals. The national sample designs implemented in TIMSS Advanced generally involved varying selection probabilities for schools, classes, and students that required specific sampling weights for each participating class, which were assigned to each individual student. The sampling weights were computed separately for each TIMSS Advanced population and within each explicit stratum.

The overall sampling weights assigned to individual students comprise a series of multiplicative components. A basic overall weight was derived from the inverse of the probability of selecting a student from the population. This basic overall weight was adjusted by multiplicative factors that account for non-responding schools, classes, and students. For some countries, additional adjustments were required to account for additional sampling steps.

Sampling weights were calculated according to a three-step procedure involving selection probabilities for schools, classes, and students. The first step consisted of calculating a school weight, which also incorporated weighting factors from any additional front-end sampling stages, such as regions in the Russian Federation. A school-level participation adjustment also was computed to compensate for any sampled schools that did not participate and were not replaced.

In a second step, a class weight was calculated, reflecting the probability of the selected class(es) being sampled among all the eligible classes in a school. This class weight was calculated independently within each participating school. A class-level participation adjustment was computed to compensate for sampled classes that did not participate, or if the participation rate among students in a class fell below 50 percent. The class participation adjustment was computed at the explicit stratum level, rather than at the school level, to reduce the potential for response bias.

The third and final step consisted of calculating a student weight, which was computed according to the particular sample design implemented in each participating country. In countries where separate school samples were selected for each population, intact classes were sampled and all students were selected with certainty. Thus, the student weight was set to 1. In countries where a single school sample was selected for both populations, and both advanced mathematics booklets and physics booklets were distributed in the sampled classes, students within the sampled classes were randomly selected for one population or the other. Thus, the student weight was calculated within each class to reflect the probability of a student being selected for a specific population. A student participation adjustment then was made to compensate for sampled students who did not take part in the assessments. This adjustment was calculated separately for each sampled class.

The basic overall sampling weight assigned to each student was the product of the three basic weights—the school weight, the class weight, and the student weight. The final overall sampling weight was the product of the basic overall sampling weight and the three participation adjustments.

4.6.1 The School Weight

In general, the school weight represents the inverse of the probability of a school being sampled at the first stage. In the national sample designs for Iran, Italy, and the Russian Federation, the school selection probabilities were proportional to school size, generally defined as the number of students in the target population. Thus, the basic school weight (with the subscript *sc*) for the i^{th} sampled school of population g (where g takes the value M for advanced mathematics and P for physics) was defined as:

$$BW_{sc}^{g,i} = \frac{MOS}{n \cdot mos_i}$$

Where n was the number of sampled schools in population g, mos_i was the measure of size for the i^{th} school, and

$$MOS = \sum_{i=1}^{N} mos_i$$

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

In order to avoid school weights being less than 1, the size of large schools (schools of size greater than the sampling interval given by MOS/n), was set as the sampling interval. As a result, these large schools were sampled with a probability of 1.

In a similar way, the measure of size for small schools was set to a constant to prevent large variance fluctuations that typically arise from the large school weights that could occur otherwise. As a result, these small schools were sampled with equal probability.

In Armenia, Lebanon, the Netherlands, Norway, the Philippines, Slovenia, and Sweden, equal probability sampling of schools, rather than PPS, was carried out, meaning that every school had the same measure of size ($mos_i = 1$). Thus the school weight for the i^{th} sampled school in population g in these countries was calculated as:

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

Special Weight Factor for Italy

As was mentioned earlier, special weight factors or adjustments were calculated to account for additional sampling steps introduced during school sampling and arising from special adaptations to national sample designs in some countries.

In Italy, while all schools sampled from the stratum of schools with both the advanced mathematics program and the advanced mathematics and physics program were assigned to the physics population, approximately half of them were randomly sub-sampled for the advanced mathematics population. Thus, an additional weight factor for the sub-sampled schools in advanced mathematics was computed as the inverse of the probability of a school sampled from this stratum being selected for advanced mathematics, and the original school weight was multiplied by this additional weight factor.

Special Weight Factors for the Russian Federation

The sample design for the Russian Federation included a preliminary sampling stage, in which regions were sampled. Thus, the school weight also incorporated the probability of selection in this preliminary stage. Hence, the school weight for all schools from the Russian Federation was the product of the "region" weight and the school weight described earlier. This region weight was computed in a manner similar to the school weight, with regions having selection probabilities proportional to their size.

An additional weight factor was required for schools sampled from the group of sampled regions and where both populations were found. This extra factor was required as these schools were randomly assigned to only one population prior to school sampling. Within each region, an additional weight factor was computed as the inverse of the probability of a school being assigned to a specific population. The school weight for these schools was multiplied by this weight factor to account for this additional sampling step.

Special Weight Factor for Sweden

In Sweden, in the stratum comprised of schools with students from both the natural science and the technology programs, an additional sampling step was introduced after school sampling. The sampled schools were randomly allocated to one of the programs and only students from the selected program took part in TIMSS Advanced 2008. An additional weight factor was computed for these sampled schools as the inverse of the probability of a school being selected for a specific program. Hence, the school weight for schools assigned to each specific program was multiplied by this additional weight factor to account for this additional sampling step.

4.6.2 School Participation Adjustment

A school-level participation adjustment was required to compensate for schools that were sampled but did not participate, and were not replaced. Sampled schools that were found to be ineligible[3] were removed from the calculation of this adjustment. The school participation adjustment was calculated separately for each explicit stratum and each population *g*, as follows:

$$A_{sc}^{g} = \begin{cases} \dfrac{n_s + n_{r1} + n_{r2} + n_{nr}}{n_s + n_{r1} + n_{r2}} & \text{for participating schools} \\ 0 & \text{for non-participating schools} \end{cases}$$

where n_s was the number of 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 and were not replaced.

The final school weight assigned to all students in the i^{th} school of population g ($g = M$ or P), corrected for non-participating schools, was:

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

3 A sampled school was ineligible if it was found to contain no eligible students. Such schools usually were in the sampling frame by mistake or were schools that had recently closed.

4.6.3 The Class Weight

The class weight is the inverse of the probability of a class being selected within a sampled school. All countries sampled classes with equal probability. For the i^{th} school sampled in population g, let $C^{g,i}$ be the total number of eligible classes and $c^{g,i}$ the number of sampled classes. The class weight (with the subscript cl) assigned to all sampled classes in the i^{th} school in population g was computed as:

$$BW_{cl}^{g,i} = \frac{C^{g,i}}{c^{g,i}}$$

For most TIMSS Advanced participants, $c^{g,i}$ took the values 1 or 2. Some TIMSS Advanced participants sampled all eligible classes in a selected school, in which case $c^{g,i}$ was equal to $C^{g,i}$.

4.6.4 Class Participation Adjustment

A class-level participation adjustment was applied to compensate for classes that did not participate, or where the student participation rate was below 50 percent. The class participation adjustment was calculated at the explicit stratum level, rather than by school, to minimize the potential for response bias. The adjustment was calculated as follows:

$$A_{cl}^{g} = \begin{cases} \dfrac{n_s + n_{r1} + n_{r2}}{\sum_{i}^{s+r1+r2} \delta^{g,i} / c^{g,i}} & \text{for participating classes} \\ 0 & \text{for non-participating classes} \end{cases}$$

where the summation was over all sampled and replacement schools, $c^{g,i}$ was the number of sampled classes in the i^{th} school, and $\delta^{g,i}$ was the number of participating classes in the i^{th} school of population g within the explicit stratum.

The final class weight assigned to all students in the sampled classes of the i^{th} school of population g was computed as:

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

4.6.5 The Student Weight

The student weight is the inverse of the probability of a student in a sampled class being selected. By design, all students within selected classes were selected for the TIMSS Advanced 2008 assessments. In most cases, they were all assigned a booklet from one subject—either advanced mathematics or physics. In countries with completely overlapping populations, however, roughly half of the students in a class were assigned a booklet in one subject and the other half in the other subject. The student weight (with the subscript *st*) for the j^{th} class in the i^{th} school of population g was calculated as follows:

$$W_{st}^{g,i,j} = \frac{n^{i,j}}{n^{g,i,j}}$$

where the $n^{i,j}$ was the total number of students in the j^{th} class of the i^{th} school and $n^{g,i,j}$ was the number of students in that class selected for population g ($g = M$ or P).

When classes were sampled for only one population, then $n^{g,i,j}$ was equal to $n^{i,j}$ and the probability of a student in a selected class being sampled for that population was equal to 1. When booklets from both populations were distributed among students from a selected class, this probability was approximately one half. In both cases, the student weight was calculated separately for each selected class and for each population.

4.6.6 Student Participation Adjustment

A student-level participation adjustment was calculated at the class level and it was calculated in the same manner, regardless whether a class was selected for a single population or for both populations. The student participation adjustment for the j^{th} class in the i^{th} school of population g was computed as:

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

where $s_{rs}^{g,i,j}$ was the number of students that responded to their assigned population g booklet in the j^{th} class of the i^{th} school, and $s_{nr}^{g,i,j}$ was the number of students that did not respond to their assigned population g booklet in the j^{th} class of the i^{th} school.

The final student weight for students selected for population g in the j^{th} class of the i^{th} school was defined as:

$$FW_{st}^{g,i,j} = A_{st}^{g,i,j} \cdot BW_{st}^{g,i,j}.$$

4.6.7 Overall Sampling Weights

The basic overall sampling weight was the product of the school weight, the class weight, and the student weight. Thus, for each study population g, this product was computed as:

$$BW^{g,i,j} = BW_{sc}^{g,i} \cdot BW_{cl}^{g,i,j} \cdot BW_{st}^{g,i,j}.$$

The final overall sampling weight was the product of the final school, class, and student weights, and was computed as:

$$FW^{g,i,j} = FW_{sc}^{g,i} \cdot FW_{cl}^{g,i,j} \cdot FW_{st}^{g,i,j}.$$

4.7 Calculating Participation Rates

Since non-participating schools, classes, and students can lead to bias in the study results, participation rates were needed to show the degree of success each TIMSS Advanced participant achieved in securing participation from their sampled schools, classes, and students. To monitor school participation, two school participation rates were computed for each population: one based on sampled schools only and one based on sampled and replacement schools. Class and student participation rates were also computed, based on sampled and replacement schools, as were overall participation rates. Both unweighted and weighted participation rates were computed.

4.7.1 Unweighted School Participation Rates

For each population, two unweighted school participation rates were computed, as follows:

$R_{unw}^{g,sc-s}$ = unweighted school participation rate from sampled schools only

$R_{unw}^{g,sc-r}$ = unweighted school participation rate from sampled and replacement schools.

For each population, the rates were defined as the ratio of the number of participating schools to the number of sampled schools, excluding any ineligible schools. A school was labeled a participating school if at least one of its sampled classes had a student participation rate of at least 50 percent. The unweighted school participation rates were calculated as follows:

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

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

4.7.2 Unweighted Class Participation Rates

The unweighted class participation rates were computed as follows:

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

where the summations were over all participating sampled and replacement schools, $c^{g,i}$ was the number of sampled classes, and $c_{*}^{g,i}$ the number of participating classes in the i^{th} school of population g.

4.7.3 Unweighted Student Participation Rates

The unweighted student participation rates were computed as follows:

$$R_{unw}^{g,st} = \frac{\sum_{i,j}^{s+r1+r2} s_{rs}^{g,i,j}}{\sum_{i,j}^{s+r1+r2} \left(s_{rs}^{g,i,j} + s_{nr}^{g,i,j} \right)}$$

where the summations were over all participating schools and classes for population g.

4.7.4 Unweighted Overall Participation Rates

Two unweighted overall participation rates were computed for each TIMSS Advanced population, as follows:

$R_{unw}^{g,ov-s}$ = unweighted overall participation rate from sampled schools only

$R_{unw}^{g,ov-r}$ = unweighted overall participation rate from sampled and replacement schools.

The overall unweighted participation rates were defined as the product of their respective unweighted school participation rates, the unweighted class participation rate, and the unweighted student participation rate. They were calculated as follows:

$$R_{unw}^{g,ov-s} = R_{unw}^{g,sc-s} \cdot R_{unw}^{g,cl} \cdot R_{unw}^{g,st}$$

$$R_{unw}^{g,ov-r} = R_{unw}^{g,sc-r} \cdot R_{unw}^{g,cl} \cdot R_{unw}^{g,st}$$

4.7.5 Weighted School Participation Rates

Two weighted school participation rates were computed for each population, as follows:

$R_{wtd}^{g,sc-s}$ = weighted school participation rate from sampled schools only

$R_{wtd}^{g,sc-r}$ = weighted school participation rate from sampled and replacement schools.

The weighted school participation rates were calculated as follows:

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

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

where the summation in the first numerator was over all responding students in sampled schools and all other summations were over all responding students in all participating schools. What distinguished the numerators from the denominators was the use of the basic school weight rather than the final, or adjusted, school weight in the denominators.

The denominator was the same in both equations and was the weighted estimate of the total enrollment in the target population. The numerators, however, were different. Only students from participating classes and participating sampled schools were included in the first equation. Students from the replacement schools were added in the second equation.

4.7.6 Weighted Class Participation Rates

The weighted class participation rates for both populations were computed as follows:

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

where the summations in both the numerator and denominator were over all responding students from participating classes and schools. The basic class weight appears in the numerator, whereas the final class weight appears in the denominator. Furthermore, the denominator in this formula is the same quantity that appears in the numerator of the weighted school participation rate for all participating schools, including sampled and replacement schools.

4.7.7 Weighted Student Participation Rates

The weighted student participation rates for each population were computed as follows:

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

where the summations in both the numerator and denominator were over all responding students from participating classes and schools. The basic student weight appears in the numerator, whereas the final student weight appears in the denominator. The denominator in this formula is the same quantity that appears in the numerator of the weighted class participation rates.

4.7.8 Weighted Overall Participation Rates

Two weighted overall participation rates were computed for each population, as follows:

$R_{wtd}^{g,ov-s}$ = weighted overall participation rate from sampled schools only

$R_{wtd}^{g,ov-r}$ = weighted overall participation rate from sampled and replacement schools.

The weighted overall participation rates were defined as the product of their respective weighted school participation rates, the weighted class participation rate, and the weighted student participation rate. They were computed as follows:

$$R_{wtd}^{g,ov-s} = R_{wtd}^{g,sc-s} \cdot R_{wtd}^{g,cl} \cdot R_{wtd}^{g,st}$$

$$R_{wtd}^{g,ov-r} = R_{wtd}^{g,sc-r} \cdot R_{wtd}^{g,cl} \cdot R_{wtd}^{g,st}$$

4.7.9 Meeting the TIMSS Advanced 2008 Sampling Participation Standards

All TIMSS Advanced 2008 participants understood that the goals for sampling participation were 100 percent for all sampled schools, classes, and students. Guidelines for reporting achievement data for TIMSS Advanced participants securing less than full participation were modeled after IEA's TIMSS and PIRLS studies. As summarized in Exhibit 4.6, countries were assigned to one of three categories on the basis of their sampling participation. Countries in Category 1 were considered to have met all the TIMSS Advanced 2008 sampling requirements and to have acceptable participation rates. Countries in Category 2 met the participation requirements only after including replacement schools. Countries that failed to meet the participation requirements even with the use of replacement schools were assigned to Category 3. An important goal for quality data in TIMSS Advanced 2008 was to have as many countries as possible achieve Category 1 status.

EXHIBIT 4.6 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 will 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 will be placed in this category 2 if: • It failed to meet the requirements for Category 1 but had a **weighted** school response rate without replacement of at least 50% (after rounding to the nearest percent) AND HAD EITHER • A **weighted** school response rate **with** replacement of at least 85% (after rounding to nearest whole percent) AND a **weighted** student response rate (after rounding) of at least 85% OR • The product of the (unrounded) **weighted** school response rate **with** replacement and the (unrounded) **weighted** student response rate of at least 75% (after rounding to the nearest whole percent). Countries in this category will 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 provided documentation to show that they complied with TIMSS sampling procedures and requirements but did not meet the requirements for Category 1 or Category 2 will be placed in Category 3. Countries in this category will appear in a separate section of the achievement tables, below the other countries, in international reports. These countries will be presented in alphabetical order.

Exhibits 4.7 through 4.10 present the school, class, student, and overall participation rates and achieved sample sizes for all TIMSS Advanced 2008 participants and for both populations.

For advanced mathematics, all participants but one met the TIMSS Advanced sampling requirements and belonged in Category 1. The Netherlands met the requirements only after including replacement schools (Category 2); and, as a consequence, their results were annotated with an asterisk in the advanced mathematics achievement exhibits of the international report. For physics, all countries but two had acceptable participation rates and belonged in Category 1. The Netherlands met the requirements only after including replacement schools (Category 2), and their results were annotated with an asterisk in the physics achievement exhibits in the international report. Slovenia with an overall participation rate of 65 percent belonged in Category 3.

Exhibit 4.7 School Participation Rates and Sample Sizes

Advanced Mathematics

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	100%	100%	38	38	38	0	38
Iran, Islamic Rep. of	99%	99%	120	120	119	0	119
Italy	97%	99%	100	92	88	3	91
Lebanon	86%	89%	240	240	203	9	212
Netherlands	77%	84%	135	133	102	10	112
Norway	94%	94%	120	120	107	0	107
Philippines	98%	98%	121	120	118	0	118
Russian Federation	100%	100%	143	143	143	0	143
Slovenia	96%	96%	87	82	79	0	79
Sweden	90%	94%	127	126	111	5	116

Physics

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	100%	100%	38	38	38	0	38
Iran, Islamic Rep. of	99%	99%	120	120	119	0	119
Italy	100%	100%	112	91	91	0	91
Lebanon	85%	88%	240	240	201	9	210
Netherlands	73%	87%	135	133	98	18	116
Norway	85%	85%	120	120	101	0	101
Russian Federation	100%	100%	149	149	149	0	149
Slovenia	83%	83%	66	64	54	0	54
Sweden	97%	97%	127	125	119	2	121

Exhibit 4.8 Student Participation Rates and Sample Sizes

Advanced Mathematics

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	95%	899	0	0	899	41	858
Iran, Islamic Rep. of	97%	2,556	55	0	2,501	76	2,425
Italy	96%	2,269	15	8	2,246	103	2,143
Lebanon	95%	1,767	36	0	1,731	116	1,615
Netherlands	92%	1,876	200	0	1,676	139	1,537
Norway	89%	2,206	17	2	2,187	255	1,932
Philippines	96%	4,253	3	0	4,250	159	4,091
Russian Federation	98%	3,269	11	0	3,258	73	3,185
Slovenia	85%	2,577	3	22	2,552	396	2,156
Sweden	89%	2,645	26	1	2,618	315	2,303

Physics

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	97%	926	0	0	926	32	894
Iran, Islamic Rep. of	97%	2,556	43	0	2,513	79	2,434
Italy	97%	1,968	18	15	1,935	74	1,861
Lebanon	94%	1,755	35	0	1,720	120	1,600
Netherlands	90%	1,911	203	3	1,705	194	1,511
Norway	86%	1,935	17	1	1,917	275	1,642
Russian Federation	97%	3,269	9	0	3,260	94	3,166
Slovenia	82%	1,404	0	6	1,398	278	1,120
Sweden	92%	2,537	29	4	2,504	213	2,291

Exhibit 4.9 Unweighted School, Class, and Student Participation Rates

Advanced Mathematics

Country	School Participation Before Replacement	School Participation After Replacement	Class Participation	Student Participation	Overall Participation Before Replacement	Overall Participation After Replacement
Armenia	100%	100%	100%	95%	95%	95%
Iran, Islamic Rep. of	99%	99%	100%	97%	96%	96%
Italy	96%	99%	100%	95%	91%	94%
Lebanon	85%	88%	99%	94%	79%	83%
Netherlands	77%	84%	100%	92%	70%	77%
Norway	89%	89%	100%	88%	79%	79%
Philippines	98%	98%	100%	96%	95%	95%
Russian Federation	100%	100%	100%	98%	98%	98%
Slovenia	95%	95%	100%	84%	80%	80%
Sweden	87%	91%	100%	88%	77%	80%

Physics

Country	School Participation Before Replacement	School Participation After Replacement	Class Participation	Student Participation	Overall Participation Before Replacement	Overall Participation After Replacement
Armenia	100%	100%	100%	97%	97%	97%
Iran, Islamic Rep. of	99%	99%	100%	97%	96%	96%
Italy	100%	100%	100%	96%	96%	96%
Lebanon	84%	88%	99%	94%	78%	82%
Netherlands	74%	87%	99%	90%	66%	78%
Norway	84%	84%	99%	86%	72%	72%
Russian Federation	100%	100%	100%	97%	97%	97%
Slovenia	83%	83%	96%	82%	65%	65%
Sweden	95%	97%	100%	91%	87%	89%

Exhibit 4.10 Weighted School, Class, and Student Participation Rates

Advanced Mathematics

Country	School Participation Before Replacement	School Participation After Replacement	Class Participation	Student Participation	Overall Participation Before Replacement	Overall Participation After Replacement
Armenia	100%	100%	100%	95%	95%	95%
Iran, Islamic Rep. of	99%	99%	100%	97%	96%	96%
Italy	97%	99%	100%	96%	93%	95%
Lebanon	86%	89%	99%	95%	81%	83%
Netherlands	77%	84%	100%	92%	71%	77%
Norway	94%	94%	100%	89%	83%	83%
Philippines	98%	98%	100%	96%	95%	95%
Russian Federation	100%	100%	100%	98%	98%	98%
Slovenia	96%	96%	100%	85%	81%	81%
Sweden	90%	94%	100%	89%	80%	84%

Physics

Country	School Participation Before Replacement	School Participation After Replacement	Class Participation	Student Participation	Overall Participation Before Replacement	Overall Participation After Replacement
Armenia	100%	100%	100%	97%	97%	97%
Iran, Islamic Rep. of	99%	99%	100%	97%	96%	96%
Italy	100%	100%	100%	97%	97%	97%
Lebanon	85%	88%	99%	94%	80%	82%
Netherlands	73%	87%	100%	90%	65%	78%
Norway	85%	85%	100%	86%	73%	73%
Russian Federation	100%	100%	100%	97%	97%	97%
Slovenia	83%	83%	98%	82%	67%	67%
Sweden	97%	97%	100%	92%	89%	89%

4.8 Trends in Student Populations

Because one of the major goals of the TIMSS Advanced 2008 assessments was to measure changes in advanced mathematics and physics achievement since 1995 for countries that participated in both assessments, it was important to track any changes in population composition and coverage that might have affected trends in student achievement across the two cycles.

Five of the countries that participated in TIMSS Advanced 2008 also participated in the assessments of advanced mathematics and physics students in their final year of schooling in 1995. The Russian Federation, Slovenia, and Sweden participated in both populations, while Italy participated only in advanced mathematics and Norway participated only in physics. Exhibit 4.11 describes the population definitions used in 1995 for advanced mathematics and physics.

Exhibit 4.11 TIMSS Advanced 1995 Populations

Advanced Mathematics	
Country	**Advanced Mathematics Population**
Italy	Students in their final year in Liceo Scientifico (classical schools) and Instituti Technici (technical schools)
Russian Federation	Students in their final year in general secondary schools who had taken advanced mathematics courses or advanced mathematics & physics courses
Slovenia	All students in their final year in general gymnasia programs
Sweden	Students in their final year in the natural science program and the technology program. It was mandatory for all students from these two programs to take the more advanced Math E course.

Physics	
Country	**Physics Population**
Norway	Students in their final year in the academic program who had taken a three-year physics course
Russian Federation	Students in their final year in general secondary schools who had taken advanced physics courses or advanced mathematics & physics courses
Slovenia	Students in general gymnasia programs who had taken the physics matura exam
Sweden	Students in their final year in the natural science program and the technology program. It was mandatory for all students from these two programs to take the more advanced Physics B course.

Exhibit 4.12 presents five attributes of the national populations sampled in 1995 and 2008: the number of years of formal schooling, the average age at time of testing, the exclusion rates, the TIMSS Advanced coverage indices, and the overall weighted participation rates. More details on the differences between the 1995 and 2008 target population definitions for these countries are provided in the *TIMSS Advanced 2008 International Report* (Martin, Mullis, Robitaille, & Foy, 2009).

Exhibit 4.12 Trends in TIMSS Advanced Student Populations

Advanced Mathematics

Country	Years of Formal Schooling*		Average Age at Time of Testing		Overall Exclusion Rates		Coverage Index (TAMCI)		Overall Participation Rate (After Replacement)	
	2008	1995	2008	1995	2008	1995**	2008	1995	2008	1995
Italy	13	13	19.0	19.1	0.5%	3.8%	19.7%	20.2%***	94.8%	67.5%
Russian Federation	10/11	11	17.0	16.9	0.0%	2.0%	1.4%	2.0%	97.6%	95.9%
Slovenia	12	12	18.8	18.9	1.3%	6.0%	40.5%	75.4%	81.4%	42.4%
Sweden	12	12	18.8	18.9	1.7%	0.2%	12.8%	16.2%	83.7%	88.6%

Physics

Country	Years of Formal Schooling*		Average Age at Time of Testing		Overall Exclusion Rates		Coverage Index (TAPCI)		Overall Participation Rate (After Replacement)	
	2008	1995	2008	1995	2008	1995**	2008	1995	2008	1995
Norway	12	12	18.8	19.0	0.5%	3.8%	6.8%	8.4%	73.0%	83.0%
Russian Federation	10/11	11	17.1	16.9	0.0%	2.0%	2.6%	1.5%	97.3%	95.1%
Slovenia	12	12	18.7	18.8	0.5%	6.0%	7.5%	38.6%	67.1%	43.0%
Sweden	12	12	18.8	18.9	2.3%	0.2%	11.0%	16.3%	89.3%	88.6%

* Represents years of schooling counting from the first year of primary or basic education (first year of ISCED Level 1).

** In 1995 exclusion rates for Advanced Mathematics and Physics were computed based on exclusion rates among all students in the final year of schooling. In the case of the Russian Federation, the figure presented in the 1995 International Report (43.0%) greatly overestimates the level of exclusions in the advanced mathematics and physics populations. The figures presented above (2.0%) include two regions, North Ossetia and Chechen Republic, as well as non-Russian speaking students.

*** The 1995 mathematics coverage index for Italy was recomputed for this report and is different from the figure reported in 1995.

References

Cochran, W.G. (1977). *Sampling techniques*, New York: John Wiley.

Garden, R.A., Lie, S., Robitaille, D.F., Angell, C., Martin, M.O., Mullis, I.V.S., Foy, P., & Arora, A. (2006). *TIMSS advanced 2008 assessment frameworks.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

IEA (2007).*Windows within-school sampling software (WinW3S)* [Computer software and manual]. Hamburg: IEA Data Processing and Research Center.

Lohr, S. (1999). *Sampling: Design and Analysis*, Pacific Grove: Duxbury Press.

Mullis, I.V.S., Martin, M.O., Robitaille, D.F., & Foy, P. (2009). *TIMSS advanced 2008 international report: Findings from IEA's study of achievement in advanced mathematics and physics in the final year of secondary school.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.


OCCASVS

Chapter 5

TIMSS Advanced 2008 Survey Operations Procedures

Ieva Johansone

5.1 Introduction

To standardize survey operations procedures, the TIMSS & PIRLS International Study Center worked with the IEA Secretariat, the IEA Data Processing and Research Center (DPC), and Statistics Canada to develop survey operations procedures for each stage of the assessment, including contacting schools and sampling classes, preparing materials for data collection, administering the assessment, scoring the assessment, and creating the data files. The survey operations were designed to be flexible enough to meet the needs of individual participants, while meeting the high quality standards of IEA.

Each National Center, under the direction of its National Research Coordinator (NRC), was responsible for the implementation of TIMSS Advanced in that country. The NRC was the contact person for all those involved in TIMSS Advanced 2008 within the country, as well as the representative of the country at the international level. The contribution that the NRCs made was crucial to successful survey administration.

Documentation and Software

The TIMSS Advanced 2008 Survey Operations Procedures were described in five units, each accompanied by additional materials (e.g., more specialized manuals and software packages), if necessary. All materials were organized and distributed according to different stages of the study.

The units and accompanying materials are listed below.

- *Unit 1, Parts 1 and 2 – Conducting the TIMSS Advanced 2008 Field Test*
- *Unit 2 – Contacting Schools and Sampling Classes for TIMSS Advanced 2008 Assessment*
- *Unit 3 – Preparing Materials for the TIMSS Advanced 2008 Data Collection and Administering the Assessment*
- *Unit 4 – Scoring the TIMSS Advanced 2008 Assessment*
- *Unit 5 – Creating the TIMSS Advanced 2008 Data Files*
- *The School Sampling Manual*
- *The School Coordinator Manual*
- *The Windows Within-school Sampling Software and Manual* (This software enabled participants to randomly select classes in each sampled school. The software also was used to track school, teacher, student, and student-teacher linkage information; prepare the survey tracking forms; and assign test instruments to students, including printing labels for all the test booklets and questionnaires.)

- *The Test Administrator Manual*
- *The International and National Quality Control Monitor Manuals*
- *The Scoring Guides for Constructed-response Items*
- *The Windows Data Entry Manager Software and Manual* (This software was used for entering, editing, and verifying the TIMSS Advanced 2008 data. Along with the software, countries also received codebooks that described the properties and the layout of the variables to be entered from each TIMSS Advanced 2008 assessment instrument.)
- *The Cross-Country Scoring Reliability Software and Manual (CCSRS)*

5.2 Arranging for Data Collection

Operationally, the data collection involved several preparatory steps, including the field test administration, contacting schools, sampling classes, and completing survey instruments.

TIMSS Advanced 2008 Field Test

The TIMSS Advanced 2008 field test was designed to test the TIMSS Advanced 2008 survey operations procedures, with a small representative sample. All instruments and operational procedures were field-tested under conditions approximating, as closely as possible, those of the final data collection. This enabled the NRCs and their staff to become acquainted with the procedures and refine their national operations, and to provide feedback that was used to improve the procedures for the data collection. The field test contributed significantly to the successful execution of TIMSS Advanced 2008.

Contacting Schools and Sampling Classes

One of the essential first steps in the TIMSS Advanced survey activities was to establish good working relationships with the schools that had been sampled to participate in the study (for more information on all sampling procedures, please refer to Chapter 4). NRCs were responsible for contacting these schools and encouraging them to take part in the assessment. This often involved obtaining support from national or regional educational authorities, depending on the national context.

In cooperation with school principals, national centers identified and trained School Coordinators for all participating schools. The School Coordinator could be a teacher or guidance counselor in the school, but not a teacher of the students who were being assessed. Several national centers had their own personnel fill this role, assigning them several schools in an area. School coordinators were provided with a *School Coordinator Manual* that described their responsibilities in detail, encouraging them to contact the NRC if they had any questions.

The responsibilities of the School Coordinator included providing the necessary information about their school; coordinating the date, time, and place for testing; identifying and training a Test Administrator; coordinating the completion of the Student Tracking Forms[1] and Teacher Tracking Forms[2]; distributing teacher and school questionnaires; and obtaining parental permission (if necessary). School Coordinators also ensured that all testing materials were received by the schools, kept secure at all times, and returned to the national center after the test administration.

A Class Listing Form was sent to each school coordinator to provide information on all the eligible classes in the school. Using this information, the national centers sampled classes within the schools. Intact classes had to be sampled, ensuring that every student

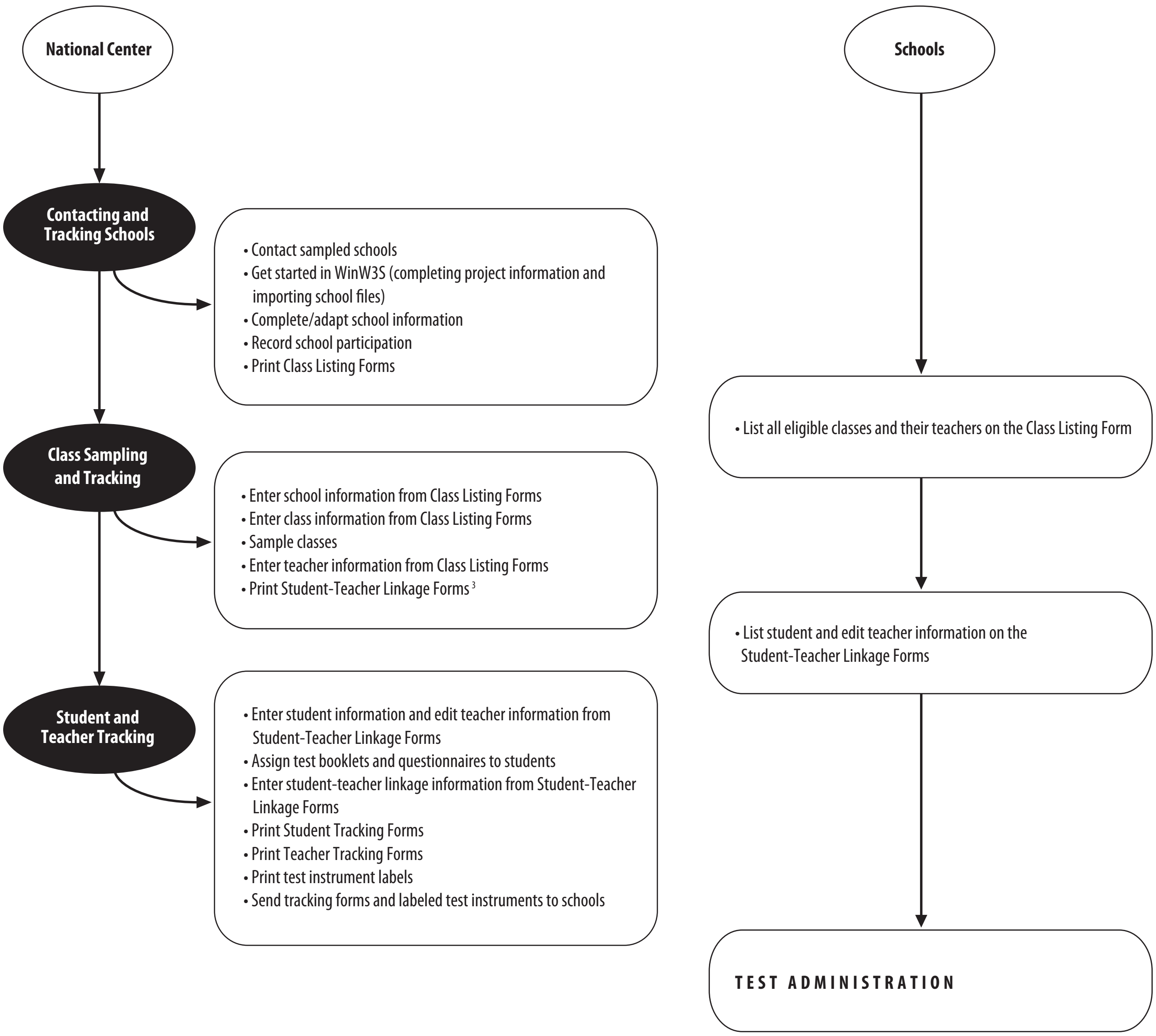
1 This form was created in WinW3S and sent to the schools with the students' test booklets and questionnaires for completion by the Test Administrators during test administration. The Test Administrators used this form to verify the assignment of test instruments to students and to indicate student participation.

2 This form was created in WinW3S and sent to School Coordinators with the teacher questionnaires. The School Coordinators used this form to indicate completion of the teacher questionnaires.

in the school was in only one class (course). Such an organization was necessary for a random sample of classes to result in a representative sample of students.

Exhibit 5.1 presents the major activities conducted by the national centers for working with schools to sample classes; track school, teacher, student, and student-teacher linkage information; and prepare for test administration.

Exhibit 5.1 Procedures for Working with Schools to Prepare for Test Administration



3 This form was created in WinW3S for each sampled class and sent to the school coordinators for completion. The school coordinators listed the names of the students and their teachers in the sampled classes; students' dates of birth, sex, and exclusion codes; and linked the students to their teachers.

Linking Students to their Schools, Classes, and Teachers

To track students, teachers, classes, and schools, a system of hierarchical identification codes (IDs) was set up. The hierarchical identification numbers that uniquely identify the selected schools, classes, students, and teachers were created by the WinW3S software, as shown in Exhibit 5.2.

Exhibit 5.2 Hierarchical Identification (ID) System

Unit	ID Components	ID Structure	Numerical Example
School	School	CCCC	0001
Class	School + Class Within School	CCCCKK	000101
Student	School + Class Within School + Student Within Class	CCCCKKSS	00010103
Teacher	School + Teacher Within School	CCCCTT	000101
Teacher / Class Combination	School + Teacher Within School with Link Number	CCCCTT-LL	000101-02

Since in some cases a teacher might have taught more than one sampled class in a given school, it was necessary to have a unique identification number to distinguish the combinations of teacher and class. This was achieved by creating a two-digit link number so that, in combination with the teacher ID, insured student data were linked to the appropriate teacher data.

Preparing the Test Instruments for Data Collection

The TIMSS Advanced achievement booklets and questionnaires were developed using the Adobe®InDesign® layout program. The TIMSS & PIRLS International Study Center provided countries with all the necessary instrument-production files, including fonts, style guides, graphics files, and explicit instructions (*TIMSS Advanced 2008 Survey Operations Procedures Unit 3*) on how to use the materials in order to produce high quality test instruments.

The goal of the test instrument preparation was to create internationally comparable assessment booklets and background questionnaires that were appropriately adapted for the national context. This began with translating the text of the test instruments from English into the language(s) used in the participating countries. All the translated contents of the test instruments (i.e., item blocks, introductory texts, cover pages, and questionnaires) were submitted to IEA for international translation verification, where independent translators provided suggested changes in the texts (for more information on translation and national adaptations of the TIMSS Advanced 2008 test instruments, please refer to Chapter 3).

Once the translation verification was done and the changes implemented into the test instrument production files, the item blocks, cover pages, and introductory texts were assembled into assessment booklets. The student questionnaires also required some assembly, since the general part was the same for students assessed in advanced mathematics and for students assessed in physics. School and teacher background questionnaires consisted of a single Adobe®InDesign® file and did not require any assembling.

TIMSS & PIRLS International Study Center Review

Before the test booklets and questionnaires were printed and administered to students, NRCs were required to submit a print-ready copy of each test instrument to the TIMSS & PIRLS International Study Center for layout verification and review of national adaptations. The national test instruments were checked against the international version to identify any deviations, and for any discrepancies in pagination, page breaks, item sequence, response options, text formats, graphics, etc.

As a result of the translation process, the test instruments from the participating countries varied in text length. The international versions, however, were designed with extra space in the margins to accommodate the use of longer texts and different sized paper without extensive layout changes. All deviations or errors were documented in the layout verification report forms and sent to the NRCs for their consideration. NRCs were expected to comment on whether or not each of the suggested changes had been made, and to include an explanation if a suggestion was not adopted.

This entire development and production process was designed to ensure that students from different countries experienced the test instruments in the same way, apart from the translation of text.

5.3 Administering the TIMSS Advanced 2008 Assessment

Once they were printed, distributing the materials to the schools required careful organization and planning on the part of the NRC. Using labels and the Student Tracking Form produced by WinW3S, each sampled student was assigned one achievement booklet. The test booklets were assigned in a systematic rotation so that each booklet was assigned to an equal number of students. Each student was also assigned a student questionnaire that was labeled so as to be linked to his or her achievement booklet. These materials were packaged for each sampled class. In addition, a teacher questionnaire was assigned and sent for each teacher listed on the Teacher Tracking Form and a School Questionnaire for the principal. The packaged materials were sent to the School Coordinator prior to the testing date who was asked to confirm the receipt of all instruments. The school questionnaire and teacher questionnaires were distributed, while the other instruments were kept in a secure room until the testing date.

Each sampled class was assigned a Test Administrator whose role was to administer the tests and student questionnaires, according to procedures described in the *Test Administrator Manual*. This person was chosen and trained by the school coordinator, although, in some cases, the School Coordinator also filled the Test Administrator role. The Test Administrator was responsible for distributing materials to the appropriate students, as well as leading students through the assessment and timing the sessions accurately. Following the assessment, they administered the student questionnaires.

The administration of the TIMSS Advanced 2008 assessment consisted of two parts. The first part concerned the achievement booklets and was followed by the completion of the student questionnaire. The time allotted for the achievement test was standardized to 90 minutes, while students were given at least 30 minutes to complete the student questionnaire, and were allowed to continue if extra time was necessary. If a student had completed the achievement test before the allotted time was over, he or she was allowed to review his or her answers or read quietly at his or her table but was not allowed to leave the testing room. The Test Administrators were required to document the starting and ending time of each part on the Test Administration Form.

The Test Administrator used the Student Tracking Form to distribute the booklets to the proper students and to document student participation. The School Coordinator used the information on the participation status to calculate the participation rate. If this was below 90 percent in any class, it was the School Coordinator's responsibility to hold a makeup session for the absent students before returning all of the testing materials to the national center.

The national centers entered the information recorded on the student and teacher tracking forms into the WinW3S software.

5.4 Scoring the TIMSS Advanced 2008 Assessment

About one third of the TIMSS Advanced 2008 assessment items were constructed-response items, and scoring them in a reliable manner was critical to the quality of the TIMSS Advanced 2008 results. Reliable scoring was accomplished through the provision of explicit scoring guides and extensive training in their use, as well as continuous monitoring of the quality of the work.

An international scoring training session was held, where the NRCs (or the country representative(s) appointed by the NRC) were trained to score the constructed-response items in the TIMSS Advanced 2008 assessment. At this session, the *TIMSS Advanced 2008 Scoring Guides for Constructed-response Items* (which are more thoroughly discussed in Chapter 2) were reviewed together with examples of student responses for each category. These example responses represented a range selected to demonstrate the guides as clearly as possible. Following this, NRCs practiced applying the scoring guides to a different set of student responses. The scores NRCs gave to these "practice papers" were shared with the group and all discrepancies discussed. NRCs were given a set of the correct scores for these example papers along with rationales.

To prepare for scoring students responses to the constructed-response items, NRCs were provided with suggestions about how to organize staff, materials, and procedures. NRCs were encouraged to hire scorers who were attentive to detail and familiar with education, particularly those with a background in mathematics or physics instruction at the senior secondary level. The TIMSS & PIRLS International Study Center also provided guidelines about how to train scorers to accurately and reliably score the constructed-response achievement items.

Documenting Scoring Reliability

To establish the reliability of the scoring within each country, two different scorers independently *scored 25 percent of* all student responses. The random sample of test booklets to be scored twice was selected 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. The scoring procedure recommended by the TIMSS & PIRLS International Study Center incorporated scoring the reliability sample together with the normal scoring activity, with both taking place simultaneously in a systematic manner. In collecting the reliability data, reliability scoring sheets were used so that scores were assigned independently.

Estimating the reliability of the scoring process *across* countries necessitated at least two scorers from the TIMSS Advanced scoring team in each country able to score student responses written in English. Computing the level of agreement across countries provides information about how consistently the scoring guides were applied from one country to the next. The student responses included in the cross-country reliability scoring were scanned, stored on CDs, and provided to the countries along with the Cross-Country Scoring Reliability Software (CCSRS), which was developed by the IEA DPC. The CD also included a manual on how to install and use the software.

5.5 Creating the TIMSS Advanced 2008 Data Files

As described earlier in this chapter, the IEA DPC provided a Windows-based program called WinDEM to facilitate data entry and verification. Detailed information on installing and using the program was provided in the manual accompanying the software. The program worked in conjunction with WinW3S software so that it was not necessary to re-enter tracking information that had been recorded into WinW3S.

WinDEM was primarily used for the entry of data from test booklets and questionnaires. The software also offered data and file management capabilities, a convenient checking and editing mechanism, interactive error detection, and reporting and quality-control procedures.

One of the very important benefits of using WinDEM was that it incorporated the international codebooks describing all variables and their characteristics, thus ensuring that the data files met the TIMSS Advanced 2008 standards for data entry. There was one codebook for each of the background questionnaires, one for the test booklets, and one for the constructed-response reliability scoring sheets. Data files for entering the TIMSS Advanced 2008 data were created based on these codebooks. The codebooks however had to match the national instruments exactly so that the answers of the respondents could be entered properly. Any adaptations to the international instruments also required adaptations of the international codebooks.

The adapted national codebooks were used for creating the TIMSS Advanced 2008 data files within each participating country. Data from the background questionnaires, achievement booklets, and reliability scoring sheets were recorded in WinDEM data files as follows:

- School background file contained responses from the *School Questionnaire*.
- Advanced mathematics teacher background file contained responses from the *Advanced Mathematics Teacher Questionnaire*.
- Physics teacher background file contained responses from the *Physics Teacher Questionnaire*.
- Advanced mathematics student background file contained responses recorded from the *Advanced Mathematics Student Questionnaire*.

- Physics student background file contained responses recorded from the *Physics Student Questionnaire*.
- Student achievement file contained responses from the test booklets.
- Reliability scoring file contained codes from the constructed-response reliability scoring sheets.

Quality control throughout the data entry process was essential in maintaining accurate data. Therefore, NRCs were responsible for performing periodic reliability checks during the data entry and for applying a series of data verification checks provided by the WinDEM software prior to submitting the data files to the IEA DPC. As part of this process, NRCs required their data entry staff to double enter at least 5 percent of each instrument type to ensure reliable data entry process. An error rate of 1 percent or less was acceptable for the background files. An error rate of 0.1 percent or less was required for the student achievement files and the reliability scoring files. If the required agreement was not reached, retraining of the data entry staff was required.

The data verification module of WinDEM identified a range of problems such as inconsistencies of identification codes and out-of-range or otherwise invalid codes. WinDEM software also allows to verify the integrity of the linkage between the students, teachers, and schools entered into the WinDEM data files and the tracking information for those specified in WinW3S.

When all data files had passed the WinDEM quality control checks, they were submitted to the IEA DPC along with data documentation for further checking and processing. For information on data processing at the IEA DPC, please refer to Chapter 7.

References

IEA. (2007a). *Cross-country scoring reliability software* [Computer software and manual]. Hamburg: IEA Data Processing and Research Center.

IEA. (2007b). *Windows data entry manager software (WinDEM)* [Computer software and manual]. Hamburg: IEA Data Processing and Research Center,

IEA. (2007c).Windows *Within-school sampling software (WinW3S*) [Computer software and manual]. Hamburg: IEA Data Processing and Research Center.

TIMSS & PIRLS International Study Center. (2006a). *TIMSS Advanced 2008 survey operations procedures unit 1, parts 1 and 2: Conducting the TIMSS Advanced 2008 field test*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

TIMSS & PIRLS International Study Center. (2007a). *TIMSS Advanced 2008 international quality control monitor manual*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

TIMSS & PIRLS International Study Center. (2007b). *TIMSS Advanced 2008 national quality control monitor manual*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

TIMSS & PIRLS International Study Center. (2007c). *TIMSS Advanced 2008 school coordinator manual*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

TIMSS & PIRLS International Study Center. (2007d). *TIMSS Advanced 2008 scoring guides for constructed-response items*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

TIMSS & PIRLS International Study Center. (2007e). *TIMSS Advanced 2008 survey operations procedures unit 2: Contacting schools and sampling classes for TIMSS Advanced 2008 assessment*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

TIMSS & PIRLS International Study Center. (2007f). *TIMSS Advanced 2008 survey operations procedures unit 3: Preparing Materials for the TIMSS Advanced 2008 Data Collection and Administering the Assessment*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

TIMSS & PIRLS International Study Center. (2007g). *TIMSS Advanced 2008 test administrator manual*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

TIMSS & PIRLS International Study Center. (2008a). *TIMSS Advanced 2008 survey operations procedures unit 4: Scoring the TIMSS Advanced 2008 assessment*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

TIMSS & PIRLS International Study Center. (2008b). *TIMSS Advanced 2008 survey operations procedures unit 5: Creating the TIMSS Advanced 2008 data files*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Chapter 6

Quality Assurance in the TIMSS Advanced 2008 Data Collection

Ieva Johansone

6.1 Introduction

To ensure the quality of the TIMSS Advanced 2008 data, considerable effort was made to develop standardized materials and survey operations procedures (see Chapter 5 for more information). In addition, the TIMSS & PIRLS International Study Center developed an ambitious international quality control program to document data collection activities in the participating countries. To implement this program, the IEA Secretariat, in cooperation with national centers, nominated an international Quality Control Monitor (QCM) in each of the participating countries.

The TIMSS & PIRLS International Study Center conducted a QCM training program that involved introducing the QCMs to the TIMSS Advanced 2008 survey operations procedures, the design of the test booklets, and the background questionnaires. During the training, each QCM received the materials necessary to complete their monitoring tasks. The materials included a copy of the *TIMSS Advanced 2008 International Quality Control Monitor Manual* (TIMSS & PIRLS International Study Center, 2007a), Classroom Observation Record, *TIMSS Advanced 2008 Survey Operations Procedures Units 2 and 3 Manuals (*TIMSS & PIRLS International Study Center, 2007d, 2007e), *TIMSS Advanced 2008 School Coordinator Manual (*TIMSS & PIRLS International Study Center, 2007c), and *TIMSS Advanced 2008 Test Administrator Manual (*TIMSS & PIRLS International Study Center, 2007f).

The major task of the international QCMs was to conduct site visits to a random sample of 24 participating schools (12 per subject, which was about 10 percent of the sampled schools). During the site-visits, the QCMs observed TIMSS Advanced classroom testing sessions and discussed the TIMSS Advanced test administration with the Test Adminstrator or School Coordinator. Where necessary, the QCMs were permitted to recruit one or more assistants in order to efficiently cover the territory and testing timetable. A total of 28 international QCMs and their assistants were trained across the 10 countries that participated in TIMSS Advanced 2008. Altogether, these monitors observed 200 testing sessions. The results of the QCM observations are reported in Section 6.2.

In addition to the international quality control program, the National Research Coordinators (NRCs) were asked to complete the *Survey Activities Questionnaire* about their experiences with the TIMSS Advanced 2008 survey operations procedures and the quality of the assessment materials.

6.2 Quality Control Observations of the TIMSS Advanced 2008 Test Administration

For each testing session observed, QCMs completed the TIMSS Advanced 2008 Classroom Observation Record. The observation record was organized into four sections, listed below.

Section A: Preliminary Activities of the Test Administrator
Section B: Test Administration Activities
Section C: Summary Observations
Section D: Interview with the School Coordinator and/or Test Administrator

6.2.1 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. QCMs 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.

Exhibit 6.1 summarizes the results for Section A. In nearly all testing sessions, Test Administrators observed the proper preparatory procedures. For those few deviations that occurred, QCMs provided reasonable explanations. For example, QCMs noted that Test Administrators who did not have a watch with a second hand had a cell phone clock or a classroom clock available to monitor the time remaining in the test session. In general, no procedural deviations were observed that were severe enough to jeopardize the integrity of the test administration.

Exhibit 6.1 Percentages of QCM Responses for Preliminary Activities of the Test Administrator

Question	Yes (%)	No (%)	Not Answered (%)
Had the Test Administrator verified adequate supplies of the test booklets were available?	95	5	0
Had the Test Administrator familiarized himself or herself with the test administration script prior to the testing?	94	6	0
Did the student identification information on the test booklets and student questionnaires correspond with the Student Tracking Form?	99	1	0
Was there adequate seating space for the students to work without distractions?	92	8	0
Was there adequate room for the Test Administrator to move around during the testing to ensure that student were following directions correctly?	93	7	0
Did the Test Administrator have a watch with a seconds hand (or a stopwatch) for timing the testing sessions accurately?	94	6	0

6.2.2 Assessment Session Activities

Section B of the Classroom Observation Record addressed the activities that took place during the assessment session and the administration of the student questionnaire. The activities, such as following the Test Administrator script, distributing and collecting test booklets, and making announcements during the testing session were reported by the QCMs and are presented in Exhibits 6.2 through 6.4.

Activities carried out during the achievement testing session are presented in Exhibit 6.2. One of the most important ways to standardize assessment administrations was following the Test Administrator's script. QCMs reported that, in almost all of their observations, the Test Administrators followed the script exactly when preparing students, distributing test materials, and giving directions. Of the few changes made, most were considered minor. Typically, the script changes were additions rather than revisions or deletions. In the 5 percent of the sessions where the testing time did not equal the time allowed, it was because students had completed the test before the allotted time had elapsed. When the allotted time was over, the Test Administrator instructed students to complete the calculator survey. In some cases, students had completed the test and the calculator survey before the testing time was over. In 98 percent of the cases, the Test Administrator made sure that students stopped working immediately after announcing the end of the testing session.

Exhibit 6.2 Percentages of QCM Responses for the Assessment Session

Question	Yes (%)	No (%)	Not Answered (%)
Did the Test Administrator follow the Test Administrator's script exactly in each of the following tasks?			
Preparing the students	85	14 (Minor changes) 1 (Major changes)	0
Distributing the materials	91	7 (Minor changes) 1 (Major changes)	1
Giving directions	81	15 (Minor changes) 2 (Major changes)	2
If the Test Administrator made changes to the script, how would you describe them?			
Additions	16	7	1 (Not Answered) 76 (Not Applicable)
Revisions	8	15	1 (Not Answered) 76 (Not Applicable)
Deletions	5	17	2 (Not Answered) 76 (Not Applicable)
Did the Test Administrator distribute the test booklets according to the booklet assignment on the Student Tracking Form?	99	0	1
Did the Test Administrator record attendance correctly on the Student Tracking Form?	93	0	7
Was the total time for testing correct as indicated in the script?	94	5	1
Did the Test Administrator announce, "You have 10 minutes left" prior to the end of the testing session?	94	6	0
Were there any other time-remaining announcements made during testing session?	10	90	0
Did the Test Administrator read the script to announce the end the testing session and to ask students to complete the Calculator Survey?	90	10	0
At the end of the testing session, did the Test Administrator make sure all students had closed booklets and stopped working?	98	2	0
Were all booklets collected and secured after the testing session?	99	1	0

After the calculator survey was completed, the Test Administrators announced a break to be followed by the student questionnaire, unless the questionnaire was to be administered on a different date. In such cases, QCMs were not required to observe the questionnaire administration. In 85 percent of all observations, the student questionnaire was administered after a short break following the assessment session.

In most cases, Test Administrators kept to the testing script for signaling a break before administering the student questionnaire. Of those who did make changes, only 3 percent were reported as major changes. There were no observations of students requesting additional time to complete the student questionnaire.

Exhibit 6.3 Percentages of QCM Responses for the Student Questionnaire Administration

Question	Yes (%)	No (%)	Not Answered (%)
When the Test Administrator read the script to end the testing session and the calculator survey, did he/she announce a break to be followed by the student questionnaire?	85	13	2
Did the Test Administrator accurately read the script to end the testing and signal a break?	78	13 (Minor changes) 3 (Major changes)	6
If there were changes, how would you describe them?			
Additions	7	8	8 (Not Answered) 77 (Not Applicable)
Revisions	8	7	8 (Not Answered) 77 (Not Applicable)
Omissions	5	9	9 (Not Answered) 77 (Not Applicable)
Did the Test Administrator distribute the student questionnaires and give directions as specified in the script?	85	5	10
Did the students ask for additional time to complete the questionnaire?	0	93	7
At the end of the testing session, prior to dismissing the students, did the Test Administrator thank the students for participating in the study?	87	11	2

Exhibit 6.4 provides observations on student compliance with instructions and the alignment of the scripted instructions with their implementation. The results show that in all of the sessions, students complied well or very well with the instructions to stop working. Almost always, the dismissal of students had been very orderly or somewhat orderly.

Exhibit 6.4 Percentages of QCM Responses for Student Cooperation

Question	Very Well (%)	Well (%)	Fairly Well (%)	Not Well at All (%)	Not Answered (%)
When the Test Administrator ended the testing session, how well did the student comply with the instruction to stop working?	76	23	0	0	1

Question	Very Orderly (%)	Somewhat Orderly (%)	Not Orderly at All (%)	Not Answered (%)
How orderly was the dismissal of the students?	82	16	1	1

6.2.3 General Observations

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

The results presented in Exhibits 6.5 and 6.6 show that, for most testing sessions, no problems were observed. In 99 percent of all cases, Test Administrators addressed students' questions adequately and as instructed in the *Test Administrator Manual*. In the very few sessions where a defective test instrument was detected, the Test Administrator replaced the instrument appropriately. There were no cases where a student refused to take the test. In 22 percent of the observed testing sessions, a student left the room for an "emergency" during the testing session. In such cases, Test Administrators were instructed to collect the student's test booklet, and give it back after he or she returned. However, in a few cases, students were instructed to close their booklets and leave them on their tables while out of the classroom.

QCMs reported no cases where students were not orderly and cooperative during the testing sessions. There were very few cases where students' orderliness or cooperation was less than perfect or very good. In all such cases, Test Administrators managed to control the situation. QCMs reported that the overall quality of all testing sessions was fair, good, very good, or, in 56 percent of the cases, excellent.

Exhibit 6.5 Percentages of QCM Responses for General Observations

Question	Yes (%)	No (%)	Not Answered (%)
Did the Test Administrator address students' questions appropriately?	99	1	0
Did you see any evidence of students attempting to cheat on the tests (e.g., by copying from a neighbor)?	3	97	0
Were any defective test books detected and replaced before the testing began?	1	99	0
Were any defective test books detected and replaced after the testing began?	1	99	0
If any defective test books were replaced, did the Test Administrator replace them appropriately?	1	1	0 (Not Answered) 98 (Not Applicable)
Did any students refuse to take the test either prior to the testing or during the testing?	0	100	0
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)?	0	0	1 (Not Answered) 99 (Not Applicable)
Did any students leave the room for an "emergency" during the testing?	22	78	0
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)?	18	4	0 (Not Answered) 78 (Not Applicable)

Exhibit 6.6 Percentages of QCM Responses for Observations of Student Behavior

Question	Extremely (%)	Moderately (%)	Somewhat (%)	Hardly (%)	Not Answered (%)	
To what extent would you describe the students as orderly and cooperative?	67	28	5	0	0	
	No, There Were No Late Students (%)	No, They Were Not Admitted (%)	Yes, Before Testing Began (%)	Yes, After Testing Began (%)	Not Answered (%)	
Were any late students admitted to the testing room?	84	3	8	4	1	
	Excellent (%)	Very Good (%)	Good (%)	Fair (%)	Poor (%)	Not Answered (%)
In general, how would you describe the overall quality of the testing session?	56	29	10	5	0	0

6.2.4 Interview with the Test Administrator or School Coordinator

As the final step of each observation, the QCMs conducted an interview with the Test Administrator or School Coordinator. Details of the interview were recorded in Section D of the Classroom Observation Record. The interview addressed activities such as shipment of assessment materials, arrangements for test administration, responsiveness of the NRC to queries, 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 Exhibits 6.7 and 6.8, show that, overall, School Coordinators considered the TIMSS Advanced 2008 administration in their schools a success. Mistakes that did occur tended to be minor and easily remedied. There were only a few cases where shipments of test materials had something missing; and, in all such cases, they were resolved before the testing date. Note that the relatively high percentages of School Coordinators or Test

Administrators not responding to questions about receiving some of the items in Exhibit 6.7 occurred because some countries did not use the particular form or shipment method.

Exhibit 6.7 Receipt of Materials and Test Administration, Percentages of Responses from QCM Interviews with the Test Administrator and/or School Coordinator

Question	Yes (%)	No (%)	Not Answered (%)
Prior to the assessment day did you have time to check your shipment of materials from your TIMSS Advanced National Research Coordinator?	96	3	1
Did you receive the correct shipment of the following items?			
School Coordinator Manual	87	12	1
Test Administrator Manual	99	0	1
Student Tracking Forms	99	0	1
Test booklets	97	1	2
Student Questionnaires	99	0	1
Teacher Questionnaires	99	0	1
School Questionnaire	99	0	1
Test Administration Form	99	0	1
Teacher Tracking Form	80	1	19
Envelopes or boxes addressed to the National Center for the purpose of returning the materials after the assessment	82	4	14
Was the National Research Coordinator responsive to your questions or concerns?	96	1	3
Was the estimated time of 45 minutes to complete the teacher questionnaire a correct estimate?	54	5 (Took longer) 11 (Took less time)	30
Were you satisfied with the accommodations (testing room) you were able to arrange for the testing?	86	4	10
Do you anticipate that a makeup session will be required at your school?	25	62	13
If you anticipate a makeup session, do you intend to conduct one?	16	9	13 (Not Answered) 62 (Not Applicable)
Did the students receive any special instructions, a motivational talk, or incentives to prepare them for the assessment?	65	21	14
Is this a complete list of the classes in this grade in this school?	79	4	17
To the best of your knowledge, are there any students in this grade level who are not in any of these classes?	2	91	7
To the best of your knowledge, are there any students in this grade level in more than one of these classes?	1	92	7
If there was another international assessment, would you be willing to serve as a school coordinator?	84	7	9

In order to better estimate the time needed to complete the teacher questionnaires, QCMs asked if the current estimate of 45 minutes was appropriate. From all cases where teacher questionnaires already were completed, 54 percent of the School Coordinators reported that the estimate of 45 minutes was about right. Five percent reported that the questionnaires took longer, and 11 percent said that they took less time to complete. In more than half the cases, School Coordinators indicated that students were given special instructions, motivational talks, or incentives by a school official or the classroom teacher prior to testing. In 25 percent of the observed classes, the School Coordinator anticipated that a make-up session was needed, and most of them were sure that they would be conducting one.

Because sampling classes required a complete list of all classes in the school at the target grade, QCMs were asked to verify that the class list did indeed include all classes. In spite of complicated course structures in some countries, almost all School Coordinators reported that the complete list of classes had been documented, and that all students appeared in one and only one of these classes. Additional comments from School Coordinators showed that some were very confused by the question: commenting, for example, that the selected class was the only eligible one in the school. Thus, a relatively high percentage of QCMs did not answer the question, marking it as not applicable.

A tribute to the planning and implementation of TIMSS Advanced 2008 was the fact that 84 percent of respondents said they would be willing to serve as a School Coordinator in future international assessments. Furthermore, the results in Exhibit 6.8 suggest that the majority of School Coordinators believed that the testing sessions went very well and that school staff members had mostly positive attitudes towards the TIMSS Advanced 2008 testing.

Exhibit 6.8 Overall Impressions, Percentages of Responses from QCM Interviews with the Test Administrator and/or School Coordinator

Question	Very Well, No Problems (%)	Satisfactorily, Few Problems (%)	Unsatisfactorily, Many Problems (%)	Not Answered (%)
Overall, how would you say the session went?	86	12	1	1

	Positive (%)	Neutral (%)	Negative (%)	Not Answered (%)
Overall, how would you rate the attitude of the other school staff members towards the survey?	77	17	2	4

	Worked Well (%)	Needs Improvement (%)	Not Answered (%)
Overall, do you feel the School Coordinator Manual worked well or does it need improvement?	79	3	18

6.3 Survey Activities Questionnaire

The Survey Activities Questionnaire was designed to elicit information about NRCs experiences in preparing for and conducting the TIMSS Advanced 2008 data collection, with a focus on identifying and selecting samples, translating test instruments, assembling and printing test materials, packing and shipping the test materials, scoring constructed-response items, entering and verifying data, and implementing the national quality assurance program.

This section reports information gathered from the Survey Activities Questionnaire, reflecting the quality of the TIMSS Advanced 2008 survey materials and procedures in the participating countries. All participating countries, except one, completed the Survey Activities Questionnaire.

6.3.1 Sampling

The first part of the Survey Activities Questionnaire asked about sampling schools and classes within the sampled schools. None of the participating countries reported problems in selecting their samples using the manuals provided by the TIMSS & PIRLS International Study Center. Two countries did not use the Windows Within-school Sampling Software (WinW3S) provided by the IEA DPC to select classes. One country chose to use their own software, because they felt their experience using this software would make the process more efficient. The other country did not have to sample classes because the survey was administered to all classes within the sampled schools.

Two NRCs encountered organizational constraints in their systems that necessitated a deviation from the sample design. In each case, the Statistics Canada sampling expert was consulted to ensure that the altered design remained compatible with TIMSS Advanced 2008 standards.

6.3.2 Preparing the Survey Instruments

In translating the survey instruments, NRCs reported mostly using a combination of their own staff and outside experts. All NRCs reported that they had gone through the process of external translation verification (organized by the IEA Secretariat) of the achievement materials and background questionnaires.

The NRCs were asked to answer some questions about assembling and printing the survey instruments, as well as issues related to checking the materials and securely storing them. All NRCs answered that they were able to assemble the achievement booklets according to the instructions provided, and only one country did not go through the process of an external review of instrument layout by the TIMSS & PIRLS International Study Center. Nearly all countries conducted the recommended quality control checks during the printing process, with missing pages being the most common in detected errors. NRCs were able to fix all of the systematic errors before sending the tests for administration.

All countries reported that they followed procedures to protect the security of the tests during assembly and printing. None of the NRCs were concerned that there might have been a breach of security.

Some questions in the questionnaire addressed the extent to which NRCs detected errors in the testing materials as they were packed for shipping to School Coordinators. A few errors were found in the materials. All errors that were discovered before distribution were remedied. In cases where errors were found after distribution, they were mainly minor and remedied by School Coordinators. In more severe cases, the provided replacement materials were used.

6.3.5 Scoring Constructed-response Items

The Survey Activities Questionnaire collected information from NRCs about preparation for scoring and scoring the constructed-response items. The scoring process was an ambitious effort, requiring recruiting and training scoring staff to score student responses including independent double scoring to verify scoring reliability. All NRCs reported that they understood the procedures of within-country reliability scoring, and only one country had difficulty understanding procedures of cross-country reliability scoring, as explained in the manuals provided by the TIMSS & PIRLS International Study Center.

Most countries used their own staff, as well as teachers and university students to score the constructed-response items. All countries used the TIMSS Advanced 2008 scoring training materials provided by the TIMSS & PIRLS International Study Center to train their scorers. Countries used anywhere between 6 and 31 scorers, and all scoring activities took up to 30 days to complete.

6.3.6 Data Entry and Verification

Within the section on entering and submitting the TIMSS Advanced 2008 data of the Survey Activities Questionnaire, NRCs reported that they mainly used their own staff or a combination of their own staff and an external data entry company. One country used university students to enter its data.

All countries, except one where the survey instruments were scanned, used the Windows Data Entry Manager Software (IEA, 2008) to enter the TIMSS Advanced 2008 data. Most countries entered the data from a percentage of all the survey instruments twice as a verification procedure. The estimated proportion of survey instruments to be entered twice ranged from 5 to 30 percent. All NRCs reported establishing a secure storage area for the returned tests after data entry.

6.3.7 National Quality Assurance Program

As part of the national quality assurance activities, NRCs were required to send National Quality Control Observers to 10 percent of the participating schools in order to observe the test administration and document compliance with prescribed procedures. The last section of the Survey Activities Questionnaire addressed preparation for and implementation of the national quality assurance program.

In carrying out the national quality assurance program, only one national center did not use the *National Quality Control Monitor Manual* provided by the TIMSS & PIRLS International Study Center as a guide. The on-site quality control observations were conducted either by an external agency, members of the national center, a combination of the two, or, in some cases, other professionals, such as inspectors, retired teachers, mathematics and science supervisors, or ministry representatives.

References

IEA (2008). *Windows data entry manager software (WinDEM)* [Computer software and manual]. Hamburg: IEA Data Processing and Research Center.

TIMSS & PIRLS International Study Center. (2007a). *TIMSS Advanced 2008 international quality control monitor manual*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

TIMSS & PIRLS International Study Center. (2007b). *TIMSS Advanced 2008 national quality control monitor manual*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

TIMSS & PIRLS International Study Center. (2007c). *TIMSS Advanced 2008 school coordinator manual*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

TIMSS & PIRLS International Study Center. (2007d). *TIMSS Advanced 2008 survey operations procedures unit 2: Contacting schools and sampling classes for the TIMSS Advanced 2008 assessment*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

TIMSS & PIRLS International Study Center. (2007e). *TIMSS Advanced 2008 survey operations procedures unit 3: Preparing materials for the TIMSS Advanced 2008 data collection and administering the assessment*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

TIMSS & PIRLS International Study Center. (2007f). *TIMSS Advanced 2008 test administrator manual*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.


OCCASVS

Chapter 7

Creating and Checking the TIMSS Advanced 2008 Database

Milena Taneva

7.1 Introduction

Creating the TIMSS Advanced 2008 database and ensuring its integrity was a complex endeavor requiring close coordination and cooperation among the staff at the IEA Data Processing and Research Center (IEA DPC), the TIMSS & PIRLS International Study Center at Boston College, Statistics Canada, and the national research centers of the 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 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 checking and database creation procedures implemented by the IEA DPC in collaboration with the TIMSS &

PIRLS International Study Center and Statistics Canada, and the steps taken to confirm the integrity of the international database.

7.2 Overview of International Database Construction

On receipt of the data files from each country, the IEA DPC assumed responsibility for checking them for completeness, for applying standard cleaning rules to verify the accuracy and consistency of the data, and for documenting any deviation from the international file structure. All queries were communicated to the national centers and modifications were made to the data files as necessary. After all modifications had been applied, the data were processed and checked again. This process of editing the data, checking the edit reports, and implementing corrections was repeated as many times as necessary until all data were consistent and comparable nationally and internationally.

In preparation for creating the international database, the TIMSS & PIRLS International Study Center provided countries with data almanacs containing international statistics summarizing the responses to the questions in the background questionnaires as well as national statistics summarizing all achievement items so that National Research Coordinators (NRCs) could examine their data in an international perspective. This was one of the most important checks toward international comparability of the data. While in a national context a particular response may seem plausible, it may raise issues when viewed in an international context. All such issues were addressed, and the corresponding variables were either recoded or removed from the international database, as appropriate. Once the achievement data had been verified and converted to the international file format, they were sent to the TIMSS & PIRLS International Study Center where basic item statistics were produced and reviewed. At the same time, the IEA DPC 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 NRCs from tracking forms and the IEA's sampling software, was used by Statistics Canada to calculate sampling weights, population coverage and exclusion statistics, and participation rates.[1]

When the item review was completed and the computation of sampling weights was finalized and verified, the TIMSS & PIRLS International Study Center proceeded to scale the TIMSS Advanced 2008 achievement data and produce proficiency scores in advanced mathematics and physics for all participating students.[2] Once the proficiency scores had been verified at the TIMSS & PIRLS International Study Center, they were sent to the IEA DPC for inclusion in the international database.

7.3 Software Support for National Centers

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

The *WinW3S* software (IEA, 2007) performed all within-school sampling operations adhering strictly to the sampling rules established by Statistics Canada and the TIMSS & PIRLS International Study Center. The software also created all necessary tracking forms and stored student and teacher tracking information—such as students' age, gender, and participation status.

The WinDEM program (IEA, 2008) enabled key-entry of all TIMSS Advanced 2008 achievement and questionnaire data in a

1 Sampling weights and participation rates are described in Chapter 4.

2 The item review process and the scaling of the TIMSS Advanced 2008 achievement data are described in Chapter 8.

standard, internationally defined format. The software also included a range of checks for data verification within and across the various data files. These checks were applied by national center staff before sending the files to the IEA DPC.

7.4 Checking the Data at the IEA DPC

All participating countries were responsible for entering their TIMSS Advanced 2008 data into the appropriate data files and submitting them to the IEA DPC, where they underwent an exhaustive process of checking and editing—a process known as data cleaning. To facilitate the data cleaning process, countries were requested to provide the IEA DPC with detailed documentation of their data in addition to the data files themselves. This data documentation included copies of all original survey tracking forms, copies of the national version of the assessment booklets and questionnaires, and the completed Survey Activities Questionnaire. To ensure that all national adaptations to the survey instruments were fully documented, countries also were required to submit National Adaptations Forms, which became a written record of all national adaptations.

Countries also were asked to send the IEA DPC all test booklets selected for double-scoring the constructed-response items (some 500 advanced mathematics booklets and 500 physics booklets per country). The students' responses to constructed-response items in these booklets were digitally scanned and preserved for use in the next cycle of TIMSS Advanced, when they will be rescored by the national scoring staff to monitor consistency in scoring practices across survey cycles.

7.4.1 Quality Control in Data Cleaning

TIMSS Advanced is a complex study with demanding standards for data quality. This required an extensive set of interrelated data checking and cleaning procedures. To ensure that all procedures were conducted in the correct sequence, that no special requirements were overlooked, and that the cleaning process was implemented independently of the persons in charge, the following steps were undertaken:

- Before being applied to real data, all data-cleaning programs were thoroughly tested using simulated data sets containing all imaginable problems and inconsistencies.
- All incoming data and documents from countries were registered in a database. The date of arrival was recorded, along with any specific issues that merited attention.
- The data cleaning was organized following strict rules. Deviations from the cleaning sequence were not allowed, and the scope for involuntary changes to the cleaning procedures was minimal.
- All corrections applied to a country's data files were documented in a country-specific cleaning report.
- Occasionally, it was necessary to make manual changes to a country's data files. All manual corrections were documented in a program which recorded all changes and allowed IEA DPC staff to undo changes, or to redo the whole manual cleaning process automatically at a later stage of the data cleaning.
- Once data-cleaning was completed for a country, all cleaning steps were repeated from the beginning with the cleaned data files to detect any problems that might have been inadvertently introduced during the cleaning process.
- IEA DPC staff worked closely with the national research centers and, at different steps of the cleaning process, countries were provided with the processed data files with accompanying documentation and statistics to review and correct any inconsistencies detected.

- 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 were recorded in a national adaptations database for the *TIMSS Advanced 2008 User Guide for the International Database* (Foy & Arora, 2009). All national deviations required recoding of the data to comply with the international data structure. However, if international comparability could not be assured, the problematic data were removed from the international database.

7.4.2 Preparing National Data Files

The main objective of the data cleaning process was to ensure that the data adhered to international formats; that school, teacher, and student information could be linked across different survey files; and that the data accurately and consistently reflected the information collected within each country. The program-based data cleaning consisted of the following steps:

- Structure check and documentation
- Identification variable cleaning
- Linkage check
- Resolving inconsistencies in background questionnaire data

These data cleaning steps are illustrated in Exhibit 7.1 and are described in the subsequent sections. As shown in the exhibit, national research centers submitted their data and documentation to the IEA DPC. These data were cleaned by the DPC, which then produced the international database and provided documentation of the cleaning process to the national centers, including reports and summary statistics. Throughout the data cleaning process, effective communication between the IEA DPC and the national centers was key to ensuring quality data in the end.

Exhibit 7.1 Overview of Data Processing at the DPC

7.4.3 Structure Check and Documentation

For every country, data cleaning began with an exploration of its data file structure and a review of its documentation: National Adaptations Forms, Student Tracking Forms, Student-Teacher Linkage Forms, Teacher Tracking Forms, and Test Administration Forms (see Chapter 5 for an overview of the TIMSS Advanced data collection forms). Most countries sent all required documentation along with their data, which greatly facilitated the data checking.

At the beginning of the cleaning process, tracking and sampling information captured in the *WinW3S* database was combined with the actual data files that contained the survey data.

The first checks implemented at the IEA DPC looked for differences between the international file structure and the national file structures. Some countries made adaptations to their background questionnaires (such as adding national variables, or omitting or modifying international variables). The extent and nature of these changes differed across countries. Some countries administered the questionnaires without any changes (apart from translation), while other countries inserted items or options within existing international variables or added entirely new national questions. To keep track of all adaptations, NRCs were asked to complete National Adaptations Forms that became a written record of all implemented national adaptations. Where necessary, the IEA DPC modified the structure of a country's data to ensure that the resulting data remained comparable with those of other countries.

Once all data files matched the international standard as specified in the international codebooks,[3] a series of standard cleaning rules were applied to the data. This was conducted using software developed at the IEA DPC that could identify and, in many cases, automatically correct inconsistencies in the data. Each problem was recorded in a database, identified by a unique problem number and with a description of the problem and the action taken by the program or by the IEA DPC staff.

When problems could not be resolved automatically, they were reported to the national centers so that original data collection instruments and tracking forms could be checked to trace the source of the errors. Whenever possible, staff at the IEA DPC suggested solutions and asked the national centers to either accept them or propose alternatives. Data files then were updated to reflect the solutions agreed upon. When the national centers could not solve problems, they were corrected by applying generalized cleaning rules.

As part of this standardization process, since a direct correspondence between the data-collection instruments and the data files was no longer necessary, the file structure was rearranged from a booklet-oriented model designed to facilitate data entry to an item-

3 The codebooks defined the structure of the data files and are described in Chapter 5.

oriented layout more suited to data analysis. Variables created purely for verification purposes during data entry were dropped at this time and provision was made for additional variables necessary for analysis and reporting (i.e., reporting variables, derived variables, sampling weights, and achievement scores).

7.4.4 Identification Variable Cleaning

Each record in a data file should have a unique identification (ID) number. The existence of records with duplicate ID numbers in a file implies an error of some kind. When two records shared the same ID number and contained exactly the same data, one of the records was deleted and the other remained in the database. When records contained different data, apart from the ID numbers, and it was not possible to identify which record contained the "true data," both records were removed from the database. The IEA DPC tried to keep such losses to a minimum, and only in rare cases were data actually deleted.

The ID cleaning focused on the student background questionnaire files, which contained most of the critical identification variables. Apart from the unique student ID numbers, variables pertaining to the students' participation and exclusion status as well as the dates of birth and dates of testing used to calculate students' age at the time of testing were important to check. The Student Tracking Forms were essential in resolving all anomalies, as was close cooperation with the national centers (since, in most cases, the Student Tracking Forms were completed in each country's national language). Once ID cleaning was completed, the *WinW3S* databases with information about student participation or exclusion were sent to Statistics Canada where they were used to calculate sampling weights, exclusion rates, and participation rates.

7.4.5 Linkage Check

In TIMSS Advanced 2008, data about students and their schools and teachers were located in several different files, so that it was crucial that the records from these files linked together correctly to provide meaningful data for analysis and reporting. The linkage across files was implemented through a hierarchical ID numbering system that incorporated a school, class, and student component[4] and was cross-checked against the tracking forms. It was required that student records in the achievement files and student background files be matched correctly, that student entries in the reliability files be properly matched to student entries in the achievement files, that teachers be linked to the correct students, and that schools be linked to the correct teachers and students.

7.4.6 Resolving Inconsistencies in Background Questionnaire Data

The number of inconsistent or implausible responses in background files varied from country to country, but no country's data were completely free of inconsistent responses. Treatment of inconsistent responses was determined on a question-by-question basis, using available documentation to make informed decisions. All background questionnaire data were checked for consistency among the responses given. For example, question 1a in the school questionnaire asked for total school enrollment in all grades, while 1b asked for enrollment in the target grade only. Clearly, the response given to 1b should not exceed the response given to 1a. All such inconsistencies were flagged, and the national centers were asked to investigate. In cases that could not be resolved, or where the data made no sense, responses were recoded as "Omitted".

Filter questions, which appear in some questionnaires, were used to direct respondents to a particular section of the questionnaire. Filter

4 The Hierarchical ID numbering system consisted of a school ID, a class ID that included the school ID, and a student ID that included the class ID, such that student 01220523 could be identified as student 23 in class 05 of school 0122.

questions and the dependent questions that followed were subject to specific cleaning rules. If the answer to a filter question was "No", or simply not given, and yet the dependent questions were answered, then the filter question was recoded as "Yes".

Split variable checks were applied to questions where the answers were coded into several variables. For example, question 5 in the student questionnaire listed a number of home possessions and asked the students to check all that applied. Student responses were captured in a series of nine variables, each one coded as "Yes" if the "Yes" box corresponding to that possession was checked and "No" if the "No" box was checked. Occasionally, students checked the "Yes" boxes for some possessions, but left the "No" boxes unchecked for others. Since in these cases it was clear that the unchecked boxes actually meant "No," they were recoded accordingly.

7.4.7 National Cleaning Documentation

NRCs received detailed reports of all problems identified in their data and of any steps applied to correct them, which also included records of all deviations from the international version of the data collection instruments and the international file structure. Additionally, the IEA DPC provided each NRC with revised data files incorporating all agreed-upon 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 national standardized scores of advanced mathematics and physics that could be used in national analyses to be conducted before the official release of the international database.

7.4.8 Handling Missing Data

When the TIMSS Advanced 2008 data were entered with WinDEM, two types of entries were possible: valid data values and missing data values. Missing data could be assigned a value of omitted or not administered 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, four missing codes are used:

- Not administered: The respondent was not administered the actual item and consequently he or she had no opportunity to provide an answer to the question.
- Omitted: The respondent had a chance to answer the question, but chose not to. This code also was used for responses that were not interpretable in either the background files or the achievement files.
- Logically not applicable (applied to filter-dependent questions): The respondent answered a filter question in a way that made the following dependent questions not applicable to him or her.
- Not reached (used only in the achievement files): An item was considered not reached when the item itself and the one immediately preceding it were not answered, and there were no other items completed in the remainder of the assessment booklet.

7.5 Data Products

Data products sent to NRCs by the IEA DPC and the TIMSS & PIRLS International Study Center included both data almanacs and data files.

7.5.1 Data Almanacs and Item Statistics

Every country received a set of data almanacs produced by the TIMSS & PIRLS International Study Center. They contained weighted national summary statistics for each variable included in the survey instruments and were sent to the participating countries for review. When necessary, they were accompanied by specific questions about the data presented in them. Countries also were provided item almanacs with national summary statistics for the achievement items, which were reviewed by the national centers as part of the data validation process.

7.5.2 Versions of the National Data Files

Building the TIMSS Advanced 2008 international database was an iterative process. The IEA DPC provided NRCs with an updated version of their country's data files whenever a major step in data processing was completed. This guaranteed that the NRCs had a chance to review their data and run their own checks to validate the contents of the data files. Prior to the release of the TIMSS Advanced 2008 international database, the participating countries received several versions of their national data files. Each country received only its own data. The first version was sent as soon as the data could be regarded as "clean" with respect to identification variables and any possible linkage issues. These first files contained interim national standardized achievement scores calculated by the IEA DPC using a Rasch-based scaling method. Documentation, with a list of the cleaning checks and corrections made to the data, also was sent to enable NRCs to review the cleaning process. Another version of the data files was sent to the countries when the sampling weights and the international achievement scores were available. This was done after all exhibits of the TIMSS Advanced 2008 international report had been verified and final updates to the data files implemented. This allowed NRCs to replicate the results presented in the international report.

7.5.3 The TIMSS Advanced 2008 International Database

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

- Data recorded for each variable were internationally comparable.

- National adaptations were reflected appropriately in all variables.
- Questions that were not internationally comparable were removed from the database.
- All entries in the database could be linked to the appropriate respondents—students, teachers, and school principals.
- Sampling weights and student achievement scores were available for international comparisons.

In a joint effort involving the IEA DPC and the TIMSS & PIRLS International Study Center, a national adaptations database, describing all adaptations made by individual countries to questionnaires and how they were handled, was constructed. All information contained in this database is provided in Supplement 2 of the *TIMSS Advanced 2008 User Guide*.

References

Foy, P., & Arora, A. (2009). *TIMSS Advanced 2008 user guide for the international database.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

IEA (2007). *Windows within-school sampling software (WinW3S)* [Computer software and manual]. Hamburg: IEA Data Processing and Research Center.

IEA (2008). *Windows data entry manager software (WinDEM)* [Computer software and manual]. Hamburg: IEA Data Processing and Research Center.


OCCASVS
CREPVSCVLVM

Chapter 8

Scaling the Data from the TIMSS Advanced 2008 Assessments

Pierre Foy, Joseph Galia, and Isaac Li

8.1 Overview

The TIMSS Advanced 2008 goals of broad coverage of the advanced mathematics and physics curricula and of measuring trends across assessments necessitated the adoption of a complex matrix-sampling booklet design,[1] with individual students responding to a subset of the advanced mathematics and physics 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 Advanced relied on Item Response Theory (IRT) scaling to describe student achievement on the assessment and to provide accurate measures of trends from the previous assessments. The TIMSS IRT scaling approach used multiple imputation—or "plausible values"—methodology to obtain proficiency scores in advanced mathematics and physics 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

1 The TIMSS Advanced 2008 assessment design is described in the *TIMSS Advanced 2008 Assessment Frameworks* (Garden, et al., 2006).

advanced scaling combined student responses to the items they were administered with information about students' backgrounds, a process known as "conditioning."

This chapter describes the steps that produced scaled scores of student achievement in advanced mathematics and physics. First, it explains the process of reviewing item statistics to validate the statistical properties of the achievement items used in the TIMSS Advanced 2008 assessments. It then provides a general explanation of the methodology used to scale the TIMSS Advanced 2008 data, and describes how this approach was applied to the 2008 assessment data and to the data from the previous TIMSS Advanced 1995 study in order to measure trends in achievement. The TIMSS Advanced scaling was conducted by the TIMSS & PIRLS International Study Center at Boston College, using software from Educational Testing Service.[2] This chapter also provides a description of the scale anchoring methodology used to describe student performance at various points on the TIMSS Advanced mathematics and physics achievement scales, and the methodology used to estimate standard errors of the estimates published for TIMSS Advanced 2008.

8.2 Item Review

For TIMSS Advanced 2008, as in the TIMSS assessments at the fourth and eighth grades, 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 ten countries that participated in TIMSS Advanced 2008. This review of item statistics was conducted before applying item response theory (IRT) scaling to the TIMSS Advanced 2008 achievement data to derive student achievement scores in advanced mathematics and physics for analysis and reporting. The review of item statistics played

2 TIMSS is indebted to Matthias Von Davier, Ed Kulick, and John Barone of Educational Testing Service for their advice and support.

a crucial role in the quality assurance of the TIMSS Advanced 2008 data, making it possible to detect 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 guide 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.

8.2.1 Statistics for Item Analysis

To begin the review process, the TIMSS & PIRLS International Study Center computed item statistics for all 143 advanced mathematics and physics achievement items that were administered in the TIMSS Advanced 2008 assessments. The properties of the items in each of the ten countries that participated were then carefully reviewed. Exhibits 8.1 and 8.2 show actual samples of the statistics calculated for a multiple-choice and a constructed-response item, respectively.

For all items, regardless of 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).[3] Also provided was an estimate of the difficulty of the item 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 for each item were displayed alphabetically by country,

3 For computing point-biserial correlations, the total score was the percentage of advanced mathematics or physics items a student answered correctly.

Exhibit 8.1 International Item Statistics for a Multiple-choice Item

Trends in International Mathematics and Science Study - TIMSS 2008 Assessment Results
International Item Statistics (Unweighted) - Review Version - 12th Grade
For Internal Review Only: DO NOT CITE OR CIRCULATE

Mathematics: Calculus / Reasoning (MA23206 - M4_06)
Label: Sign of derivative function
Type: MC Key: D

				Percentages							Point Biserials							
Country	N	Diff	Disc	Pct_A	Pct_B	Pct_C	Pct_D	Pct_E	Pct_OM	Pct_NR	PB_A	PB_B	PB_C	PB_D	PB_E	PB_OM	RDIFF	Flags
Armenia	439	41.5	0.45	32.8	9.1	8.4	41.5	.	8.2	0.2	-0.14	-0.22	-0.20	0.45	.	-0.13	-0.77	____F_
Iran, Islamic Rep. of	1195	59.7	0.54	13.6	5.4	9.5	59.7	.	11.7	0.1	-0.28	-0.10	-0.22	0.54	.	-0.26	-0.78	____F_
Italy	1080	44.3	0.47	44.4	2.2	4.1	44.3	.	5.1	0.1	-0.39	-0.09	-0.06	0.47	.	-0.06	-0.70	____F_
Lebanon	801	80.9	0.37	15.7	0.6	2.1	80.9	.	0.6	0.0	-0.32	-0.08	-0.10	0.36	.	-0.07	-1.36	__E_F_
Netherlands	763	87.5	0.25	5.4	1.3	5.6	87.5	.	0.1	0.3	-0.17	-0.05	-0.16	0.25	.	-0.03	-1.94	__E_F_
Norway	966	52.3	0.38	28.8	8.6	8.1	52.3	.	2.3	0.0	-0.33	-0.03	-0.09	0.38	.	-0.08	-1.15	____F_
Philippines	2057	17.6	0.23	61.1	9.7	10.3	17.6	.	1.3	0.0	-0.16	-0.07	0.05	0.23	.	-0.02	0.15	_CH_F_
Russian Federation	1588	68.1	0.54	24.5	0.9	5.5	68.1	.	1.0	0.0	-0.42	-0.08	-0.22	0.54	.	-0.11	-0.64	__H_F_
Slovenia	1090	38.3	0.46	41.2	4.4	12.3	38.3	.	3.9	0.0	-0.36	-0.04	-0.10	0.46	.	-0.04	-0.26	__H_F_
Sweden	1141	67.3	0.48	17.0	4.8	9.6	67.3	.	1.3	0.3	-0.35	-0.11	-0.19	0.48	.	-0.11	-2.01	__E_F_
International Avg.	1112	55.8	0.42	28.4	4.7	7.5	55.8	.	3.6	0.1	-0.29	-0.09	-0.13	0.42	.	-0.09	-0.95	____F_

Keys: Diff: Percent correct score; Disc: Item discrimination; Pct_A...E: Percent choosing option; Pct_OM, NR: Percent Omitted and Not Reached;
PB_A...E: Point Biserial for option; PB_OM: Point Biserial for Omittted. RDIFF= Rasch difficulty.

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 8.2 International Item Statistics for a Constructed-response Item

Trends in International Mathematics and Science Study - TIMSS 2008 Assessment Results
International Item Statistics (Unweighted) - Review Version - 12th Grade
For Internal Review Only: DO NOT CITE OR CIRCULATE

Mathematics: Calculus / Reasoning (MA23157 - M5_06)
Label: Find point of maxima and inflection
Type: CR Key: X

				Percentages					Point Biserials					Reliability			
Country	N	Diff	Disc	Pct_0	Pct_1	Pct_2	Pct_OM	Pct_NR	PB_0	PB_1	PB_2	PB_OM	RDIFF	Cases	Score	Code	Flags
Armenia	407	11.5	0.57	24.1	13.3	4.9	57.7	4.9	-0.14	0.43	0.36	-0.32	1.02	91	100.0	100.0	____F_
Iran, Islamic Rep. of	1132	33.4	0.68	27.0	23.8	21.6	27.7	6.8	-0.25	0.13	0.61	-0.41	0.52	294	100.0	98.3	______
Italy	1032	16.2	0.55	27.8	17.6	7.4	47.2	3.4	-0.12	0.20	0.49	-0.29	0.83	139	100.0	100.0	____F_
Lebanon	784	39.5	0.47	26.4	50.5	14.3	8.8	2.1	-0.32	0.12	0.38	-0.18	0.82	188	99.5	99.5	______
Netherlands	757	43.8	0.34	22.3	58.0	14.8	4.9	1.2	-0.19	0.04	0.28	-0.18	0.60	198	94.9	93.9	______
Norway	936	28.6	0.39	28.5	45.0	6.1	20.4	2.6	-0.21	0.10	0.39	-0.17	0.41	235	100.0	99.1	__E_F_
Philippines	2033	16.7	0.38	50.2	31.0	1.2	17.6	0.2	-0.21	0.24	0.36	-0.12	0.89	507	95.1	93.9	__H_F_
Russian Federation	1582	57.2	0.47	12.2	41.2	36.6	10.1	0.9	-0.19	-0.10	0.42	-0.30	0.02	402	98.8	98.8	__E___
Slovenia	1058	20.1	0.43	44.6	31.8	4.3	19.4	1.9	-0.19	0.26	0.32	-0.21	1.04	283	100.0	99.3	__H_F_
Sweden	1112	20.7	0.54	46.8	21.8	9.8	21.7	4.1	-0.31	0.21	0.46	-0.24	0.36	289	96.9	94.5	__E_F_
International Avg.	1083	28.8	0.48	31.0	33.4	12.1	23.5	2.8	-0.21	0.16	0.41	-0.24	0.65	.	98.5	97.7	______

Keys: Diff: Percent correct score; Disc: Item discrimination; Pct_0...3: Percent obtaining score level; Pct_OM, NR: Percent Omitted and Not Reached;
PB_0...3: Point Biserial for score level; PB_OM: Point Biserial for Omittted; RDIFF= Rasch difficulty;
Reliability: (Cases) Responses double scored; (Score) Percent agreement on score; (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.

together with the international average for each statistic in the last row of the exhibits.

Statistics displayed for multiple-choice items included the percentage of students that chose each response option, as well as the percentage of students that omitted or did not reach the item, and the point-biserial correlations for each response option. Statistics displayed for constructed-response items (which could have 1 or 2 score points) included the difficulty and discrimination of each score level. Constructed-response item displays also provided information about the reliability with which each item was scored in each country, showing the total number of double-scored cases, the percentage of code agreement between the scorers, and the percentage of score agreement between scorers.

The definitions and detailed descriptions of the statistics that were calculated are given below for the examples shown in Exhibits 8.1 and 8.2. The statistics were calculated separately for advanced mathematics and physics. The statistics are listed in order of their appearance in the item analysis outputs:

N: 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 item analysis.[4]

Diff: The item difficulty is the average percent correct. For a 1-point item, including multiple-choice items, it is the percentage of students providing a fully correct response to the item. For 2-point items, it is the average percentage of points; i.e., if all students scored 1 point on a 2-point item, then the average percent correct for such an item would be 50 percent. For this statistic, not-reached items were treated as not administered.

4 For item review and scaling, items not reached by a student were treated as if they had not been administered. For estimating student proficiency, however, not reached items were treated as incorrect.

Disc: The item discrimination was computed as the correlation between a correct response to the item and the overall score on all of the advanced mathematics or physics items administered to a student.[5] Items exhibiting good measurement properties should have a moderately positive correlation, indicating that the more able students get the item right, the less able get it wrong.

PCT_A, PCT_B, PCT_C, PCT_D, and PCT_E: Used for multiple-choice items only (see Exhibit 8.1). 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.

PCT_0, PCT_1, and PCT_2: Used for constructed-response items only (see Exhibit 8.2). 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.

PCT_OM: Percentage of students who, having reached the item, did not provide a response. No reached items were excluded from the denominator.

PCT_NR: Percentage of students who 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 also omitted.

PB_A, PB_B, PB_C, PB_D, and PB_E: Used for multiple-choice items only. These columns show the point-biserial correlations between choosing each of the response options (A, B, C, D, or E) and the overall score on all of the advanced mathematics or physics items administered to a student. Items with good

5 For constructed-response items, the discrimination is the correlation between the number of score points and total score.

psychometric properties have moderately positive correlations for the correct option and negative correlations for the distractors (the incorrect options).

PB_0, PB_1, and PB_2: Used for constructed-response items only. These columns present the point-biserial correlations between the score levels on the item (0, 1, or 2) and the overall score on all of the mathematics or science items the student was administered. For items with good measurement properties, the correlation coefficients should increase from negative to positive as the score on the item increases.

PB_OM: The point-biserial correlation between a binary variable, indicating an omitted response to the item, and the overall score on all of the mathematics or physics items administered to a student. This correlation should be negative or near zero.

RDIFF: An estimate of the difficulty of an item based on a Rasch one-parameter IRT model applied the achievement data for a given country. The difficulty estimate is expressed in the logit metric (with a positive logit indicating a difficult item) and was scaled so that the average Rasch item difficulty across all items within each country was zero.

Reliability (Cases): To provide a measure of the reliability of the scoring of constructed-response items, items in approximately 25 percent of the test booklets in each country were independently scored by two scorers. This column indicates the number of responses that were double-scored for a given item in a country.

Reliability (Score): This column contains the percentage of agreement on the score value of the two-digit codes assigned by the two independent scorers.

Reliability (Code): This column contains the percentage of exact agreement on the two-digit codes assigned by the two independent scorers.

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

- Item discrimination was less than 0.15 (flag D).
- Item difficulty was less than 25 percent for four-option multiple-choice items, or less than 20 percent for five-option multiple choice items (flag C).
- Item difficulty exceeded 95 percent (flag V).
- The Rasch difficulty estimate for a given country made the item either easier (flag E) or harder (flag H) relative to the international average Rasch difficulty of that item.
- The point-biserial correlation for at least one distractor in a multiple choice item was positive, or the estimated mean abilities across the score levels of a constructed-response item were not ordered (flag A).
- The percentage of students selecting one of the response options for a multiple-choice item or of one of the score values for a constructed-response item was less than 10 percent (flag F).
- Scoring reliability for agreement on the score value of a constructed-response item was less than 80 percent (flag R).

Although not all of these conditions necessarily indicated a problem, the flags were a useful way to draw attention to potential sources of concern.

In order to measure trends, TIMSS Advanced 2008 included items from the 1995 assessments.[6] For these trend items, the review included an examination of changes in item statistics between the 1995 and 2008 administrations. An example is shown in Exhibit 8.3. The information in this exhibit is different from that presented in Exhibits 8.1 and 8.2, and includes countries' statistics from both the 1995 and 2008 assessments. In reviewing these item statistics, the aim was to detect any unusual changes in item properties between assessments that might indicate a problem in using the item to measure trends.

Exhibit 8.3 International Item Statistics for a Trend Item

Trends in International Mathematics and Science Study - TIMSS Advanced 2008 Assessment Results
Percent of Responses by Item Category (Mathematics) - Trend Items - Final Year of Secondary School

Mathematics: Geometry / Applying (MA13026A - M3_06A)
Label: Triangle abc/reflection
Type: CR Key: X

COUNTRY	Year	N	10	70	71	79	99	INVALID	OMIT	NOT REACHED	V1	1.GIRL % Right	2.BOY % Right
Italy	1995	126	65.8	4.3	5.1	11.8	13.1	1.5	5.0	6.6	65.8	65.8	65.7
	2008	1070	54.9	2.5	6.7	21.2	14.8	0.0	8.8	6.0	54.9	56.5	54.0
Russian Federation	1995	468	75.4	3.8	5.3	9.2	6.3	0.0	1.2	5.1	75.4	74.4	76.3
	2008	1588	75.0	2.1	6.2	10.7	5.9	0.0	4.4	1.6	75.0	74.5	75.5
Slovenia	1995	452	69.2	1.4	0.3	21.9	7.3	2.0	2.2	3.1	69.2	66.7	71.6
	2008	1083	69.7	7.6	2.7	16.9	3.1	0.0	0.9	2.2	69.7	68.7	71.2
Sweden	1995	244	39.5	10.0	1.6	25.4	23.5	0.4	13.6	9.5	39.5	25.4	47.1
	2008	1148	18.8	8.3	5.2	34.5	33.2	0.0	26.0	7.2	18.8	17.9	19.4
International Avg.	1995	323	62.4	4.8	3.1	17.1	12.6	1.0	5.5	6.1	62.4	58.1	65.2
	2008	1222	54.6	5.1	5.2	20.8	14.3	0.0	10.0	4.2	54.6	54.4	55.0

V1 = Percent scoring 1 or better V2 = Percent scoring 2 or better
Percent right for boys and girls corresponds to the percent obtaining the maximum score on the item.
Because of missing gender information, some totals may appear inconsistent.

6 Information on trend items is available in Chapter 2.

8.2.2 Item-by-Country Interaction

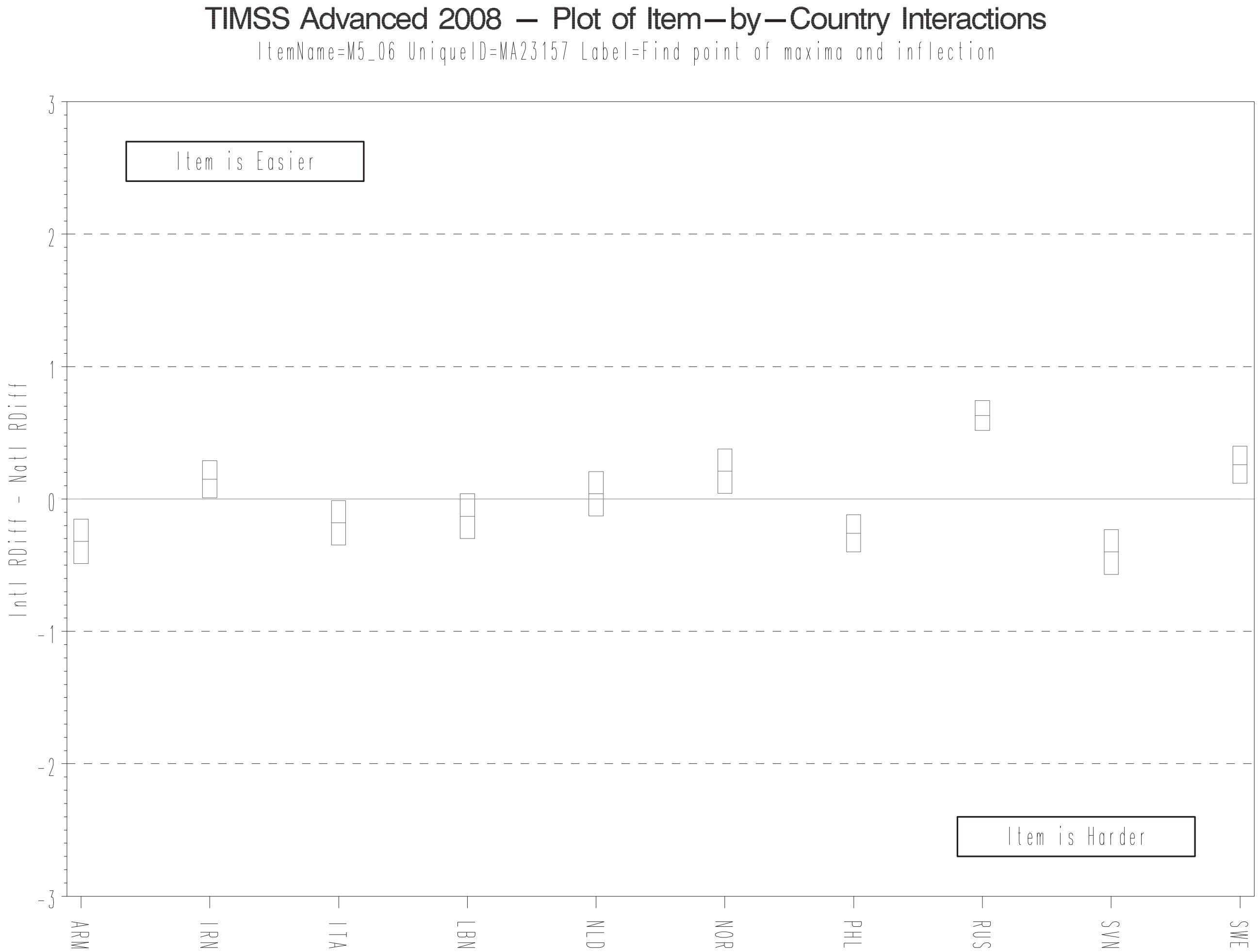
Although countries are expected to exhibit some variation in performance across items, in general countries with high average performance on the assessment should perform relatively well on all of the items, and low-scoring countries should do less well on all of items. When this does not occur (i.e., when a high-performing country has a low performance on an item on which other countries are doing well), there is said to be an item-by-country interaction. When large, such item-by-country interactions may be a sign that an item is flawed in some way, and steps should be taken to address the problem.

To assist in detecting sizeable item-by-country interactions, the TIMSS & PIRLS International Study Center produced a graphical display for each item showing the difference between each country's Rasch item difficulty and the average Rasch item difficulty across all countries. Exhibit 8.4 provides an example of a TIMSS Advanced 2008 item-by-country interaction display. The difference in Rasch item difficulty for each country is presented as a 95 percent confidence interval, which includes a built-in Bonferroni correction for multiple comparisons. The limits for this confidence interval were computed as follows:

$$\text{Upper Limit} = RDIFF_{i.} - RDIFF_{ik} + SE(RDIFF_{ik}) \cdot Z_b$$
$$\text{Lower Limit} = RDIFF_{i.} - RDIFF_{ik} - SE(RDIFF_{ik}) \cdot Z_b$$

Where $RDIFF_{ik}$ was the Rasch difficulty of item i in country k, $RDIFF_{i}.$ was the average difficulty of item i across all countries, SE($RDIFF_{ik}$) was the standard error of the Rasch difficulty of item i in country k , and Z_b was the critical value from the Z distribution corrected for multiple comparisons using the Bonferroni procedure.

Exhibit 8.4 Sample Plot of Item-by-Country Interaction for a TIMSS Advanced 2008 Item



8.2.3 Trend Item Analysis

Because an important part of the TIMSS Advanced 2008 assessment was measuring trends across the 1995 and 2008 assessment cycles, an additional review step ensured that the trend items had similar characteristics in both cycles (i.e., an item that was relatively easy in 1995 should have been relatively easy in 2008). The comparison between cycles was made in a number of ways. For each trend country, almanacs of item statistics displayed the percentage of students within each score category (or response option for multiple-choice items) for each cycle, as well as the difficulty of the item and the percent correct by gender. While some changes were anticipated as countries' overall achievement may have improved or declined, items were noted if the difference between the Rasch difficulties across the two cycles for a particular country was greater than 2 logits.

The TIMSS & PIRLS International Study Center used two different graphical displays to examine the differences item difficulties. The first of these, shown in Exhibit 8.5, displays the difference in Rasch item difficulties between 1995 and 2008. A positive difference indicated that an item was relatively easier in a country in 2008, and a negative difference indicated that an item was relatively more difficult. The second, Exhibit 8.6, shows the performance of a given country on all trend items simultaneously. For each country, the graph plotted the 1995 Rasch difficulty of every trend item in 1995 against its 2008 Rasch difficulty. Where there were no differences between the difficulties in 1995 and 2008, the data points aligned on or near the diagonal.

Exhibit 8.5 Sample Plot of Difference in Rasch Item Difficulties for a Trend Item

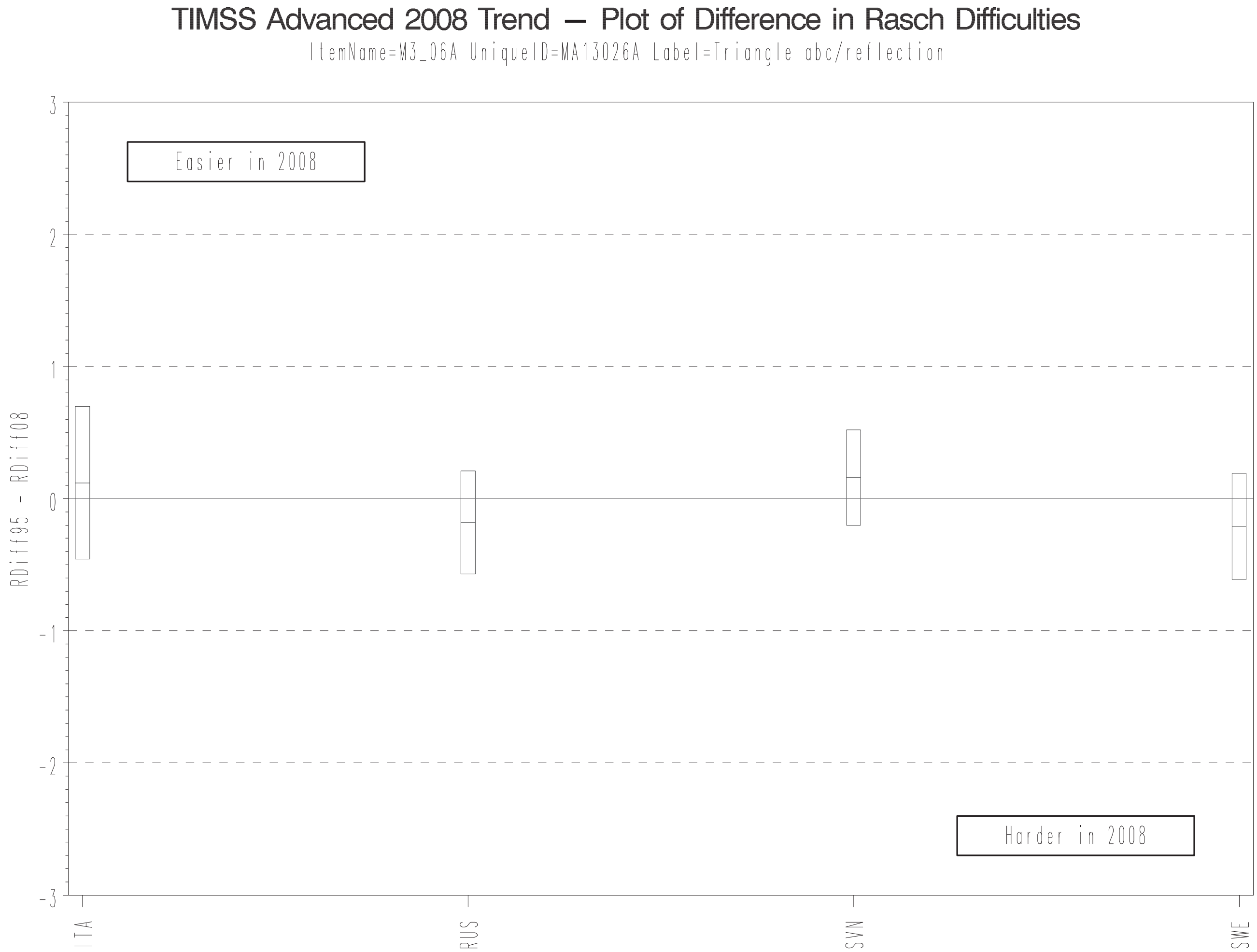
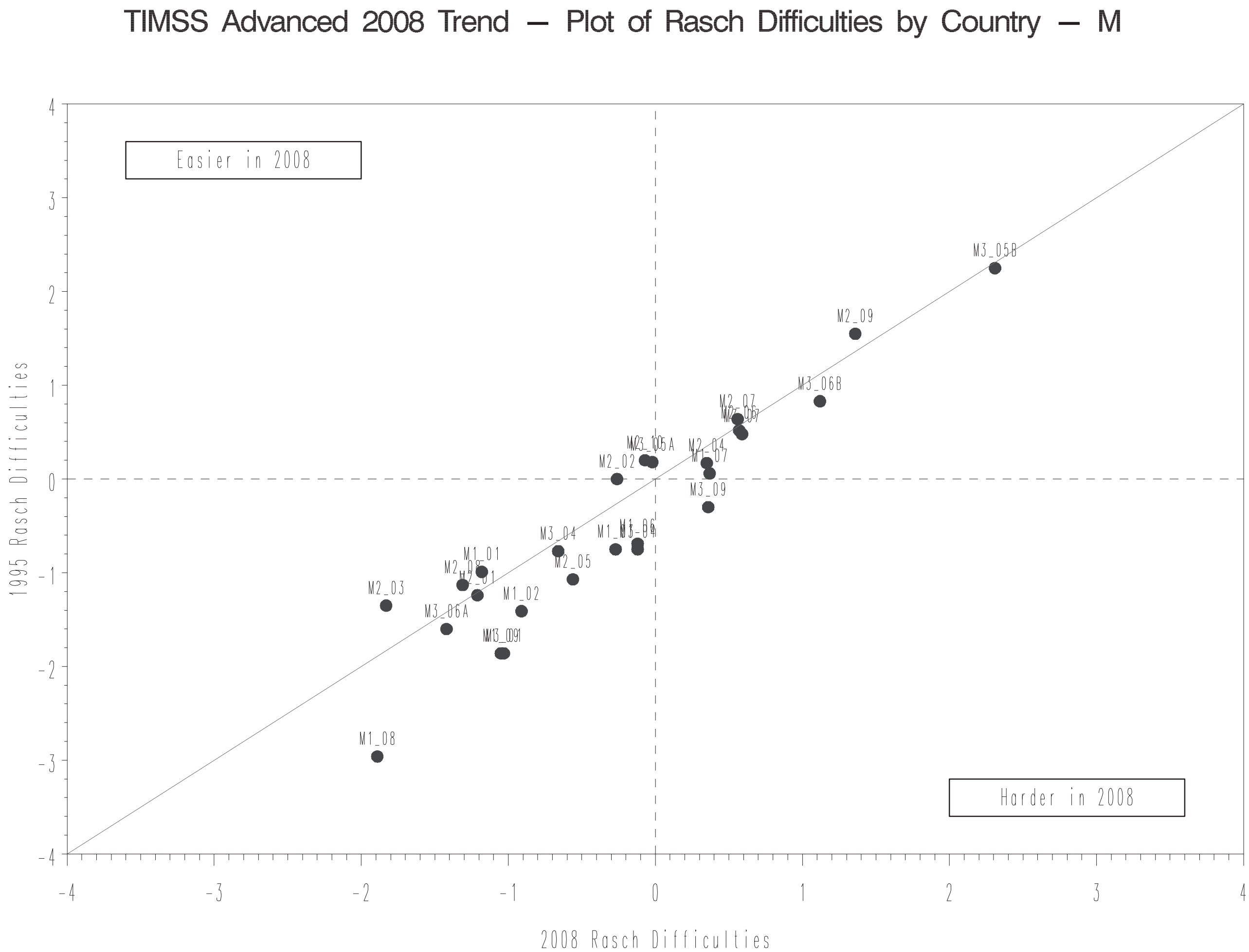


Exhibit 8.6 Sample Plot of Rasch Item Difficulties across Trend Items by Country



8.2.4 Reliability

Gauging the reliability of the TIMSS Advanced 2008 assessments was a critical quality control step in reviewing the items. There were two aspects of reliability under review. The set of items selected as part of the advanced mathematics and physics assessments needed to constitute a cohesive whole measuring their respective domains, a quality known as test reliability. Also, the scoring of the constructed-response items had to meet specific reliability criteria in terms of consistent within-country and cross-country scoring.

8.2.4.1 Test Reliability

Exhibit 8.7 displays the advanced mathematics and physics test reliability coefficients for every country. These coefficients are the median Cronbach's alpha reliability across the four test booklets of advanced mathematics and physics. In general, median reliabilities were relatively high in both subjects with the international median at 0.82 for advanced mathematics and 0.80 for physics. All median reliabilities were at least 0.70, except for physics in Lebanon, where the median reliability was 0.68.

Exhibit 8.7 Cronbach's Alpha Reliability Coefficients for TIMSS Advanced 2008

Country	Reliability Coefficient	
	Advanced Mathematics	Physics
Armenia	0.87	0.82
Iran, Islamic Rep. of	0.90	0.85
Italy	0.84	0.75
Lebanon	0.80	0.68
Netherlands	0.70	0.74
Norway	0.78	0.79
Philippines	0.79	—
Russian Federation	0.88	0.88
Slovenia	0.83	0.81
Sweden	0.80	0.80
International Median	0.82	0.80

8.2.4.2 Scoring Reliability for Constructed-response Items

About one third of the items in the TIMSS Advanced 2008 assessment were constructed-response items, comprising nearly half of the score points for the assessment.[7] 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 Advanced 2008, measures taken to ensure that the constructed-response items were scored reliably in all countries, and these measures included developing scoring guides for each constructed-response question (that provided descriptions of acceptable responses for each score point value)[8] as well as providing extensive training in the application of the scoring guides.

Within-Country Scoring Reliability

To gather and document information about the within-country agreement among scorers, a random sample of approximately 25 percent of the assessment booklets was selected to be scored independently by two scorers. The inter-scorer agreement for each item in each country was examined as part of the item review process. The average and range of the within-country percentage agreement across all items for both grades are presented in Exhibit 8.8 for both advanced mathematics and physics.

Scoring reliability was high on average across countries. The percent agreement on the correctness score across all countries was 98 percent in advanced mathematics and 97 percent in physics. All countries had an average percent agreement on the correctness score above 94 percent in advanced mathematics and above 91 percent in physics.

7 The development of the TIMSS Advanced 2008 assessment items is described in Chapter 2.

8 A discussion of the development of the scoring guides for constructed-response items is provided in Chapter 2.

Exhibit 8.8 Within-country Scoring Reliability for TIMSS Advanced 2008 Constructed-Response Items

Advanced Mathematics

Countries	Correctness Score Agreement			Diagnostic Score Agreement		
	Average of Percent Agreement Across Items	Range of Percent Agreement		Average of Percent Agreement Across Items	Range of Percent Agreement	
		Min	Max		Min	Max
Armenia	100	98	100	97	86	100
Iran, Islamic Rep. of	98	90	100	95	89	100
Italy	100	97	100	98	94	100
Lebanon	100	99	100	99	97	100
Netherlands	94	72	100	91	65	99
Norway	99	98	100	98	95	100
Philippines	98	93	100	95	85	100
Russian Federation	97	86	100	95	86	100
Slovenia	100	99	100	99	97	100
Sweden	97	88	100	93	83	99
International Average	98	92	100	96	88	100

Physics

Countries	Correctness Score Agreement			Diagnostic Score Agreement		
	Average of Percent Agreement Across Items	Range of Percent Agreement		Average of Percent Agreement Across Items	Range of Percent Agreement	
		Min	Max		Min	Max
Armenia	99	93	100	97	93	100
Iran, Islamic Rep. of	96	91	100	90	71	99
Italy	99	94	100	97	86	100
Lebanon	99	93	100	98	92	100
Netherlands	91	80	99	85	71	97
Norway	97	90	100	94	87	100
Russian Federation	96	89	100	93	83	99
Slovenia	100	98	100	99	95	100
Sweden	97	89	99	93	83	99
International Average	97	91	100	94	85	99

Cross-Country Scoring Reliability

Because of the different languages used by the countries participating in TIMSS Advanced 2008, establishing the reliability of constructed-response scoring across all countries was not feasible. However, TIMSS Advanced 2008 did conduct a cross-country study of scoring using English as a common language. A sample of student responses from a pilot study carried out in English was provided to countries. It included 100 student responses to each of nine advanced mathematics items and nine physics items. This set of 1,800 student responses in English was then scored independently in each country that had two scorers proficient in English. In all, 14 scorers from 7 countries participated in the study. Scoring for this study took place shortly after the other scoring reliability activities were completed. Making all possible pair-wise comparisons among scorers gave 91 comparisons for each student response to each item. This resulted in 9,100 total comparisons when aggregated across all 100 student responses to an item. Agreement across countries was defined in terms of the percentage of these comparisons that were in agreement.

Exhibit 8.9 shows that scorer reliability across countries was high for advanced mathematics, with the percent agreement averaging 94 percent across the nine items for the correctness score and 90 percent for the diagnostic score. For physics, the percent agreement averaged 88 percent across the nine items for the correctness score and 80 percent for the diagnostic score.

Exhibit 8.9 Cross-Country Scoring Reliability for TIMSS Advanced 2008

Advanced Mathematics

Item Label	Total Valid Comparisons	Percent Agreement	
		Correctness Score Agreement	Diagnostic Score Agreement
M4_04 - MA23201	9100	88	88
M4_08 - MA23043	9100	95	90
M5_03 - MA23054	9100	94	92
M5_05 - MA23131A	9100	99	99
M5_05 - MA23131B	9100	98	97
M5_10 - MA23094	9100	89	81
M6_07 - MA23198	9100	94	92
M7_03 - MA23141	9100	98	83
M7_11 - MA23170	9100	90	84
Average Percent Agreement		94	90

Physics

Item Label	Total Valid Comparisons	Percent Agreement	
		Correctness Score Agreement	Diagnostic Score Agreement
P4_07 - PA23053	9100	86	82
P4_09 - PA23119	9100	80	67
P4_11 - PA23066	9100	83	81
P5_03 - PA23035	9100	95	85
P5_05 - PA23012	9100	88	74
P5_07 - PA23051	9100	94	87
P6_05 - PA23022	9100	83	76
P7_05 - PA23034	9100	97	95
P7_06 - PA23044	9100	90	70
Average Percent Agreement		88	80

8.2.5 Summary of TIMSS Advanced 2008 Item Statistics Review

Based on the information from the comprehensive collection of item analyses and reliability data that were computed and summarized for TIMSS Advanced 2008, the TIMSS & PIRLS International Study Center thoroughly reviewed all item statistics for every participating country to ensure that the items were performing comparably across countries. Specifically, items with the following problems were considered for possible deletion from the international database:

- An error was detected during the TIMSS Advanced 2008 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 an item to have a negative point-biserial, or, for an item with more than 1 score point, a non-monotonic relationship between score level and total score.
- The item-by-country interaction results showed a large negative interaction for a particular country.
- For constructed-response items, the within-country scoring reliability data showed a score agreement of less than 70 percent.
- For trend items, an item performed substantially differently in 2008 compared to 1995, or an item was not included in the 1995 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 was consulted before deciding how the item should be

9 Chapter 3 describes the process of translation verification applied to the TIMSS Advanced 2008 instruments.

treated. If a problem was detected by the TIMSS & PIRLS 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 Advanced 2008 achievement data involved a review of 143 items for 10 countries and resulted in the detection of very few items that were inappropriate for international comparisons. The few items singled out in the review process were mostly items with differences attributable to either translation or printing problems. The following is a list of deleted items as well as a list of recodings made to constructed-response items.

Advanced Mathematics

Items deleted

ALL COUNTRIES

M2_04 (MA13014) – attractive distractor

LEBANON

M5_08 (MA23082) – printing error

Constructed-response items needing category recoding

ALL COUNTRIES

M3_06A (MA13026A) – recode 11 to 71
M3_06B (MA13026B) – recode 11 to 72
M3_08 (MA13028) – recode 20 to 10, 10 to 70, 11 to 71, 12 to 72
M4_07 (MA23166) – recode 20 to 10, 21 to 11, 10 to 70, 11 to 71, 70 to 72

Physics

Items deleted

ALL COUNTRIES

P1_07 (PA13007) – attractive distractor
P2_10 (PA13020) – low discrimination
P3_07B (PA13027B) – percent omitted too high

SWEDEN

P4_04 (PA23104) – negative discrimination

Constructed-response items needing category recoding

ALL COUNTRIES

P3_03 (PA13023) – recode 11 to 72
P3_05 (PA13025) – recode 29 to 19
P3_06 (PA13026) – recode 20 to 10, 29 to 19, 10 to 72, 19 to 79
P3_07A (PA13027A) – recode 20 to 10, 21 to 11, 22 to 19, 29 to 19, 10 to 70, 11 to 71
P4_10 (PA23088) – recode 11 to 71
P6_04 (PA23072) – recode 20 to 10, 10 to 11
P6_07 (PA23078) – recode 20 to 10, 10 to 11, 11 to 12

8.3 The TIMSS Advanced 2008 Scaling Methodology[10]

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.

10 A more detailed description of the TIMSS scaling methodology is given in Chapter 11 of the *TIMSS 2007 Technical Report* (Foy, Galia, & Li, 2008).

Three distinct IRT models, depending on item type and scoring procedure, were used in the analysis of the TIMSS Advanced 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 student'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, that were scored as correct or incorrect; and a two-parameter model, for constructed-response items with two response options, that also were scored as correct or incorrect. Since each of these item types has 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 response options.

8.3.1 Proficiency Estimation Using Plausible Values

Most cognitive testing endeavors to assess the performance of individual students 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 estimates of individual student ability is negligible under these conditions, the distribution of student ability, or its joint distribution with other variables, can be approximated using each individual student's estimated ability.

For the distribution of proficiencies in large populations, more efficient estimates can be obtained from a matrix-sampling design such as that used in TIMSS Advanced. This design solicits relatively few responses from each sampled student while maintaining a wide range of content representation when responses are aggregated across all students. With this approach, the advantage of estimating population characteristics more efficiently is offset to some degree by the inability to make precise statements about individuals. The uncertainty associated with individual student ability estimates becomes too large to be ignored.

Plausible values methodology was developed as a way to address this issue. Instead of first computing estimates of individual student abilities and then aggregating these to estimate population parameters, the plausible values approach uses all available data—students' responses to the items they were administered together with all background data—to estimate directly the characteristics of student populations and subpopulations. Although these directly estimated population characteristics could be used for reporting purposes, the usual plausible values approach generates multiple imputed scores, called plausible values, from the estimated ability distributions and uses these in analyses and reporting, making use of standard statistical software. By including all the available background data in the model, a process known as "conditioning," relationships between these background variables and the estimated proficiencies are appropriately accounted for in the plausible values. Because of this, analyses conducted using plausible values provide an accurate representation of these underlying relationships.

Plausible values are not intended to be estimates of individual student scores, but rather are imputed scores for similar students—students with similar response patterns and background characteristics in the sampled population—that may be used to estimate population characteristics correctly. When the underlying model is correctly specified, plausible values 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. Taking the average of the plausible values does not yield suitable estimates of individual student scores.[11]

8.4 Implementing the Scaling Procedures for the TIMSS Advanced Assessment Data

The application of IRT scaling and plausible values methodology to the data from the TIMSS Advanced 2008 assessments involved four major tasks: calibrating the achievement test items (estimating model parameters for each item), creating principal components from the student questionnaire data for use in conditioning, generating proficiency scores for advanced mathematics and for physics, and placing these proficiency scores on the scales—one for advanced mathematics and one for physics—used to report the results from the TIMSS Advanced assessments in 1995.

Before scaling the 2008 assessment data, however, the data from the 1995 assessments had to be rescaled from the one-parameter Rasch model used in 1995 to the multi-parameter models that have been in use in TIMSS since 1999.

8.4.1 Rescaling the Data from the TIMSS Advanced 1995 Assessments

The students' responses to the achievement items and to the questions in the student background questionnaire from the TIMSS Advanced 1995 international database provided the data for rescaling the TIMSS Advanced 1995 data. The TIMSS Advanced 1995 assessments included 68 items for advanced mathematics and 66 items for physics. These items were classified into the content and cognitive domains defined in the *TIMSS Advanced 2008 Assessment Frameworks* (Garden, et al., 2006) in preparation for trend scaling. Of the 134 items, 10 advanced

11 For further discussion, see Mislevy, Beaton, Kaplan, and Sheehan (1992) and von Davier, Gonzalez, and Mislevy (2009).

mathematics items and 4 physics items did not fit any framework classification and thus were omitted from the TIMSS Advanced 1995 rescaling since they were no longer appropriate for the domains specified in the 2008 frameworks. Also, one advanced mathematics item and three physics items were omitted for the reasons given in Section 8.2.5. Finally, one physics item released after the 1995 assessment (PA13052), was omitted because of poor discrimination. Some trend items that required recoding in the 2008 assessments also were recoded in the 1995 database.

All countries that participated in TIMSS Advanced 1995 were included in the item calibrations. Exhibit 8.10 presents the sample sizes for the countries included in the TIMSS Advanced 1995 item calibrations.[12]

Exhibit 8.10 Sample Sizes for Item Calibrations of the TIMSS Advanced 1995 Assessments

Country	Advanced Mathematics	Physics
Australia	645	661
Austria	782	777
Canada	2,781	2,367
Cyprus	391	368
Czech Republic	1,101	1,087
France	1,071	1,110
Germany	2,296	723
Greece	456	459
Italy	398	—
Latvia	—	708
Lithuania	734	—
Norway	—	1,048
Russian Federation	1,638	1,233
Slovenia	1,536	747
Sweden	1,001	1,012
Switzerland	1,404	1,371
United States	2,785	3,114
Total	**19,019**	**16,785**

12 Because Denmark and Israel failed to satisfy the 1995 sampling guidelines, they were not included in the item calibrations for the rescaling, as was also the case for the original scaling.

The item calibrations were conducted separately for each subject by the TIMSS & PIRLS International Study Center using the commercially-available Parscale software (Muraki & Bock, 1991; version 4.1). The two- and three-parameter and polytomous IRT models were fitted to the data to produce item parameter estimates. These new estimated values can be found in Exhibits D.1 and D.2 of Appendix D. These parameter estimates then became part of the input for producing proficiency scores.

A principal components analysis was run separately within each country to generate input for the conditioning step. The estimated proficiency scores are conditioned on the student background variables to improve the reliability of sub-population reporting. Principal components analysis is used to reduce the number of conditioning variables to a manageable size. The usual TIMSS approach retains the number of principal components that account for at least 90 percent of the variability in the student background data. Since most countries in 1995 had small sample sizes, the 90 percent criterion was reduced to 70 percent to minimize over-specification in the conditioning model, provided the number of components retained did not exceed 10 percent of the sample size—in which case the number of components was limited to 10 percent of the sample size. Exhibit 8.11 displays the total number of variables considered for conditioning and the number of principal components selected for each country.

The generation of IRT proficiency scores was conducted separately for each country and for each subject using Educational Testing Service's MGROUP program (Sheehan, 1985; version 3.2).[13] MGROUP 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

13 The MGROUP program was provided by ETS under contract to the TIMSS & PIRLS International Study Center at Boston College. It is now commercially available as DESI.

Exhibit 8.11 Number of Variables and Principal Components for Conditioning in Rescaling the TIMSS Advanced 1995 Assessments

Countries	Advanced Mathematics			Physics		
	Number of Primary Conditioning Variables	Total Number of Principal Components	Number of Principal Components Retained	Number of Primary Conditioning Variables	Total Number of Principal Components	Number of Principal Components Retained
Australia	2	553	64	2	560	66
Austria	2	542	78	2	545	77
Canada	3	573	148	3	570	134
Cyprus	2	562	39	2	565	36
Czech Republic	2	565	110	2	549	108
Denmark	2	545	123	2	533	65
France	2	434	107	2	435	111
Germany	2	542	147	2	526	72
Greece	2	551	45	2	550	45
Israel	2	589	95	2	583	85
Italy	2	531	39	—	—	—
Latvia	—	—	—	2	579	70
Lithuania	2	542	73	—	—	—
Norway	—	—	—	3	579	104
Russian Federation	2	602	152	2	598	123
Slovenia	2	582	141	2	573	74
Sweden	2	576	100	2	571	101
Switzerland	4	544	127	4	546	125
United States	2	612	154	2	618	166

that represent student proficiency. Exhibit 8.12 shows the sample sizes of the countries for which proficiency scores were generated.[14]

The reporting metrics for the rescaled 1995 data were established to give the distribution of TIMSS Advanced 1995 proficiency scores in advanced mathematics and in physics a mean of 500 and a standard deviation of 100, with all 1995 countries included in the item calibrations contributing equally. Extreme scale values were truncated, giving plausible values a minimum of 5 and a maximum of 995.

14 Denmark and Israel, which had been excluded from the item calibrations, were included among the countries for which proficiency scores were produced.

Exhibit 8.12 Sample Sizes for Proficiency Estimation of the TIMSS Advanced 1995 Assessments

Country	Advanced Mathematics	Physics
Australia	645	661
Austria	782	777
Canada	2,781	2,367
Cyprus	391	368
Czech Republic	1,101	1,087
Denmark	1,388	654
France	1,071	1,110
Germany	2,296	723
Greece	456	459
Israel	953	853
Italy	398	—
Latvia	—	708
Lithuania	734	—
Norway	—	1,048
Russian Federation	1,638	1,233
Slovenia	1,536	747
Sweden	1,001	1,012
Switzerland	1,404	1,371
United States	2,785	3,114
Total	**21,360**	**18,292**

8.4.2 Calibrating the TIMSS Advanced 2008 Assessment Data

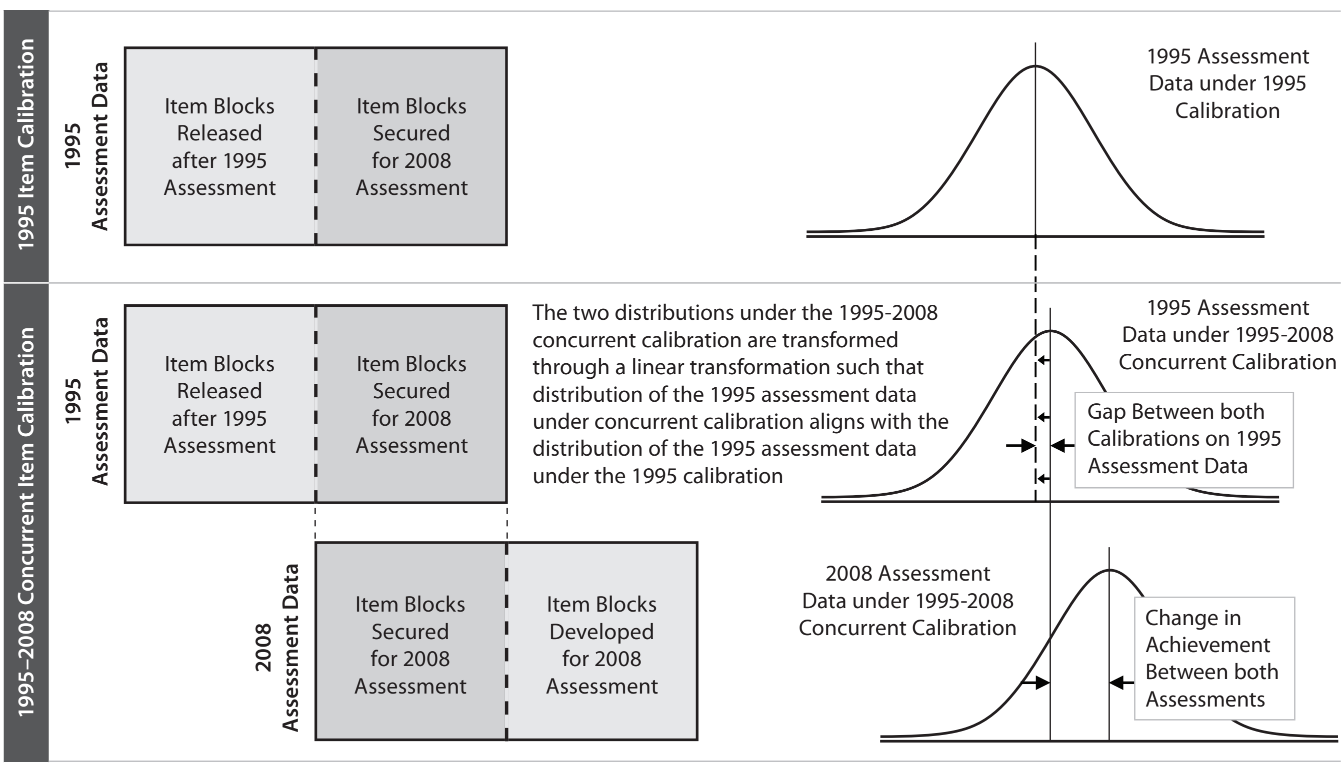
As described in the *TIMSS Advanced 2008 Assessment Frameworks* (Garden, et al., 2006), the TIMSS Advanced 2008 assessments consisted of a total of seven advanced mathematics blocks and seven physics blocks, distributed across eight assessment booklets. Each block contained either advanced mathematics or physics items, drawn from a range of content and cognitive domains. The seven mathematics blocks were designated M1 through M7, and the seven physics blocks P1 through P7. Blocks M1 through M3 and P1 through P3 contained items that were previously used in the 1995 assessments, whereas blocks M4 through M7 and P4 through P7 consisted of newly-developed items for the 2008 assessments. Each assessment booklet contained three blocks of either all advanced mathematics items or all physics items. The booklets were distributed among the students in each sampled class according to a scheme that ensured equivalent random samples of students responding to each booklet.

Separate IRT scales were constructed for reporting overall student achievement in advanced mathematics and in physics. Concurrent item calibrations were conducted by the TIMSS & PIRLS International Study Center using Parscale, and included data from the TIMSS Advanced 2008 assessments and the TIMSS Advanced 1995 assessments to measure trends from 1995. The calibrations used all available data from each country's TIMSS Advanced student samples, which were weighted such that each country contributed equally.

The first step in constructing the scales for TIMSS Advanced 2008 was to estimate the IRT model item parameters for each item on each of the scales through a concurrent calibration of both sets of assessment data—1995 and 2008. It was then possible to obtain the mean and standard deviation of the latent ability distributions of students in both assessments using item paramters from the concurrent calibration. The difference between these two distributions was the change in achievement from 1995 to 2008. The second step was to find the linear transformation that transformed the distribution of the 1995 assessment data under the 1995-2008 concurrent calibration to match the distribution of these data under the 1995 calibration. The third step was to apply this same transformation to the 2008 assessment data scaled using the concurrent calibration. This placed the 2008 assessment data on the metric of the 1995 assessment—i.e., a scale with a mean of 500 and a standard deviation of 100.

Exhibit 8.13 illustrates how the concurrent calibration approach was applied in the context of TIMSS Advanced 2008 trend scaling. The observed gap between the distribution of the 1995 data under the 1995 item calibration and the 1995 data under the 1995-2008 concurrent calibration was small and arose from slight differences in the item parameter estimations, which in turn were due mostly to the 1995 assessment data being calibrated with the 2008 assessment data. The linear transformation removed this gap by shifting the two distributions from the concurrent calibration such that the distribution of the 1995 assessment data from the concurrent calibration aligned with the distribution of the 1995 assessment data from the 1995 calibration, while preserving the gap between the 1995 and 2008 assessment data under the concurrent calibration. This latter gap was the change in achievement between the previous and current assessments that TIMSS Advanced set out to measure as its trend.

Exhibit 8.13 Concurrent Calibration Model Used for TIMSS Advanced 2008



Having estimated the item parameters from the 1995-2008 concurrent calibration, new achievement distributions were generated by applying these item parameters to the 1995 assessment and to the 2008 assessment data. Following the procedure outlined above, the next step was to identify the linear transformation that transformed the 1995 assessment distribution generated by the 1995-2008 concurrent calibration item parameters to match the 1995 assessment distribution generated by the item parameters from the 1995 rescaling, and to apply this same transformation to the 2008 assessment data distribution (also generated by the concurrent calibration item parameters).

Exhibit 8.14 shows the distribution of items included in the TIMSS Advanced 2008 concurrent calibrations for reporting trends in overall advanced mathematics and physics. All the data from both the 1995 and 2008 assessments were included. Items were categorized as unique to the 1995 assessment, common to both assessments, or unique to the 2008 assessment. For advanced mathematics, the 2008 assessment contributed 45 items worth 51 score points that were unique to 2008 and 26 items worth 28 score points that also were included in the 1995 assessment. The 1995 assessment also contributed 31 items worth 40 score points that were released in 1995. For physics, the 2008 assessment contributed 45 items worth 51 score points that were unique to 2008 and 23 items worth 26 score points that also were included in the 1995 assessment. The 1995 assessment also contributed 35 items worth 42 score points that were released in 1995.

Exhibit 8.14 Items Included in the TIMSS Advanced 2008 Concurrent Item Calibrations

TIMSS 2008 Trend Scales	Items Unique to the TIMSS Advanced 2008 Assessments		Items Common to the TIMSS Advanced 2008 and 1995 Assessments		Items Unique to the TIMSS Advanced 1995 Assessments		TOTAL	
	Number	Points	Number	Points	Number	Points	Number	Points
Advanced Mathematics	45	51	26	28	31	40	102	119
Physics	45	51	23	26	35	42	103	119

Because of the small number of countries that participated in both TIMSS Advanced assessments, concurrent item calibrations were conducted using data from all the countries that participated in either the 1995 assessments or the 2008 assessments. To construct the advanced mathematics scale, the calibration included 19,019 students from 15 countries in the 1995 assessment and 22,242 students from 10 countries in the 2008 assessment. The item parameters established in this calibration were used subsequently for estimating student proficiency scores in advanced mathematics for the 10 countries that participated in TIMSS Advanced 2008. The national samples included in the calibration for reporting trends in advanced mathematics are presented in Exhibit 8.15.

Similarly, to construct the physics scale, the item calibration was conducted using data from countries that participated in either the 1995 assessment or the 2008 assessment. The physics concurrent calibration included 16,785 students from 15 countries in the 1995 assessment and 16,489 students from 9 countries in the 2008 assessment. The item parameters obtained in this calibration were used subsequently for estimating student proficiency scores in physics for the nine countries that participated in TIMSS Advanced 2008. Exhibit 8.16 presents the national samples included in the calibration for reporting trends in Physics.

Exhibits D.3 and D.4 of Appendix D display the item parameters for advanced mathematics and physics, respectively, generated from the concurrent calibration of the 1995 and the 2008 data. As a by-product of the calibrations, interim scores in advanced mathematics and physics were produced for use in constructing conditioning variables.

Exhibit 8.15 Sample Sizes for Concurrent Item Calibration of Advanced Mathematics for the TIMSS Advanced 1995 and 2008 Assessments

Country	1995 Assessment	2008 Assessment
Countries in Both Cycles		
Italy	398	2,143
Russian Federation	1,638	3,185
Slovenia	1,536	2,156
Sweden	1,001	2,303
Countries in 1995		
Australia	645	—
Austria	782	—
Canada	2,781	—
Cyprus	391	—
Czech Republic	1,101	—
France	1,071	—
Germany	2,296	—
Greece	456	—
Lithuania	734	—
Switzerland	1,404	—
United States	2,785	—
Countries in 2008		
Armenia	—	858
Iran, Islamic Rep. of	—	2,425
Lebanon	—	1,612
Netherlands	—	1,537
Norway	—	1,932
Philippines	—	4,091
Total	**19,019**	**22,242**

Exhibit 8.16 Sample Sizes for Concurrent Item Calibration of Physics for the TIMSS Advanced 1995 and 2008 Assessments

Country	1995 Assessment	2008 Assessment
Countries in Both Cycles		
Norway	1,048	1,640
Russian Federation	1,233	3,166
Slovenia	747	1,097
Sweden	1,012	2,291
Countries in 1995		
Australia	661	—
Austria	777	—
Canada	2,367	—
Cyprus	368	—
Czech Republic	1,087	—
France	1,110	—
Germany	723	—
Greece	459	—
Latvia	708	—
Switzerland	1,371	—
United States	3,114	—
Countries in 2008		
Armenia	—	894
Iran, Islamic Rep. of	—	2,434
Italy	—	1,861
Lebanon	—	1,595
Netherlands	—	1,511
Total	**16,785**	**16,489**

8.4.3 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 not reached when the item itself and the item immediately preceding it were not answered, and there were no other items completed in the remainder of the booklet.

In TIMSS Advanced 2008, as in TIMSS Advanced 1995 and previous TIMSS assessments, not-reached items were treated differently in estimating item parameters than they were in generating student proficiency scores. In estimating the values of the item parameters, items in the TIMSS Advanced assessment booklets that were considered not to have been reached by students were treated as if they had not been administered. This approach was considered optimal for parameter estimation. However, not-reached items were always considered as incorrect responses when student proficiency scores were generated.

8.4.4 Evaluating the Fit of IRT Models to the TIMSS Advanced 2008 Data

After the concurrent item calibrations were completed, checks were performed to verify that the item parameters obtained from Parscale adequately reproduced the observed distribution of student responses across the proficiency continuum. The fit of the IRT models to the TIMSS Advanced data was examined by comparing the 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 proficiencies for each student that responded to an item. When the empirical results fall near the fitted curves for any given item, the IRT model fits the data well and leads to more accurate and reliable measurement of the underlying proficiency scale. Graphical plots of these response function curves are called item characteristic curves (ICC).

Exhibit 8.17 shows an ICC of the empirical and fitted item response functions for a dichotomous multiple-choice item. In the graph, the horizontal axis represents the proficiency scale; and the vertical axis, the probability of a correct response. The fitted curve based on the estimated item parameters is shown as a solid line. Empirical results are represented by circles. The empirical results were obtained by dividing the proficiency scale into intervals of equal size and then counting the number of students responding to the item whose estimated a-priori (EAP) scores from Parscale fell in each interval. Then the proportion of students in each interval that responded correctly to the item was calculated. In the exhibit, the center of each circle represents this empirical proportion of correct responses. The size of each circle is proportional to the number of students contributing to the estimation of its empirical proportion correct.

Exhibit 8.17 Example Item Response Function for a TIMSS Advanced 2008 Dichotomous Item

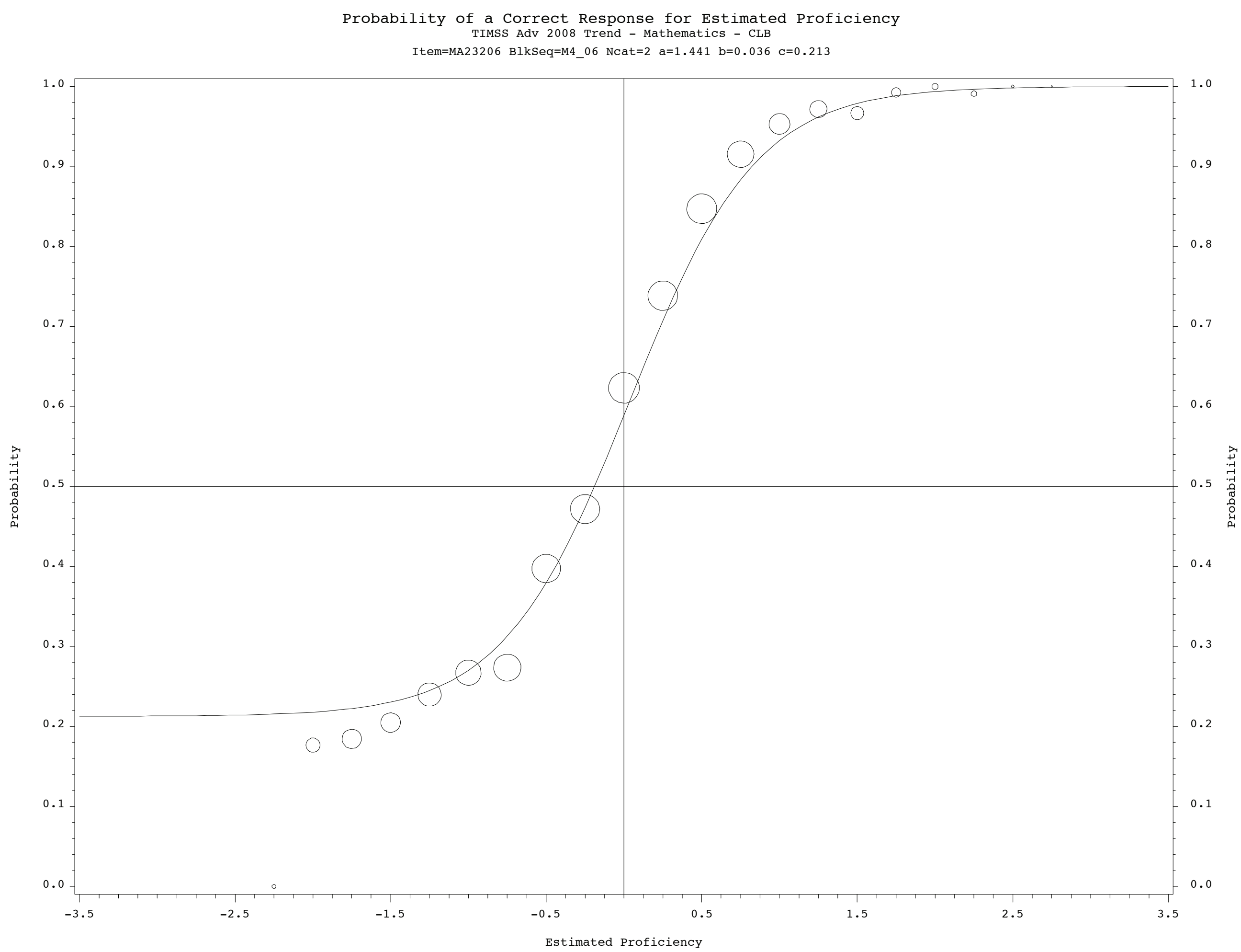
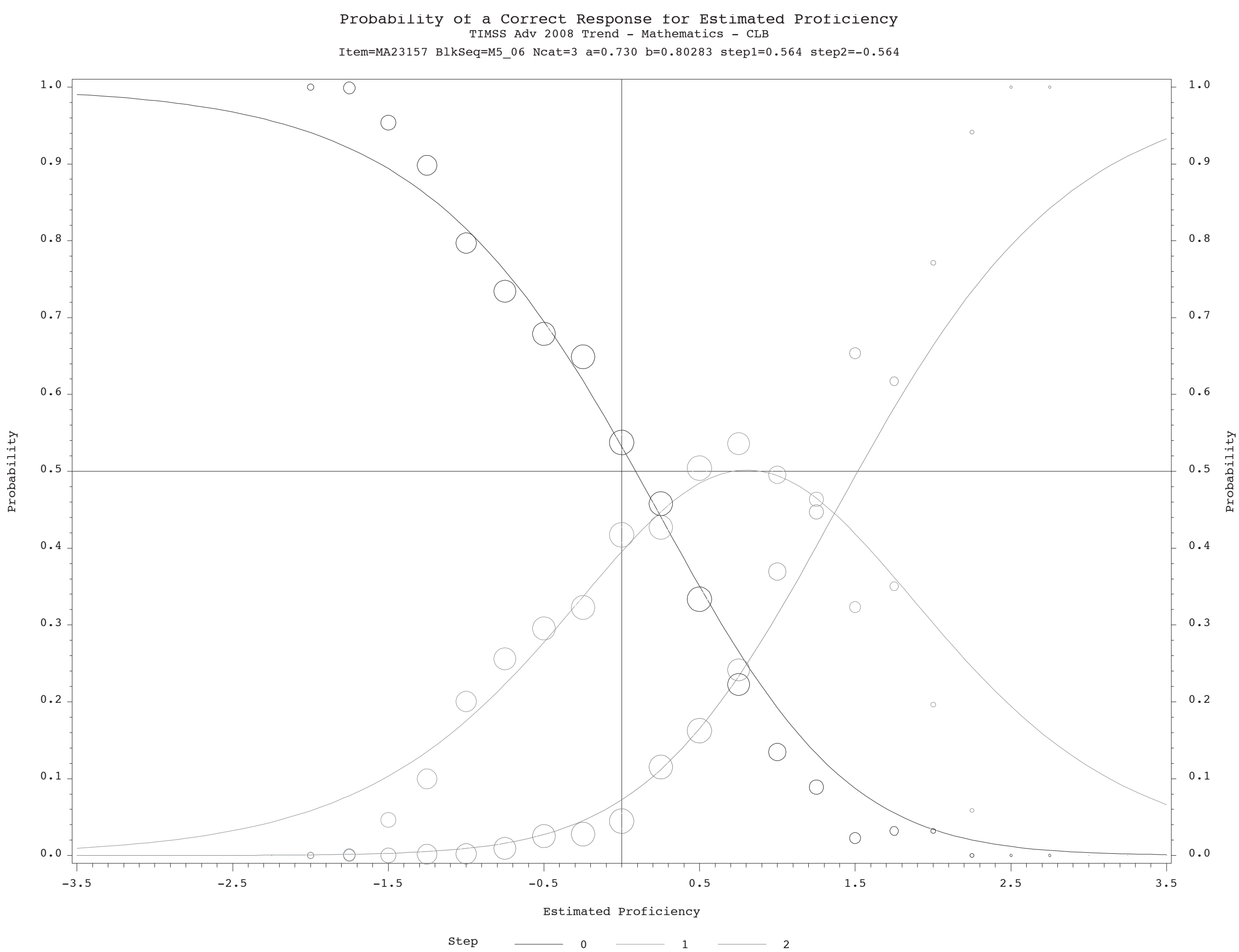


Exhibit 8.18 contains an ICC of the empirical and fitted item response functions for a polytomous constructed-response item with three response categories—0, 1, and 2 points. As for the dichotomous item plot, the horizontal axis represents the proficiency scale, but the vertical axis represents the probability of having a response in a given response category. The fitted curves based on the estimated item parameters are shown as solid lines. Empirical results are represented by circles. The interpretation of the circles is the same as in Exhibit 8.17. The curve starting at the top left of the chart plots the probability of a score of zero on the item, which decreases as proficiency increases.

Exhibit 8.18 Example Item Response Function for a TIMSS Advanced 2008 Polytomous Item



The bell-shaped curve shows the probability of a score of 1 point—starting low for low-ability students, reaching a maximum for medium-ability students, and decreasing for high-ability students. The curve ending at the top right corner of the chart shows the probability of a score of 2 points—full credit, starting low for low-ability students and increasing as proficiency increases.

8.4.5 Variables for Conditioning the TIMSS Advanced 2008 Data

Because there were so many background variables that could be used in conditioning, TIMSS Advanced followed the practice established by NAEP and followed by 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 Advanced background data were constructed as follows:

- For categorical variables (questions with a small number of fixed response options), a "dummy coded" variable was created for each response option, with a value of 1 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.
- Background variables with numerous categories (such as year of birth or time spent doing homework) were recoded using criterion scaling.[15] This was done by replacing each response option with an interim achievement score. For the overall advanced mathematics scale, the interim achievement score was the advanced mathematics score produced from the item calibration. For the overall physics scale, the interim achievement score was the physics score produced from the item calibration.
- Separately for each TIMSS country, all the dummy-coded and criterion-scaled variables were included in a principal components

15 The process of generating criterion-scaled variables is described in Beaton (1969).

analysis. Those principal components accounting for 90 percent of the variance of the background variables were retained for use as conditioning variables. However, if the selected number of principal components exceeded 5 percent of the student sample size, the number of selected principal components was reduced to 5 percent of the student sample size. Because the principal components analysis was performed separately for each country, different numbers of principal components were required to account for 90 percent of the common variance in each country's background variables.

In addition to the principal components, student gender (dummy coded), the language of the test (dummy coded), and an indicator of the class in the school to which the student belonged (criterion scaled) were included as primary conditioning variables, thereby accounting for most of the variance among students and preserving the between- and within-class variance structure in the scaling model. Conditioning variables were needed for all the TIMSS Advanced 2008 participants, as well as for all the TIMSS Advanced 1995 countries. Exhibits 8.19 and 8.20 show the total number of variables that were considered for conditioning and the number of principal components selected for each country for advanced mathematics and physics, respectively.

Exhibit 8.19 Number of Variables and Principal Components for Conditioning Advanced Mathematics in TIMSS Advanced 2008

Countries	1995 Assessment			2008 Assessment		
	Number of Primary Conditioning Variables	Total Number of Principal Components	Number of Principal Components Retained	Number of Primary Conditioning Variables	Total Number of Principal Components	Number of Principal Components Retained
Armenia	—	—	—	2	271	42
Australia	2	553	64	—	—	—
Austria	2	542	78	—	—	—
Canada	3	573	148	—	—	—
Cyprus	2	562	39	—	—	—
Czech Republic	2	565	110	—	—	—
Denmark	2	545	123	—	—	—
France	2	434	107	—	—	—
Germany	2	542	147	—	—	—
Greece	2	551	45	—	—	—
Iran, Islamic Rep. of	—	—	—	2	279	121
Israel	2	589	95	—	—	—
Italy	2	531	39	2	270	107
Lebanon	—	—	—	3	277	80
Lithuania	2	542	73	—	—	—
Netherlands	—	—	—	2	267	76
Norway	—	—	—	2	270	96
Philippines	—	—	—	2	276	156
Russian Federation	2	602	152	2	277	157
Slovenia	2	582	141	2	270	107
Sweden	2	576	100	2	268	115
Switzerland	4	544	127	—	—	—
United States	2	612	154	—	—	—

Exhibit 8.20 Number of Variables and Principal Components for Conditioning Physics in TIMSS Advanced 2008

Countries	1995 Assessment			2008 Assessment		
	Number of Primary Conditioning Variables	Total Number of Principal Components	Number of Principal Components Retained	Number of Primary Conditioning Variables	Total Number of Principal Components	Number of Principal Components Retained
Armenia	—	—	—	2	275	44
Australia	2	560	66	—	—	—
Austria	2	545	77	—	—	—
Canada	3	570	134	—	—	—
Cyprus	2	565	36	—	—	—
Czech Republic	2	549	108	—	—	—
Denmark	2	533	65	—	—	—
France	2	435	111	—	—	—
Germany	2	526	72	—	—	—
Greece	2	550	45	—	—	—
Iran, Islamic Rep. of	—	—	—	2	282	121
Israel	2	583	85	—	—	—
Italy	—	—	—	2	205	93
Lebanon	—	—	—	3	281	79
Latvia	2	579	70	—	—	—
Netherlands	—	—	—	2	272	75
Norway	3	579	104	2	270	82
Russian Federation	2	598	123	2	283	158
Slovenia	2	573	74	2	272	54
Sweden	2	571	101	2	268	114
Switzerland	4	546	125	—	—	—
United States	2	618	166	—	—	—

8.4.6 Generating IRT Proficiency Scores for the TIMSS Advanced 2008 Data

MGROUP was used to generate the IRT proficiency scores. Exhibit 8.21 shows the student sample sizes—from the 1995 assessments and the 2008 assessments—for which proficiency scores, using the item parameters obtained from the concurrent calibration, were generated on the overall advanced mathematics and physics scales.

Exhibit 8.21 Sample Sizes for TIMSS Advanced Proficiency Estimation

Country	Advanced Mathematics		Physics	
	1995 Assessment	2008 Assessment	1995 Assessment	2008 Assessment
Armenia	—	858	—	894
Australia	645	—	661	—
Austria	782	—	777	—
Canada	2,781	—	2,367	—
Cyprus	391	—	368	—
Czech Republic	1,101	—	1,087	—
France	1,071	—	1,110	—
Germany	2,296	—	723	—
Greece	456	—	459	—
Iran, Islamic Rep. of	—	2,425	—	2,434
Italy	398	2,143	—	1,861
Latvia	—	—	708	—
Lebanon	—	1,612	—	1,595
Lithuania	734	—	—	—
Netherlands	—	1,537	—	1,511
Norway	—	1,932	1,048	1,640
Philippines	—	4,091	—	—
Russian Federation	1,638	3,185	1,233	3,166
Slovenia	1,536	2,156	747	1,097
Sweden	1,001	2,303	1,012	2,291
Switzerland	1,404	—	1,371	—
United States	2,785	—	3,114	—
Total	**19,019**	**22,242**	**16,785**	**16,489**

8.4.7 Transforming the Advanced Mathematics and Physics Scores to Measure Trends

As part of rescaling the data from the TIMSS Advanced 1995 assessments using 2- and 3-parameter models as described in Section 8.4.1, the TIMSS Advanced reporting scales were established by setting the average of the mean scores of the countries included in the rescaling item calibrations to 500 and the standard deviation to 100. To provide results for the 2008 assessments that would be directly comparable to the results from the 1995 assessments, the 2008 proficiency scores (plausible values) for advanced mathematics and physics had to be transformed to the TIMSS Advanced scales established with the 1995 data. This was accomplished through a linear transformation of the proficiency scores from the 1995-2008 concurrent calibration such that the 1995 proficiency distribution from the concurrent calibration aligned itself with the 1995 proficiency distribution from the 1995 rescaling calibration.

The means and standard deviations of the 1995 advanced mathematics and physics scores produced in 2008—the plausible values from the TIMSS Advanced 1995 assessment data based on the 1995-2008 concurrent item calibrations—were made to match the means and standard deviations of the scores calculated for the TIMSS Advanced 1995 assessment—the plausible values produced using the item calibration from scaling the 1995 assessment data—by applying the appropriate linear transformations. These linear transformations were given by:

$$PV_{k,i}^{*} = A_{k,i} + B_{k,i} \cdot PV_{k,i}$$

where

$PV_{k,i}$ was plausible value i for scale k prior to transformation,

$PV_{k,i}^{*}$ was plausible value i for scale k after transformation, and $A_{k,i}$ and $B_{k,i}$ were the linear transformation constants for plausible value i of scale k.

The linear transformation constants were obtained by first computing the international means and standard deviations of the proficiency scores for the overall advanced mathematics and physics scales using the plausible values from the 1995 scaling for the 1995 countries included in the concurrent calibration. Next, the same calculation was done using the plausible values from the 1995 assessment data based on the 1995-2008 concurrent calibration for the same set of countries. The linear transformation constants were thus defined as:

$$B_{k,i} = \sigma^*_{k,i} / \sigma_{k,i}$$

$$A_{k,i} = \mu^*_{k,i} - B_{k,i}\,\mu_{k,i}$$

where

$\mu_{k,i}{}^*$ was the international mean of scale k based on plausible value i obtained from scaling the 1995 assessment data;

$\mu_{k,i}$ was the international mean of scale k based on plausible value i from the TIMSS Advanced 1995 assessment data based on the 1995-2008 concurrent item calibrations;

$\sigma_{k,i}{}^*$ was the international standard deviation of scale k based on plausible value i obtained from scaling the 1995 assessment data;

$\sigma_{k,i}$ was the international standard deviation of scale k based on plausible value i from the TIMSS Advanced 1995 assessment data based on the 1995-2008 concurrent item calibrations.

Exhibit 8.22 shows the linear transformation constants that were computed for TIMSS Advanced 2008 for the advanced mathematics and physics scales. Once these linear transformation constants had been established, all of the 2008 advanced mathematics and physics proficiency scores—the plausible values generated from the 2008 assessment data—for all the participating countries were transformed by applying the linear transformations. This provided advanced mathematics and physics student achievement scores for the 2008 assessments that were directly comparable to the rescaled scores from the 1995 assessment data.

Exhibit 8.22 Linear Transformation Constants Applied to the TIMSS Advanced 2008 Proficiency Scores

Scale	Plausible Value	TIMSS Advanced 1995 Scores Using 1995 Item Calibrations		TIMSS Advanced 1995 Scores Using 1995–2008 Concurrent Calibrations		$A_{k,i}$	$B_{k,i}$
		Mean	Standard Deviation	Mean	Standard Deviation		
Advanced Mathematics	PV1	500.00109	99.99451	0.11465	0.88983	487.11733	112.37442
	PV2	500.00010	99.99951	0.10189	0.89226	488.58124	112.07392
	PV3	500.00047	99.99754	0.11694	0.89312	486.90788	111.96396
	PV4	500.00004	99.99981	0.10578	0.88989	488.11346	112.37280
	PV5	500.00107	99.99447	0.10317	0.89075	488.41976	112.25904
Physics	PV1	500.00000	100.00000	-0.02718	0.97905	502.77604	102.14010
	PV2	500.00000	100.00000	-0.02970	0.98882	503.00383	101.13043
	PV3	500.00000	100.00000	-0.02392	0.97793	502.44621	102.25631
	PV4	500.00000	100.00000	-0.02718	0.98387	502.76209	101.63923
	PV5	500.00000	100.00000	-0.03582	0.97809	503.66228	102.23961

Note: The means and standard deviations for advanced mathematics based on the 1995 item calibrations are affected by rare cases of very low scores that were truncated.

8.5 The TIMSS Advanced 2008 International Benchmarks of Student Achievement in Advanced Mathematics and Physics

To describe student performance at various points along the TIMSS 2008 advanced mathematics and physics achievement scales, TIMSS Advanced 2008 used scale anchoring to summarize and describe student achievement at three points on the advanced mathematics and physics scales—Advanced International Benchmark (625), High

International Benchmark (550), and Intermediate International Benchmark (475). For a description of performance at the international benchmarks, please see the *TIMSS Advanced 2008 International Report* (Mullis, Martin, Robitaille, & Foy, 2009).

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 those anchor points (the international benchmarks) can answer correctly. The items so identified were grouped by content domain within benchmarks for review by mathematics and physics experts. The committee members[16] examined the content of each item and determined the kind of mathematics or physics knowledge or skill demonstrated by students who responded correctly to the item. They then summarized the detailed list of item competencies in a brief description of achievement at each international benchmark. This procedure resulted in a content-referenced interpretation of the achievement results that can be considered in light of the TIMSS Advanced 2008 advanced mathematics and physics frameworks.

As the first step, students scoring within 20 scale score points of each benchmark were identified for the benchmark analysis. The score ranges around each international benchmark and the number of students scoring in each range for advanced mathematics and physics are shown in Exhibit 8.23. The range of 20 points above and below a benchmark provided an adequate sample in each group, yet was small enough so that performance at each benchmark anchor point was still distinguishable from the next.

Exhibit 8.23 Range Around Each International Benchmark and Number of Students Within Each Range

	Intermediate (475)	High (550)	Advanced (625)
Range of Scale Scores	455–495	530–570	605–645
Advanced Mathematics	2,826	2,752	1,138
Physics	2,201	2,369	1,327

16 In addition to Robert A. Garden, the TIMSS Advanced Mathematics Coordinator, and Svein Lie, the TIMSS Physics Coordinator, committee members included Carl Angell, Wolfgang Dietrich, Liv Sissel Gronmo, Torgeir Onstad, and David F. Robitaille.

Having identified the number of students scoring within each benchmark range, the next step was conducting the data analysis to determine which items anchored at each of the international benchmarks. An important feature of the scale anchoring method is that it yields descriptions of the performance demonstrated by students reaching each of the international benchmarks on the scales, and that the descriptions reflect demonstrably different accomplishments by students reaching each successively higher benchmark. Because the process entails the delineation of sets of items that students at each international benchmark are likely to answer correctly and that discriminate between one benchmark and the next, the criteria for identifying the items that anchor considers performance at more than one benchmark.

For *multiple-choice* items, a criterion of 65 percent was used for each benchmark being analyzed, since students would be likely (about two thirds of the time) to answer the item correctly. A criterion of less than 50 percent was used for the next lower benchmark, because with this response probability, students were more likely to have answered the item incorrectly than correctly. The criteria for each benchmark are outlined below.

- A multiple-choice item anchored at the Intermediate International Benchmark (475) if at least 65 percent of students scoring in the range answered the item correctly. Because this was the lowest benchmark described, there were no further criteria.
- A multiple-choice item anchored at the High International Benchmark (550) if at least 65 percent of students scoring in the range answered the item correctly, and less than 50 percent of students at the Intermediate International Benchmark answered the item correctly.

- A multiple-choice item anchored at the Advanced International Benchmark (625) if at least 65 percent of students scoring in the range answered the item correctly, and less than 50 percent of students at the High International Benchmark answered the item correctly.

To include all of the multiple-choice items in the anchoring process and provide information about content domains and cognitive processes that might not otherwise have had many anchor items, the concept of items that "almost anchored" was introduced. These were items that met slightly less stringent criteria for being answered correctly. The criteria to identify multiple-choice items that "almost anchored" were that at least 55 percent of students scoring in the range answered the item correctly and less than 50 percent of students at the next lowest benchmark answered correctly. To be completely inclusive for all items, items that met only the criterion that at least 55 percent of the students answered correctly (regardless of the performance of students at the next lower point) were also identified. The categories of items were mutually exclusive, and ensured that all of the items were available to inform the descriptions of student achievement at the anchor levels. A multiple-choice item was considered to be "too difficult" to anchor if less than 55 percent of students at the advanced benchmark answered the item correctly.

A somewhat less strict criterion was used for all the *constructed-response* items, because students had much less scope for guessing. For constructed-response items, the criterion of 50 percent was used for the benchmark without any discrimination criterion for the next lower benchmark. A constructed-response item anchored at one of the international benchmarks if at least 50 percent of students at that benchmark answered the item correctly. A constructed-response item was considered to be "too difficult" to anchor if less than 50 percent of students at the advanced benchmark answered the item correctly.

For students scoring in the range around each international benchmark, the percentage of those students that answered each 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 Advanced 2008 items were scored 1 point for a correct answer and 0 points for other answers. For these items, the percentage of students at each benchmark who answered each item correctly was computed. For the relatively few constructed-response items scored for partial or full credit, percentages were computed for the students receiving full credit. Then the criteria described above were applied to identify the items that anchored, almost anchored, and met only the 55 to 65 percent criteria. Exhibit 8.24 presents the number of advanced mathematics and physics items that anchored at each international benchmark.

Exhibit 8.24 Number of Items Anchoring at Each International Benchmark

	Intermediate (475)	High (550)	Advanced (625)	Too Difficult to Anchor	Total
Advanced Mathematics	16	23	21	11	71
Physics	17	14	22	15	68

In preparation for the committee review, the advanced mathematics and physics items were organized into separate binders. The items were grouped by international benchmark and, within benchmark, 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 55 to 65 percent criteria. The following information was included for each item: content area, cognitive domain, maximum points, answer key, release status, percent correct at each benchmark, and overall international percent correct. For constructed-response items, the scoring guides were included.

The TIMSS & PIRLS International Study Center staff convened the committee for a three-day meeting in Boston to complete three tasks: The committee (1) worked through each item and arrived at a short description of the knowledge, understanding, or skills demonstrated by students who answered the item correctly; (2) developed a description (in detailed and summary form) of the level of advanced mathematics or physics proficiency demonstrated by students at each of the three international benchmarks to publish in the TIMSS Advanced 2008 International Report; and (3) selected example items that supported and illustrated the benchmark descriptions to publish together with the descriptions.

8.6 Capturing the Uncertainty in the TIMSS Advanced 2008 Student Achievement Scores

To obtain estimates of students' proficiency in advanced mathematics and physics that were both accurate and cost-effective, TIMSS Advanced 2008 made extensive use of probability sampling techniques to sample students from national 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 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 Advanced 2008 International Report* (Mullis, Martin, Robitaille, & Foy, 2009) 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 student populations (sampling variance), and to inferring students' performance on the entire assessment from their performance on the subset of items that they took (imputation variance).

8.6.1 Estimating the Sampling Variance

The TIMSS Advanced 2008 sample design applied a stratified multi-stage 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, such a complex sample design complicates the task of computing standard errors to quantify sampling variability.

When, as in TIMSS Advanced, the sample design involves multi-stage cluster sampling, there are several options for estimating sampling errors that avoid the assumption of simple random sampling (Wolter, 1985). The jackknife 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 Advanced 2008 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 the entire original sample, and once again for each jackknife pseudo-replicate sample. The variation between the estimates for each of the jackknife replicate samples and the original sample estimate is the jackknife estimate of the sampling error of the statistic.

8.6.2 Constructing Sampling Zones

To apply the JRR technique used in TIMSS Advanced 2008, successive sampled schools were 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 participating schools to the first sampling zone, the third and fourth participating schools to the second zone, and so on. A maximum of 75 zones were used, although most countries had fewer because they generally sampled less than 150 schools. When more than 75 zones were constructed, as was the case in Lebanon, they were collapsed to keep the total number to 75.

Sampling zones were constructed within explicit strata. When there was an odd number of schools in an explicit stratum, either by design or because of school non-response, the students in the not-paired school were randomly divided to make up two "quasi" schools for the purposes of calculating the jackknife standard error.[17] Each sampling zone then consisted of a pair of schools or "quasi" schools. Exhibit 8.25 shows the number of sampling zones in each country.

Within each sampling zone, each school was assigned an indicator (u_j), coded randomly to 0 or 1, such that one school had a value of zero, and the other a value of 1. This indicator determined whether the weights for the sampled students in the school in this zone were to be doubled ($u_j = 1$) or zeroed ($u_j = 0$) for the purposes of creating the pseudo-replicate samples.

17 If the not-paired school consisted of 2 sampled classrooms, each classroom became a "quasi" school.

Exhibit 8.25 Number of Sampling Zones Used in the TIMSS Advanced 2008 Countries

Country	Advanced Mathematics	Physics
Armenia	55	55
Italy	46	46
Iran, Islamic Rep. of	60	60
Lebanon	75	75
Netherlands	56	58
Norway	55	52
Philippines	61	—
Russian Federation	45	47
Slovenia	53	66
Sweden	59	61

8.6.3 Computing the Sampling Variance Using the JRR Method

The formula for the sampling variance of a statistic t, based on the JRR algorithm used in TIMSS Advanced 2008, is given by the following equation:

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

where H is the total number of sampling zones in the sample of the country under consideration. The term $t(S)$ corresponds to the statistic of interest for the whole sample computed with the overall sampling weights.[18] The term $t(J_h)$ denotes the same statistic using the h^{th} jackknife replicate sample J_h and its set of replicate sampling weights, which are identical to the overall sampling weights, except for the students in the h^{th} sampling zone. In the h^{th} zone, all students belonging to one of the randomly selected schools of the pair were removed, and the students belonging to the other school in the zone were included twice. In practice, this was accomplished by recoding to zero the sampling weights for the students in the school to be excluded

18 The sampling weights are described in Chapter 4.

from the replication, and multiplying by 2 the sampling weights of the remaining students within the h^{th} pair. Each sampled student was assigned a vector of 75 replicate sampling weights W_{hi}, where h took values from 1 to 75. If W_{oi} was the overall sampling weight of student i, the h replicate weights for that student were computed as:

$$W_{hi} = W_{0i} \cdot k_{hi}$$

where

$$k_{hi} = \left\{ \begin{array}{ll} 2 \cdot u_j & \text{if student } i \text{ is in school } j \text{ of sampling zone } h \\ 1 & \text{otherwise} \end{array} \right\}$$

The school-level indicators u_j determined which students in a sampling zone would get zero weights and which ones would get double weights, on the basis of the school within the pair from which the students were sampled. The process of setting the k_{hi} values for all sampled students and across all sampling zones is illustrated in Exhibit 8.26. Thus, the computation of the JRR variance estimate for any statistic in TIMSS Advanced 2008 required the computation of the statistic up to 76 times for any given country: once to obtain the statistic for the full sample based on the overall weights W_{oi}, and up to 75 times to obtain the statistics for each of the jackknife replicate samples J_h using a set of replicate weights W_{hi}.

Exhibit 8.26 Construction of Replicate Weights Across Sampling Zones in TIMSS Advanced 2008

Sampling Zone	u_j	k_{hi} for Computing JRR Replicate Sampling Weights						
		1	2	3	…	h	…	75
1	0	0	1	1	…	1	…	1
	1	2						
2	0	1	0	1	…	1	…	1
	1		2					
3	0	1	1	0	…	1	…	1
	1			2				
⋮	⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮
h	0	1	1	1	…	0	…	1
	1					2		
⋮	⋮	⋮	⋮	⋮	…	⋮	⋱	⋮
75	0	1	1	1	…	1	…	0
	1							2

In the TIMSS Advanced 2008 analyses, 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 additional replicate weights where *h* was greater than the number of zones within the country were all set equal to the overall sampling weight. Although this involved some redundant computations, having 75 replicate weights for each country had no effect on the magnitude of the error variance computed using the jackknife formula, and it simplified the computation of standard errors for numerous countries at the same time. All standard errors presented in the TIMSS Advanced 2008 international report were computed using SAS programs developed at the TIMSS & PIRLS International Study Center.

8.6.4 Estimating the Imputation Variance

The TIMSS Advanced 2008 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.[19] The results for all of the booklets were then aggregated using item response theory to provide results for the entire assessment. Because each student responded to just a subset of the assessment items, it was necessary to use multiple imputation (the generation of plausible values) to derive reliable estimates of student performance on the assessment as a whole. Since every student's proficiency estimate incorporates some uncertainty arising from this imputation, TIMSS Advanced followed the customary procedure of generating five estimates for each student and using the variability among them as a measure of the imputation uncertainty, or error. In the TIMSS Advanced 2008 international report, the imputation error for each variable has been combined with the sampling error for that variable to provide a standard error that incorporates both.

The general procedure for estimating the imputation variance using plausible values is described in Mislevy, Beaton, Kaplan, and Sheenan (1992). First, compute the statistic t for each set of M plausible values. The statistics t_m, where $m = 1, 2, \ldots, M$, can be anything estimable from the data, such as a mean, the difference between means, percentiles, and so forth.

Once the statistics t_m are computed, the imputation variance of the statistic t is then calculated as:

$$Var_{imp}(t) = \left(1 + \frac{1}{M}\right) Var(t_1, \ldots t_M)$$

19 The TIMSS Advanced 2008 assessment design is described in the *TIMSS Advanced 2008 Assessment Frameworks* (Garden, et al., 2006).

where M is the number of plausible values used in the calculation, and $Var(t_1,\ldots,t_M)$ is the usual variance of the M estimates computed using each plausible value.

8.6.5 Combining the Sampling and Imputation Variance

The standard errors of all the proficiency statistics reported by TIMSS Advanced include both sampling and imputation variance components. These standard errors were computed using the following formula:

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

where $Var_{jrr}(t_1)$ is the sampling variance computed for the first plausible value[20] and $Var_{imp}(t)$ is the imputation variance. The *TIMSS Advanced 2008 User Guide for the International Database* (Foy & Arora, 2009) contains programs in SAS and SPSS that compute each of these variance components for the TIMSS Advanced 2008 data. Furthermore, the IEA IDB Analyzer—software provided with the international database—automatically computes standard errors as described in this section.

Exhibit 8.27 shows basic summary statistics for overall advanced mathematics and physics achievement in the TIMSS Advanced 2008 assessments. It presents the student sample size, the mean and standard deviation averaged across the five plausible values, the jackknife sampling error for the mean, and the overall standard error for the mean, which includes the imputation error.

20 Under ideal circumstances and with unlimited computing resources, the JRR sampling variance would be computed for each of the plausible values and the imputation variance as described here. This would require 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 sampling variance component using only one plausible value (the first one), and then the imputation variance using the five plausible values. Using this approach, a statistic needs to be computed only 80 times.

Exhibit 8.27 Summary Statistics and Standard Errors for Proficiency in TIMSS Advanced 2008

Advanced Mathematics

Country	Sample Size	Mean Proficiency	Standard Deviation	Jackknife Sampling Error	Overall Standard Error
Armenia	858	432.760	95.466	3.090	3.675
Iran, Islamic Rep. of	2,425	496.750	98.767	6.306	6.369
Italy	2,143	448.779	95.468	6.870	7.142
Lebanon	1,612	544.726	60.472	2.258	2.318
Netherlands	1,537	552.470	45.797	1.917	2.647
Norway	1,932	439.224	86.910	4.679	4.990
Philippines	4,091	355.189	105.545	5.374	5.522
Russian Federation	3,185	560.984	90.972	7.138	7.213
Slovenia	2,156	457.316	84.850	3.890	4.151
Sweden	2,303	412.806	103.265	5.370	5.571

Physics

Country	Sample Size	Mean Proficiency	Standard Deviation	Jackknife Sampling Error	Overall Standard Error
Armenia	894	495.067	100.284	5.125	5.363
Iran, Islamic Rep. of	2,434	459.856	115.728	7.145	7.204
Italy	1,861	422.238	102.558	7.510	7.624
Lebanon	1,595	443.542	78.324	2.615	2.990
Netherlands	1,511	582.474	53.723	3.547	3.703
Norway	1,640	534.142	78.147	4.182	4.212
Russian Federation	3,166	521.220	120.490	10.140	10.172
Slovenia	1,097	534.941	80.247	1.507	1.941
Sweden	2,291	496.950	91.865	5.509	5.651

References

Beaton, A.E. (1969). Criterion scaling of questionnaire items. *Socio-Economic Planning Sciences*, 2, 355–362.

Foy, P., & Arora, A. (2009). *TIMSS advanced 2008 user guide for the international database*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Foy, P., Galia, J., & Li, I. (2008). Scaling the data from the TIMSS 2007 mathematics and science assessments. Chapter 11 in J.F. Olson, M.O. Martin, and I.V.S. Mullis (Eds.), *TIMSS 2007 technical report* (pp. 225-280). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Garden, R.A., Lie, S., Robitaille, D.F., Angell, C., Martin, M.O., Mullis, I.V.S., Foy, P., & Arora, A. (2006), *TIMSS advanced 2008 assessment frameworks.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Johnson, E.G., & Rust, K.F. (1992). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics*, 17(2), 175–190.

Mislevy, R.J., Beaton, A., Kaplan, B.A., & Sheehan, K. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29(2), 133–161.

Mullis, I.V.S., Martin, M.O., Robitaille, D.F., & Foy, P. (2009), *TIMSS advanced 2008 international report: Findings from IEA's of achievement in advanced mathematics and physics in the final year of secondary school*, Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Muraki, E., & Bock, R.D. (1991). *PARSCALE: Parameter scaling of rating data* [Software Version 4.1]. Chicago, IL: Scientific Software, Inc.

Sheehan, K. M. (1985). *M-GROUP*: *Estimation of group effects in multivariate models* [Software Version 3.2]. Princeton, NJ: Educational Testing Service.

von Davier, M., Gonzalez, E., & Mislevy, R.J. (2009). What are plausible values and why are they useful? *IERI monograph series: Issues and methodologies in large-scale assessments*, (Vol. 2, pp. 9-36). Hamburg/Princeton NJ: IEA-ETS Research Institute.

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


OCCASVS
10
12

Appendix A

Organizations and Individuals Responsible for TIMSS Advanced 2008

Introduction

TIMSS Advanced 2008 was a collaborative effort involving many individuals around the world. This appendix recognizes the individuals and organizations for their contributions. Given the work on TIMSS Advanced 2008 has spanned approximately four years and has involved so many people and organizations, this list may not include all who contributed. Any omission is inadvertent.

Of the first importance, TIMSS Advanced 2008 is deeply indebted to the students, teachers, and school principals who contributed their time and effort to the study.

Management and Coordination

TIMSS Advanced 2008 was conducted by the TIMSS & PIRLS International Study Center at Boston College, which has responsibility for the overall direction and management of IEA's TIMSS and PIRLS projects. Headed by Ina V.S. Mullis and Michael O. Martin, the study center is located in the Lynch School of Education. In carrying out the project, the TIMSS & PIRLS International Study Center worked closely with the IEA Secretariat in Amsterdam, which provided guidance overall and was responsible for verification of all translations produced by the participating countries. The IEA Data Processing and Research Center in Hamburg was responsible for processing and verifying the internal consistency and accuracy of the data submitted by the participants. Statistics Canada in Ottawa was responsible for school and student sampling activities. Educational Testing Service (ETS) in Princeton, New Jersey provided psychometric methodology recommendations addressing calibration and scaling, and also made available software for scaling the achievement data.

The Project Management Team, comprised of the Directors and Senior Management from the TIMSS & PIRLS International Study Center, the IEA Secretariat, the IEA Data Processing and Research Center, Statistics Canada, and ETS, met twice a year throughout the study to discuss progress, procedures, and schedule. In addition, the Directors of the TIMSS & PIRLS International Study Center met with members of IEA's Technical Executive Group twice yearly to review technical issues.

Dr. Robert Garden from New Zealand was the TIMSS Advanced 2008 Mathematics Coordinator and Dr. Svein Lie, from the University of Oslo, was the TIMSS Advanced 2008 Physics Coordinator. Together with the Physics and Mathematics task force, a panel of internationally recognized experts in mathematics and physics research, curriculum, instruction, and assessment, they provided excellent guidance throughout TIMSS Advanced 2008.

To work with the international team and coordinate within-country activities, each participating country designated one or two individuals to be the TIMSS National Research Coordinator or Co-Coordinators, known as the NRCs. The NRCs had the complicated and challenging task of implementing the TIMSS Advanced 2008 study in their countries in accordance with TIMSS guidelines and procedures. The quality of the TIMSS Advanced 2008 assessment and data depends on the work of the NRCs and their colleagues in carrying out the very complex sampling, data collection, and scoring tasks involved.

Continuing the tradition of truly exemplary work established in other TIMSS assessments, the TIMSS Advanced 2008 NRCs performed their many tasks with dedication, competence, energy, and goodwill, and have been commended by the IEA Secretariat, the TIMSS & PIRLS International Study Center, the IEA Data Processing and Research Center, and Statistics Canada for their commitment to the project and the high quality of their work.

Funding

Funding for TIMSS Advanced 2008 was provided through a generous grant from the Norwegian Ministry of Education, fees from participating countries and through IEA's own resources. The financial support provided by Boston College is gratefully acknowledged.

IEA Secretariat

Hans Wagemaker, Executive Director
Barbara Malak, Manager Membership Relations
Juriaan Hartenberg, Financial Manager

TIMSS & PIRLS International Study Center at Boston College

Ina V.S. Mullis, Co-Director
Michael O. Martin, Co-Director
Pierre Foy, Director of Sampling and Data Analysis
Alka Arora, TIMSS Advanced 2008 Project Coordinator

Ryan Auster, TIMSS Graduate Assistant
Marcie Bligh, Manager of Office Administration
Susan Farrell, Co-Manager of Publications
Joseph Galia, Senior Statistician/Programmer
Christine Hoage, Manager of Finance
Jiefang Hu, TIMSS Graduate Assistant
Ieva Johansone, Survey Operations Coordinator
Isaac Li, Statistician/Programmer
Jennifer Moher, Data Graphics Specialist
Mario Pita, Co-Manager of Publications
Corinna Preuschoff, TIMSS Research Associate
Ruthanne Ryan, Data Graphics Specialist
Steven Simpson, Data Graphics Specialist

IEA Data Processing and Research Center

Dirk Hastedt, Co-Director
Milena Taneva, Manager, TIMSS Advanced Data Processing
Dirk Oehler, Researcher
Tim Daniel, Researcher
Alexander Konn, Programmer
Olaf Zuehlke, Researcher

Statistics Canada

Sylvie LaRoche, Senior Methodologist
Marc Joncas, Senior Methodologist

Educational Testing Service

Matthias Von Davier, Principal Research Scientist

Sampling Referee

Keith Rust, Vice President and Associate Director of the Statistical Group, Westat, Inc.

TIMSS Advanced 2008 Task Force

Robert Garden, TIMSS Advanced 2008 Mathematics Coordinator
Svein Lie, TIMSS Advanced 2008 Physics Coordinator
Carl Angell, Norway
Wolfgang Dietrich, Sweden
Liv Sissel Gronmo, Norway
Helen Lye, Australia
Torgeir Onstad, Norway
David Robitaille, Canada

TIMSS Advanced 2008 National Research Coordinators (NRCs)

Armenia

Arsen Baghdasaryan
Yerevan State University

Iran, Islamic Republic of

Abdol'azim Karimi
Ministry of Education
Research Institute for Education (RIE)

Italy

Anna Maria Caputo
Instituto Nazionale per la Valuatazione del Sistema Educativo di Istruzione e di Formazione

Lebanon

Leila Maliha Fayad
Center for Educational Research and Development
Ministry of Education and Higher Education

Netherlands

Marjolein Drent
Martina Meelissen
Centre for Applied Research in Education
University of Twente

Norway

Liv Sissel Grønmo
University of Oslo, ILS

Philippines

Ester B. Ogena
Science Education Institute
Department of Science and Technology

Merle C. Tan
National Institute for Science and Mathematics Education Development
University of Philippines

Russian Federation

Galina Kovaleva
Center for Evaluating the Quality of Secondary General Education
Institute of Content and Methods of Learning
Russian Academy of Education

Slovenia

Barbara Japelj Pavesic
Educational Research Institute

Sweden

Marie Eklund
National Agency for Education

Appendix B

Characteristics of National Samples

B.1 Armenia

A single school sample was used for both advanced mathematics and physics.

Advanced mathematics and physics

Coverage and Exclusions

- Coverage of the national desired target populations was 100%.
- There were no school-level exclusions prior to sampling.
- There were no within-sample exclusions.

Sample Design

- All schools with eligible students were selected for the data collection sample.
- There was no explicit or implicit stratification.
- Two classes were sampled per school, whenever possible.
- Half of the students in the selected classes were assigned randomly an advanced mathematics booklet; the other half were assigned a physics booklet.

Field Test Sample

- Six schools were sampled for the field test and used for both populations.
- Field test schools also were included in the data collection sample.

Notes on Sampling Weights

- The target population definitions in Armenia were finalized after the school sample had been selected. As a result, a smaller than expected sample of schools was available for analysis.
- In schools where all eligible classes were selected, classes were treated as strata and were randomly divided into two replicates for variance estimation.
- In schools where eligible classes were sampled, schools were treated as strata and classes as replicates for variance estimation.

Exhibit B.1 School Sample Allocation in Armenia

Advanced Mathematics and Physics

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Excluded Schools	Participating Schools			Non-Participating Schools
				Sampled	First Replacement	Second Replacement	
Armenia	38	0	0	38	0	0	0
Total	**38**	**0**	**0**	**38**	**0**	**0**	**0**

B.2 Islamic Republic of Iran

A single school sample was used for both advanced mathematics and physics.

Advanced mathematics and physics

Coverage and Exclusions

- Coverage of the national desired target populations was 100%.
- There were no school-level exclusions prior to sampling.
- There were no within-sample exclusions.

Sample Design

- Schools were stratified explicitly by school type (public, private) and gender (boys, girls), for a total of four explicit strata.
- Schools were stratified implicitly by province (31 provinces), for a total of 124 implicit strata.
- Because many sampled schools had been merged or closed, replacement of these ineligible schools was necessary—and approved by the sampling consultants—to prevent a significant drop in sample size.
- Two classes were sampled in schools with at least 90 eligible students; one class was sampled per school otherwise.
- Half of the students in the selected classes were assigned randomly an advanced mathematics booklet; the other half were assigned a physics booklet.

Field Test Sample

- Thirty schools were sampled for the field test at the same time as the data collection sample, thus no schools were selected for both activities.

Notes on Sampling Weights

- A school sampling weight adjustment was computed to account for the 10 ineligible schools that were replaced such that the school sampling weights of the 119 participating schools would represent only the 110 eligible schools.

Exhibit B.2 School Sample Allocation in Iran

Advanced Mathematics and Physics

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Excluded Schools	Participating Schools			Non-Participating Schools
				Sampled	First Replacement	Second Replacement	
Public schools for girls	44	3	0	41	3	0	0
Public schools for boys	48	3	0	45	3	0	0
Private schools for girls	10	1	0	9	1	0	0
Private schools for boys	18	3	0	14	3	0	1
Total	**120**	**10**	**0**	**109**	**10**	**0**	**1**

B.3 Italy

The sample design for Italy differed from the standard TIMSS Advanced design. First, a single sample of schools was selected for both study populations. Then, the sampled schools were allocated randomly to advanced mathematics, physics, or both.

Advanced mathematics and physics

Coverage and Exclusions

- Coverage of the national desired target populations was 100%.
- There were no school-level exclusions prior to sampling.
- Within-sample exclusions consisted of an excluded class, special needs students, and non-native language speakers, for exclusion rates of 0.53% for advanced mathematics and 0.95% for physics.
- Physics students were found in one school from the first explicit stratum, where only advanced mathematics students were expected. As a result, the reported physics exclusion rate is underestimated. Unfortunately, the sample design did not allow estimation of this specific type of exclusion.

Sample Design

- The measure of size for sample selection was total enrollment in the 13th grade.
- Schools were stratified explicitly by the presence of eligible classes from the two study populations:
  - Schools with an advanced mathematics program;
  - Schools with an advanced mathematics program and an advanced mathematics and physics program;
  - Schools with an advanced mathematics and physics program.

- Schools were stratified implicitly by region (20 regions), for a total of 56 implicit strata.
- After school sampling, the schools in the second and third explicit strata were allocated to the two study populations as follows:
  - All 82 schools in the second explicit stratum were allocated to physics, but only 42 of them were allocated randomly to advanced mathematics.
  - Of the 36 schools in the third explicit stratum, 6 schools were allocated randomly to advanced mathematics, and the remaining 30 schools were allocated to physics.
- Class sampling and booklet assignment varied across the explicit strata and from the random allocation of sampled schools to the study populations, as follows:
  - In the first explicit stratum, one advanced mathematics class was sampled per school and only the advanced mathematics booklets were administered.
  - In the 40 schools of the second explicit stratum that were allocated only to physics, two physics classes were sampled, whenever possible, and only the physics booklets were administered.
  - In the 42 schools of the second explicit stratum that were allocated to both populations, one class from the advanced mathematics program and one class from the advanced mathematics and physics program were sampled. In the class selected from the advanced mathematics program, the advanced mathematics booklets were administered; in the class selected from the advanced mathematics and physics program, half of the students were assigned randomly an advanced mathematics booklet and the other half were assigned a physics booklet.

- In the third explicit stratum, one class was sampled per school. The advanced mathematics booklets were administered in the 6 schools allocated to advanced mathematics and the physics booklets were administered in the 30 schools allocated to physics.

Field Test Sample

- A field test was administered in Italy, but the data were not available for international analysis. Schools in the field test were eligible for selection in the data collection sample.

Notes on Sampling Weights

- Many sampled schools were found to be ineligible as they did not have any advanced mathematics or physics students. In many other schools, the number of eligible students was smaller than expected. As a result, the estimated student population sizes are smaller than the population sizes estimated from the sampling frame.

Special Note

- In Italy, the booklet assignment did not follow strictly the rules in classes where both advanced mathematics and physics booklets were distributed, but booklets still were distributed in a random manner.

Exhibit B.3 School Sample Allocation in Italy

Advanced Mathematics

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Excluded Schools	Participating Schools			Non-Participating Schools
				Sampled	First Replacement	Second Replacement	
Schools with an advanced mathematics program	52	3	0	46	2	0	1
Schools with an advanced mathematics program and an advanced mathematics and physics program	42	4	0	38	0	0	0
Schools with an advanced mathematics and physics program	6	2	0	4	0	0	0
Total	**100**	**9**	**0**	**88**	**2**	**0**	**1**

Physics

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Excluded Schools	Participating Schools			Non-Participating Schools
				Sampled	First Replacement	Second Replacement	
Schools with an advanced mathematics program and an advanced mathematics and physics program	82	14	0	68	0	0	0
Schools with an advanced mathematics and physics program	30	7	0	23	0	0	0
Total	**112**	**21**	**0**	**91**	**0**	**0**	**0**

B.4 Lebanon

A single school sample was used for both advanced mathematics and physics.

Advanced mathematics and physics

Coverage and Exclusions

- Coverage of the national desired target populations was 100%.
- School-level exclusions prior to sampling consisted of 33 very small schools (less than 4 eligible students), for a total of 64 students and an exclusion rate of 1.25%.
- There were no within-sample exclusions.

Sample Design

- Schools were stratified explicitly by school size (large, small), for a total of two explicit strata.
- Schools were stratified implicitly by region (6 regions) and school type (public, private), for a total of 24 implicit strata.
- All schools from the large schools explicit stratum were selected; schools from the small schools explicit stratum were sampled with equal probabilities.
- All eligible classes in the selected schools were sampled.
- Half of the students in the selected classes were assigned randomly an advanced mathematics booklet; the other half were assigned a physics booklet.

Field Test Sample

- Lebanon did not carry out a field test.

Notes on Sampling Weights

- In all but three schools in the large schools explicit stratum, all classes were selected and classes were treated as strata and were randomly divided into two replicates for variance estimation. The other three schools were treated as strata and the classes as replicates for variance estimation.
- Two schools in the large schools explicit stratum did not meet the class-level student participation requirements for physics and consequently were treated as non-participating schools for physics.

Exhibit B.4 School Sample Allocation in Lebanon

Advanced Mathematics

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Excluded Schools	Participating Schools			Non-Participating Schools
				Sampled	First Replacement	Second Replacement	
Large schools	102	0	0	92	0	0	10
Small schools	138	0	0	111	9	0	18
Total	**240**	**0**	**0**	**203**	**9**	**0**	**28**

Physics

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Excluded Schools	Participating Schools			Non-Participating Schools
				Sampled	First Replacement	Second Replacement	
Large schools	102	0	0	90	0	0	12
Small schools	138	0	0	111	9	0	18
Total	**240**	**0**	**0**	**201**	**9**	**0**	**30**

B.5 Netherlands

Two separate school samples were selected for advanced mathematics and physics.

Advanced mathematics

Coverage and Exclusions

- Coverage of the national desired target population was 100%.
- School-level exclusions prior to sampling consisted of schools that had participated in the field test, for a total of 179 students and an exclusion rate of 2.5%.
- There were no within-sample exclusions.
- A small number of students were allowed to take their final examinations prior to the TIMSS Advanced assessments and thus were excluded from the advanced mathematics assessment because they were no longer in their advanced mathematics classes. The size of this exclusion could not be estimated.

Sample Design

- There was no explicit nor implicit stratification.
- Schools were sampled with equal probabilities.
- All eligible advanced mathematics classes in the selected schools were sampled.

Field Test Sample

- A convenience sample of six schools was selected for the field test and used for both populations. These schools were excluded from the sampling frame prior to selecting the schools for the data collection sample.

Physics

Coverage and Exclusions

- Coverage of the national desired target population was 100%.
- School-level exclusions prior to sampling consisted of the six schools that had participated in the field test, for a total of 179 students and an exclusion rate of 2.5%.
- Within-sample exclusions consisted of special needs students, for an exclusion rate of 0.2%.
- A small number of students were allowed to take their final examinations prior to the TIMSS Advanced assessments and were excluded from the physics assessment because they were no longer in their physics classes. The size of this exclusion could not be estimated.

Sample Design

- There was no explicit nor implicit stratification.
- Schools were sampled with equal probabilities.
- All eligible physics classes in the selected schools were sampled.

Field Test Sample

- A convenience sample of six schools was selected for the field test and used for both populations. These schools were excluded from the sampling frame prior to selecting the schools for the data collection sample.

Exhibit B.5 School Sample Allocation in the Netherlands

Advanced Mathematics

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Excluded Schools	Participating Schools			Non-Participating Schools
				Sampled	First Replacement	Second Replacement	
Netherlands	135	2	0	102	9	1	21
Total	**135**	**2**	**0**	**102**	**9**	**1**	**21**

Physics

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Excluded Schools	Participating Schools			Non-Participating Schools
				Sampled	First Replacement	Second Replacement	
Netherlands	135	2	0	98	16	1	18
Total	**135**	**2**	**0**	**98**	**16**	**1**	**18**

B.6 Norway

All eligible schools in Norway were selected for TIMSS Advanced 2008, but each school was selected for only one population, resulting in two separate school samples for advanced mathematics and physics.

Advanced Mathematics

Coverage and Exclusions

- Coverage of the national desired target population was 100%.
- School-level exclusions prior to sampling consisted of 15 very small schools, for a total of 66 students and an exclusion rate of 0.88%.
- Within-sample exclusions consisted of non-native language speakers, for an exclusion rate of 0.14%.

Sample Design

- Schools were stratified explicitly by the presence of eligible classes from the two study populations (advanced mathematics classes, advanced mathematics classes and physics classes) and the number of eligible physics students (less than 10, between 10 and 19, between 20 and 34, 35 or more), for a total of 5 explicit strata.
- There was no implicit stratification.
- All schools from the explicit stratum of advanced mathematics classes were selected.
- Schools were sampled with equal probabilities in the remaining explicit strata.
- Two advanced mathematics classes were sampled in schools with at least 60 eligible students; one advanced mathematics class was sampled per school otherwise.

Field Test Sample

- A sample of 25 schools was selected for the field test and used for both populations. Schools in the field test were eligible for selection in the data collection sample.

Physics

Coverage and Exclusions

- Coverage of the national desired target population was 100%.
- School-level exclusions prior to sampling consisted of 15 very small schools, for a total of 18 students and an exclusion rate of 0.4%.
- Within-sample exclusions consisted of special needs students, for an exclusion rate of 0.05%.

Sample Design

- Schools were stratified explicitly by the presence of eligible classes from the two study populations (physics classes, advanced mathematics classes and physics classes) and the number of eligible physics students (less than 10, between 10 and 19, between 20 and 34, 35 or more), for a total of five explicit strata.
- There was no implicit stratification.
- All schools from the explicit stratum of physics classes were selected.
- Schools were sampled with equal probabilities in the remaining explicit strata.
- All eligible physics classes in the selected schools were sampled.

Field Test Sample

- A sample of 25 schools was selected for the field test and used for both populations. Schools in the field test were eligible for selection in the data collection sample.

Exhibit B.6 School Sample Allocation in Norway

Advanced Mathematics

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Excluded Schools	Participating Schools			Non-Participating Schools
				Sampled	First Replacement	Second Replacement	
Schools with advanced mathematics classes	12	0	0	9	0	0	3
Schools with advanced mathematics classes and physics classes and 35 or more physics students	15	0	0	15	0	0	0
Schools with advanced mathematics classes and physics classes and 20 to 34 physics students	24	0	0	22	0	0	2
Schools with advanced mathematics classes and physics classes and 10 to 19 physics students	37	0	0	36	0	0	1
Schools with advanced mathematics classes and physics classes and less than 10 physics students	32	0	0	25	0	0	7
Total	**120**	**0**	**0**	**107**	**0**	**0**	**13**

Physics

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Excluded Schools	Participating Schools			Non-Participating Schools
				Sampled	First Replacement	Second Replacement	
Schools with physics classes	1	0	0	1	0	0	0
Schools with advanced mathematics classes and physics classes and 35 or more physics students	18	0	0	16	0	0	2
Schools with advanced mathematics classes and physics classes and 20 to 34 physics students	29	0	0	24	0	0	5
Schools with advanced mathematics classes and physics classes and 10 to 19 physics students	42	0	0	35	0	0	7
Schools with advanced mathematics classes and physics classes and less than 10 physics students	30	0	0	25	0	0	5
Total	**120**	**0**	**0**	**101**	**0**	**0**	**19**

B.7 Philippines

The Philippines participated in the advanced mathematics assessment only.

Advanced Mathematics

Coverage and Exclusions

- Coverage of the national desired target population was 100%.
- There were no school-level exclusions prior to sampling.
- There were no within-sample exclusions.

Sample Design

- Schools were stratified explicitly by their orientation (science and technology, other) and information about school size (size known, size unknown), for a total of four explicit strata.
- Schools were stratified implicitly by type (Philippines science high school system, regional science high schools, other public science high schools, other private high schools, university rural high schools and laboratory schools) in the explicit stratum of other schools, for a total of 10 implicit strata.
- Schools were sampled with equal probabilities in all explicit strata.
- One advanced mathematics class was sampled per school.

Field Test Sample

- The Philippines did not carry out a field test.

Notes on Sampling Weights

- From the original sample of 126 schools, a sub-sample of 123 schools was selected to reduce the sample size. The school sampling weights were adjusted to account for this sub-sampling.
- After the school sample was selected, two duplicate schools were found in the sample. The sampling weights were corrected to account for this duplication upon confirmation from the NRC that there were no other such cases on the sampling frame.
- Two schools, identified as outliers in terms of achievement, had their school weights set to 1 to stabilize variance estimation.

Exhibit B.7 School Sample Allocation in the Philippines

Advanced Mathematics

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Excluded Schools	Participating Schools			Non-Participating Schools
				Sampled	First Replacement	Second Replacement	
Science and technology schools of known size	53	0	0	53	0	0	0
Science and technology schools of unknown size	23	1	0	22	0	0	0
Other schools of known size	32	0	0	32	0	0	0
Other schools of unknown size	13	0	0	11	0	0	2
Total	**121**	**1**	**0**	**118**	**0**	**0**	**2**

B.8 Russian Federation

Two separate school samples were selected for advanced mathematics and physics.

Advanced Mathematics

Coverage and Exclusions

- Coverage of the national desired target population was 100%.
- There were no school-level exclusions prior to sampling.
- There were no within-sample exclusions.

Sample Design

- In a preliminary sampling stage, a sample of 43 regions out of 85 was selected with probabilities proportional to size. The 19 largest regions were selected with certainty.
- In the 19 certainty regions, schools were stratified explicitly by the presence of eligible classes from the two study populations (advanced mathematics classes, advanced mathematics classes and physics classes). They also were stratified implicitly by the 19 regions and 9 levels of urbanization.
- Each of the remaining sampled regions became an explicit stratum. Half of the schools that offered both advanced mathematics classes and physics classes were allocated randomly to the advanced mathematics assessment and added to the schools with only advanced mathematics classes prior to sampling (the other half was allocated to the physics assessment). These schools were stratified implicitly by seven levels of urbanization.
- One advanced mathematics class was sampled per school.

Field Test Sample

- A convenience sample of 23 schools from 3 regions was selected for the field test and used for both populations. Schools in the field test were eligible for selection in the data collection sample.

Notes on Sampling Weights

- School weights were adjusted to take into account the sampling of regions.
- In the sampled regions, the sampling weights of schools with both advanced mathematics classes and physics classes were adjusted to account for their random allocation to the advanced mathematics assessment.
- The sampled regions were treated as replicates for variance estimation.

Physics

Coverage and Exclusions

- Coverage of the national desired target population was 100%.
- There were no school-level exclusions prior to sampling.
- There were no within-sample exclusions.

Sample Design

- In a preliminary sampling stage, a sample of 43 regions out of 85 was selected with probabilities proportional to size. The 19 largest regions were selected with certainty.
- In the 19 certainty regions, schools were stratified explicitly by the presence of eligible classes from the two study populations (physics

classes, advanced mathematics classes and physics classes). They also were stratified implicitly by the 19 regions and 9 levels of urbanization.

- Each of the remaining sampled regions became an explicit stratum. Half of the schools that offered both advanced mathematics classes and physics classes were allocated randomly to the physics assessment and added to the schools with only physics classes prior to sampling (the other half was allocated to the advanced mathematics assessment). These schools were stratified implicitly by seven levels of urbanization.
- One physics class was sampled per school.

Field Test Sample

- A convenience sample of 23 schools from 3 regions was selected for the field test and used for both populations. Schools in the field test were eligible for selection in the data collection sample.

Notes on Sampling Weights

- School weights were adjusted to take into account the sampling of regions.
- In the sampled regions, the sampling weights of schools with both advanced mathematics classes and physics classes were adjusted to account for their random allocation to the physics assessment.
- The sampled regions were treated as replicates for variance estimation.

Exhibit B.8 School Sample Allocation in the Russian Federation

Advanced Mathematics

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Excluded Schools	Participating Schools			Non-Participating Schools
				Sampled	First Replacement	Second Replacement	
Schools in certainty regions with advanced mathematics classes	15	0	0	15	0	0	0
Schools in certainty regions with advanced mathematics classes and physics classes	50	0	0	50	0	0	0
Schools in sampled regions with advanced mathematics classes	78	0	0	78	0	0	0
Total	**143**	**0**	**0**	**143**	**0**	**0**	**0**

Physics

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Excluded Schools	Participating Schools			Non-Participating Schools
				Sampled	First Replacement	Second Replacement	
Schools in certainty regions with physics classes	42	0	0	42	0	0	0
Schools in certainty regions with advanced mathematics classes and physics classes	27	0	0	27	0	0	0
Schools in sampled regions with physics classes	80	0	0	80	0	0	0
Total	**149**	**0**	**0**	**149**	**0**	**0**	**0**

B.9 Slovenia

All schools in Slovenia with eligible students were selected for participation in TIMSS Advanced 2008. There were 87 schools with eligible students in advanced mathematics. There were 66 schools with eligible students in physics, which also were in the advanced mathematics sample, and all eligible physics students took part in the physics assessment. Since all 66 physics schools also were eligible for the advanced mathematics assessment, some students were selected for both assessments. The two assessments were scheduled on different days to accommodate this and a random mechanism determined which assessment was administered first in each school.

Advanced Mathematics

Coverage and Exclusions

- Coverage of the national desired target population was 100%.
- There were no school-level exclusions prior to sampling.
- Within-sample exclusions consisted of one excluded school—a Waldorf school—and special needs students, for an exclusion rate of 1.3%.

Sample Design

- All schools with eligible students were selected.
- Schools were stratified explicitly by the presence of eligible classes from the two study populations (advanced mathematics classes, advanced mathematics classes and physics classes) and the number of mathematics experts (many, few), for a total of three explicit strata. Schools with "many experts" were defined as those schools in which a high proportion of students—25 percent or more—registered for an advanced-level test in their final-year examinations during the 2007/2008 school year.

- There was no implicit stratification.
- Two advanced mathematics classes were sampled per school in the explicit stratum of advanced mathematics classes and physics classes with many experts; one advanced mathematics class was sampled per school otherwise.

Field Test Sample

- A sample of 22 schools with advanced mathematics classes and physics classes was selected and used for both populations. Schools in the field test also were part of the data collection sample.

Notes on Sampling Weights

- In schools where classes were sampled, schools were treated as strata and classes as replicates for variance estimation. If only one class was sampled in a school, the selected class was randomly divided into two replicates for variance estimation.
- In schools where all eligible classes were selected, classes were treated as strata and were randomly divided into two replicates for variance estimation.

Physics

Coverage and Exclusions

- Coverage of the national desired target population was 100%.
- There were no school-level exclusions prior to sampling.
- Within-sample exclusions consisted of one excluded school—a Waldorf school—and special needs students, for an exclusion rate of 0.5%.

Sample Design

- All schools with eligible students were selected.
- Schools were stratified explicitly by the number of mathematics experts (many, few), for a total of two explicit strata.
- There was no implicit stratification.
- All eligible physics classes in the selected schools were sampled.

Field Test Sample

- A sample of 22 schools with advanced mathematics classes and physics classes was selected and used for both populations. Schools in the field test also were part of the data collection sample.

Notes on Sampling Weights

- In all participating schools, classes were treated as strata and were randomly divided into two replicates for variance estimation.

Exhibit B.9 School Sample Allocation in Slovenia

Advanced Mathematics

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Excluded Schools	Participating Schools			Non-Participating Schools
				Sampled	First Replacement	Second Replacement	
Schools with advanced mathematics classes and physics classes and many experts	15	0	0	15	0	0	0
Schools with advanced mathematics classes and physics classes and few experts	51	1	1	48	0	0	1
Schools with advanced mathematics classes	21	3	0	16	0	0	2
Total	**87**	**4**	**1**	**79**	**0**	**0**	**3**

Physics

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Excluded Schools	Participating Schools			Non-Participating Schools
				Sampled	First Replacement	Second Replacement	
Schools with advanced mathematics classes and physics classes and many experts	15	0	0	13	0	0	2
Schools with advanced mathematics classes and physics classes and few experts	51	1	1	41	0	0	8
Total	**66**	**1**	**1**	**54**	**0**	**0**	**10**

B.10 Sweden

All eligible schools in Sweden were selected for TIMSS Advanced 2008, but each school was selected for only one population, resulting in two separate school samples for advanced mathematics and physics.

Advanced Mathematics

Coverage and Exclusions

- Coverage of the national desired target population was 100%.
- School-level exclusions prior to sampling consisted of 25 very small schools (less than 7 students from the natural science program or less then 5 students from the technology program) and 2 additional small schools, for a total of 177 students and an exclusion rate of 1.5%.
- Within-sample exclusions consisted of one excluded school—with an alternate curriculum—and special needs students, for an exclusion rate of 0.3%.

Sample Design

- Schools were stratified explicitly by the presence of eligible classes from the two study populations (advanced mathematics classes, advanced mathematics classes and physics classes), the presence of the two eligible programs (natural science, technology, both programs) in the stratum of schools with advanced mathematics classes and physics classes, and school type (public, private), for a total of six explicit strata. Private schools offered only the natural science program.
- There was no implicit stratification.

- All schools in the explicit stratum of advanced mathematics classes were selected.
- Schools were sampled with equal probabilities in the remaining explicit strata.
- In the explicit stratum of schools with advanced mathematics classes and physics classes and both programs, the sampled schools provided classes from only one of the two programs through a random mechanism.
- Since it was not always possible to identify all eligible advanced mathematics students from the final-year advanced mathematics classes (in some schools, students took the advanced mathematics course in the 11^{th} grade or at the beginning of their final year), they were identified through home classes in their final year of secondary school (12^{th} grade). A home class from the natural science program consisted entirely of eligible students, whereas a home class from the technology program included ineligible students that were removed from the advanced mathematics assessment.
- Two advanced mathematics classes were sampled in schools with at least 100 eligible students; one advanced mathematics class was sampled per school otherwise. Occasionally, two advanced mathematics classes were sampled in additional schools to increase the student sample size.

Field Test Sample

- A sample of 25 schools was selected for the field test and used for both populations. Field test schools were eligible as replacement schools in the data collection sample.

Notes on Sampling Weights

- In the explicit stratum of schools with advanced mathematics classes and physics classes and both programs, the school sampling weights were adjusted to account for the random mechanism that determined which program provided eligible classes for sampling.

Physics

Coverage and Exclusions

- Coverage of the national desired target population was 100%.
- School-level exclusions prior to sampling consisted of 23 very small schools (less than seven students from the natural science program or less then five students from the technology program), for a total of 218 students and an exclusion rate of 2.15%.
- Within-sample exclusions consisted of special students, for an exclusion rate of 0.16%.

Sample Design

- Schools were stratified explicitly by the presence of the two eligible programs (natural science, technology, both programs) and school type (public, private), for a total of four explicit strata. Private schools offered only the natural science program.
- There was no implicit stratification.
- Schools were sampled with equal probabilities.
- In the explicit stratum of schools with advanced mathematics classes and physics classes and both programs, the sampled schools provided classes from only one of the two programs through a random mechanism.

- Since it was not always possible to identify all eligible physics students from the final-year physics classes (in some schools, students took the physics course at the beginning of their final year), they were identified through home classes in their final year of secondary school (12^{th} grade). A home class from the natural science program consisted entirely of eligible students, whereas a home class from the technology program included ineligible students that were removed from the physics assessment.
- Two physics classes were sampled in schools with at least 75 eligible students; one physics class was sampled per school otherwise. Occasionally, two physics classes were sampled in additional schools to increase the student sample size.

Field Test Sample

- A sample of 25 schools was selected for the field test and used for both populations. Field test schools were eligible as replacement schools in the data collection sample.

Notes on Sampling Weights

- In the explicit stratum of schools with advanced mathematics classes and physics classes and both programs, the school sampling weights were adjusted to account for the random mechanism that determined which program provided eligible classes for sampling.

Exhibit B.10 School Sample Allocation in Sweden

Advanced Mathematics

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Excluded Schools	Participating Schools			Non-Participating Schools
				Sampled	First Replacement	Second Replacement	
Public schools with advanced mathematics classes from the natural science program	6	0	0	6	0	0	0
Private schools with advanced mathematics classes from the natural science program	5	0	0	3	0	0	2
Schools with advanced mathematics classes and physics classes from both programs	52	0	0	47	4	0	1
Public schools with advanced mathematics classes and physics classes from the natural science program	38	0	0	35	1	0	2
Private schools with advanced mathematics classes and physics classes from the natural science program	20	0	1	14	0	0	5
Schools with advanced mathematics classes and physics classes from the technology program	6	0	0	6	0	0	0
Total	**127**	**0**	**1**	**111**	**5**	**0**	**10**

Physics

Explicit Stratum	Total Sampled Schools	Ineligible Schools	Excluded Schools	Participating Schools			Non-Participating Schools
				Sampled	First Replacement	Second Replacement	
Schools with advanced mathematics classes and physics classes from both programs	63	1	0	60	0	0	2
Public schools with advanced mathematics classes and physics classes from the natural science program	36	0	0	35	1	0	0
Private schools with advanced mathematics classes and physics classes from the natural science program	20	0	0	18	0	0	2
Schools with advanced mathematics classes and physics classes from the technology program	8	1	0	6	1	0	0
Total	**127**	**2**	**0**	**119**	**2**	**0**	**4**

Appendix C

Item Parameters for IRT Analyses of TIMSS Advanced 2008 Data

Exhibit C.1 IRT Parameters for the TIMSS Advanced 1995 Re-scaling of Advanced Mathematics

Item		Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
MA13001	I01	1.30 (0.05)	–0.10 (0.02)	0.10 (0.01)			
MA13002	I02	1.44 (0.06)	0.01 (0.03)	0.18 (0.01)			
MA13003	I03	0.42 (0.02)	–0.51 (0.11)	0.07 (0.03)			
MA13004	I04	1.24 (0.06)	0.20 (0.03)	0.27 (0.02)			
MA13006	I06	0.90 (0.04)	0.12 (0.03)	0.07 (0.01)			
MA13007	I07	0.80 (0.04)	0.15 (0.05)	0.20 (0.02)			
MA13008	I08	0.69 (0.04)	–0.95 (0.09)	0.19 (0.03)			
MA13009	I09	0.61 (0.03)	–0.39 (0.06)	0.09 (0.02)			
MA13011	J01	1.18 (0.09)	0.02 (0.05)	0.15 (0.03)			
MA13012	J02	1.39 (0.17)	0.88 (0.05)	0.20 (0.02)			
MA13013	J03	0.98 (0.08)	–0.07 (0.07)	0.15 (0.03)			
MA13015	J05	1.39 (0.10)	0.00 (0.05)	0.19 (0.03)			
MA13016	J06	0.84 (0.11)	1.08 (0.08)	0.15 (0.02)			
MA13017	J07	0.91 (0.08)	0.52 (0.06)	0.12 (0.03)			
MA13018	J08	0.52 (0.08)	0.03 (0.26)	0.32 (0.07)			
MA13019	J09	1.07 (0.17)	1.59 (0.10)	0.15 (0.02)			
MA13020	J10	1.87 (0.24)	1.01 (0.04)	0.26 (0.02)			
MA13021	J11	0.39 (0.04)	–0.83 (0.12)	0.20 (0.00)			
MA13024	J14	1.33 (0.11)	0.27 (0.05)	0.20 (0.03)			
MA13025A	J15A	0.77 (0.04)	0.06 (0.04)				
MA13025B	J15B	0.98 (0.08)	1.96 (0.11)				
MA13026A	J16A	0.49 (0.04)	–0.84 (0.08)				
MA13026B	J16B	0.78 (0.05)	1.21 (0.07)				
MA13027	J17	0.44 (0.02)	0.75 (0.04)		–1.89 (0.11)	1.89 (0.12)	
MA13028	J18	0.99 (0.07)	1.17 (0.06)				
MA13029	J19	0.43 (0.02)	0.17 (0.04)		–1.53 (0.11)	1.53 (0.11)	
MA13064	K01	1.14 (0.09)	–0.98 (0.09)	0.23 (0.04)			
MA13065	K02	0.99 (0.11)	1.17 (0.06)	0.09 (0.02)			
MA13066	K03	1.08 (0.15)	0.50 (0.09)	0.45 (0.03)			
MA13067	K04	1.03 (0.12)	1.10 (0.06)	0.10 (0.02)			
MA13068	K05	2.25 (0.27)	0.87 (0.04)	0.31 (0.02)			
MA13069	K06	1.12 (0.11)	0.32 (0.07)	0.25 (0.03)			
MA13070	K07	0.49 (0.08)	0.43 (0.19)	0.22 (0.05)			
MA13073	K10	1.01 (0.16)	1.66 (0.11)	0.12 (0.02)			
MA13074	K11	0.85 (0.10)	0.56 (0.08)	0.20 (0.03)			
MA13075	K12	0.45 (0.03)	–0.01 (0.06)				
MA13076	K13	0.49 (0.04)	1.42 (0.12)				
MA13077	K14	1.03 (0.08)	1.58 (0.09)				
MA13078	K15	0.66 (0.03)	1.05 (0.04)		–1.66 (0.10)	1.66 (0.11)	
MA13079	K16	0.30 (0.01)	–0.16 (0.04)		–1.09 (0.15)	1.07 (0.15)	0.02 (0.12)
MA13080	K17	0.56 (0.02)	0.57 (0.02)		–1.93 (0.11)	0.11 (0.15)	1.82 (0.12)

Exhibit C.1 IRT Parameters for the TIMSS Advanced 1995 Re-scaling of Advanced Mathematics (Continued)

Item		Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
MA13081	K18	0.38 (0.02)	0.19 (0.05)		–1.28 (0.11)	1.28 (0.11)	
MA13082	L01	1.06 (0.11)	–0.14 (0.10)	0.37 (0.04)			
MA13083	L02	1.10 (0.09)	0.03 (0.07)	0.22 (0.03)			
MA13084	L03	1.15 (0.10)	0.59 (0.05)	0.14 (0.02)			
MA13085	L04	0.80 (0.09)	0.56 (0.08)	0.18 (0.03)			
MA13086	L05	1.29 (0.08)	0.32 (0.04)	0.06 (0.02)			
MA13087	L06	1.04 (0.09)	0.86 (0.05)	0.08 (0.02)			
MA13088	L07	0.81 (0.10)	1.17 (0.08)	0.14 (0.03)			
MA13089	L08	0.76 (0.07)	0.43 (0.08)	0.14 (0.03)			
MA13090	L09	0.69 (0.08)	0.41 (0.12)	0.23 (0.04)			
MA13093	L12	0.53 (0.07)	–0.13 (0.22)	0.26 (0.06)			
MA13094	L13	0.92 (0.05)	0.89 (0.05)				
MA13095	L14	0.68 (0.04)	0.04 (0.04)				
MA13097	L16	0.45 (0.02)	0.41 (0.03)		–0.50 (0.09)	–0.19 (0.11)	0.69 (0.10)
MA13098	L17	0.73 (0.03)	0.95 (0.03)		–1.39 (0.09)	1.39 (0.09)	
MA13099	L18	0.60 (0.04)	–0.25 (0.06)				

Exhibit C.2 IRT Parameters for the TIMSS Advanced 1995 Re-scaling of Physics

Item		Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})
PA13001	E01	0.35 (0.04)	–0.81 (0.47)	0.36 (0.08)		
PA13002	E02	0.91 (0.06)	0.47 (0.05)	0.33 (0.02)		
PA13003	E03	0.81 (0.04)	–0.43 (0.07)	0.18 (0.03)		
PA13004	E04	0.62 (0.04)	–1.63 (0.19)	0.25 (0.07)		
PA13005	E05	0.46 (0.04)	–1.28 (0.28)	0.28 (0.08)		
PA13006	E06	1.42 (0.14)	1.67 (0.04)	0.30 (0.01)		
PA13009	E09	1.01 (0.07)	1.26 (0.04)	0.21 (0.01)		
PA13011	F01	0.75 (0.17)	1.54 (0.14)	0.39 (0.03)		
PA13012	F02	1.60 (0.18)	1.46 (0.05)	0.09 (0.01)		
PA13013	F03	0.96 (0.12)	1.13 (0.07)	0.22 (0.02)		
PA13014	F04	1.23 (0.11)	0.40 (0.05)	0.20 (0.02)		
PA13015	F05	0.59 (0.05)	–0.08 (0.06)	0.25 (0.00)		
PA13016	F06	1.05 (0.10)	1.05 (0.05)	0.10 (0.02)		
PA13017	F07	0.27 (0.06)	0.74 (0.51)	0.23 (0.09)		
PA13018	F08	0.65 (0.12)	1.25 (0.13)	0.28 (0.04)		
PA13019	F09	0.81 (0.17)	2.03 (0.16)	0.19 (0.02)		
PA13021	F11	1.08 (0.11)	0.85 (0.06)	0.16 (0.02)		
PA13022	F12	0.55 (0.02)	1.44 (0.05)		–1.10 (0.09)	1.10 (0.11)
PA13023	F13	0.40 (0.03)	–0.43 (0.08)			
PA13024	F14	0.62 (0.03)	1.03 (0.04)		–0.51 (0.07)	0.51 (0.08)
PA13025	F15	0.78 (0.03)	1.17 (0.04)		–0.74 (0.07)	0.74 (0.08)
PA13026	F16	0.91 (0.07)	1.72 (0.08)			
PA13027A	F17A	0.58 (0.04)	0.62 (0.06)			
PA13049	G01	1.05 (0.15)	1.31 (0.08)	0.28 (0.02)		
PA13050	G02	0.99 (0.08)	–0.10 (0.09)	0.20 (0.04)		
PA13051	G03	1.75 (0.38)	1.75 (0.09)	0.38 (0.01)		
PA13053	G05	0.70 (0.12)	1.38 (0.11)	0.21 (0.03)		
PA13054	G06	0.90 (0.11)	0.51 (0.10)	0.32 (0.03)		
PA13055	G07	0.90 (0.13)	1.52 (0.09)	0.16 (0.02)		
PA13056	G08	0.98 (0.11)	1.10 (0.07)	0.17 (0.02)		
PA13057	G09	0.91 (0.18)	2.12 (0.16)	0.13 (0.02)		
PA13058	G10	1.32 (0.17)	1.37 (0.06)	0.21 (0.02)		
PA13059	G11	0.41 (0.04)	2.79 (0.27)			
PA13060	G12	0.48 (0.02)	0.35 (0.04)		–1.03 (0.08)	1.03 (0.09)
PA13062	G14	1.19 (0.06)	0.84 (0.03)			
PA13063	G15	1.14 (0.07)	1.32 (0.05)			
PA13064	G16	0.48 (0.03)	–1.14 (0.09)			
PA13065	G17	0.38 (0.03)	1.46 (0.13)			
PA13066	G18	0.52 (0.03)	1.57 (0.07)		–0.42 (0.08)	0.42 (0.11)
PA13067	G19	0.62 (0.02)	1.20 (0.04)		–1.05 (0.08)	1.05 (0.10)
PA13068	H01	0.98 (0.32)	2.32 (0.25)	0.37 (0.02)		

Exhibit C.2 IRT Parameters for the TIMSS Advanced 1995 Re-scaling of Physics (Continued)

Item		Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})
PA13069	H02	0.55 (0.10)	0.83 (0.19)	0.28 (0.05)		
PA13070	H03	1.12 (0.13)	1.09 (0.06)	0.21 (0.02)		
PA13071	H04	1.38 (0.16)	1.20 (0.05)	0.19 (0.02)		
PA13072	H05	0.66 (0.10)	1.06 (0.12)	0.24 (0.04)		
PA13073	H06	1.44 (0.28)	1.86 (0.10)	0.26 (0.01)		
PA13074	H07	0.84 (0.09)	0.82 (0.07)	0.13 (0.03)		
PA13075	H08	0.99 (0.10)	1.08 (0.06)	0.11 (0.02)		
PA13076	H09	0.86 (0.09)	1.32 (0.07)	0.08 (0.02)		
PA13077	H10	1.34 (0.12)	1.00 (0.04)	0.12 (0.01)		
PA13079	H12	0.93 (0.05)	1.02 (0.05)			
PA13080	H13	1.22 (0.06)	0.58 (0.03)			
PA13081	H14	0.62 (0.05)	2.11 (0.14)			
PA13083	H16	0.46 (0.02)	1.08 (0.05)		–2.14 (0.13)	2.14 (0.14)
PA13084	H17	0.47 (0.02)	1.38 (0.06)		–2.39 (0.15)	2.39 (0.16)
PA13085	H18	0.40 (0.03)	1.21 (0.07)		0.17 (0.08)	–0.17 (0.11)
PA13086A	H19A	0.57 (0.03)	0.76 (0.04)		–0.02 (0.07)	0.02 (0.08)
PA13086B	H19B	0.50 (0.04)	–0.59 (0.09)			

Exhibit C.3 IRT Parameters for the TIMSS Advanced 1995–2008 Concurrent Scaling of Advanced Mathematics

Item		Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
MA13064	K01	1.15 (0.13)	–0.92 (0.13)	0.20 (0.06)			
MA13065	K02	0.98 (0.14)	1.26 (0.08)	0.09 (0.02)			
MA13066	K03	1.05 (0.20)	0.57 (0.12)	0.44 (0.04)			
MA13067	K04	1.01 (0.15)	1.19 (0.07)	0.10 (0.02)			
MA13068	K05	2.20 (0.43)	0.96 (0.05)	0.31 (0.02)			
MA13069	K06	1.10 (0.15)	0.39 (0.09)	0.24 (0.04)			
MA13070	K07	0.49 (0.09)	0.48 (0.23)	0.21 (0.07)			
MA13073	K10	0.99 (0.22)	1.76 (0.15)	0.12 (0.02)			
MA13074	K11	0.84 (0.12)	0.63 (0.10)	0.20 (0.04)			
MA13075	K12	0.47 (0.04)	0.08 (0.06)				
MA13076	K13	0.50 (0.05)	1.49 (0.13)				
MA13077	K14	1.02 (0.10)	1.67 (0.10)				
MA13078	K15	0.65 (0.03)	1.14 (0.04)		–1.69 (0.10)	1.69 (0.11)	
MA13079	K16	0.31 (0.02)	–0.07 (0.05)		–1.06 (0.15)	1.05 (0.15)	0.01 (0.11)
MA13080	K17	0.55 (0.02)	0.65 (0.03)		–1.97 (0.11)	0.12 (0.16)	1.85 (0.12)
MA13081	K18	0.38 (0.02)	0.27 (0.05)		–1.26 (0.11)	1.26 (0.11)	
MA13082	L01	1.01 (0.14)	–0.11 (0.14)	0.35 (0.06)			
MA13083	L02	1.08 (0.13)	0.09 (0.10)	0.21 (0.05)			
MA13084	L03	1.13 (0.14)	0.67 (0.06)	0.14 (0.03)			
MA13085	L04	0.78 (0.11)	0.63 (0.11)	0.17 (0.04)			
MA13086	L05	1.27 (0.11)	0.39 (0.04)	0.06 (0.02)			
MA13087	L06	1.02 (0.12)	0.95 (0.06)	0.08 (0.02)			
MA13088	L07	0.79 (0.13)	1.25 (0.10)	0.14 (0.03)			
MA13089	L08	0.74 (0.09)	0.50 (0.10)	0.13 (0.04)			
MA13090	L09	0.68 (0.10)	0.47 (0.15)	0.22 (0.05)			
MA13093	L12	0.52 (0.08)	–0.11 (0.28)	0.24 (0.08)			
MA13094	L13	0.92 (0.06)	0.98 (0.05)				
MA13095	L14	0.68 (0.05)	0.13 (0.05)				
MA13097	L16	0.45 (0.02)	0.49 (0.03)		–0.51 (0.09)	–0.19 (0.11)	0.69 (0.10)
MA13098	L17	0.71 (0.03)	1.04 (0.04)		–1.42 (0.09)	1.42 (0.10)	
MA13099	L18	0.62 (0.05)	–0.16 (0.06)				
MA13001	M1_01	1.18 (0.05)	–0.10 (0.04)	0.12 (0.02)			
MA13002	M1_02	1.47 (0.07)	0.11 (0.03)	0.21 (0.02)			
MA13003	M1_03	0.44 (0.03)	–0.30 (0.10)	0.07 (0.03)			
MA13004	M1_04	1.16 (0.07)	0.25 (0.04)	0.27 (0.02)			
MA13006	M1_06	0.88 (0.04)	0.21 (0.04)	0.07 (0.02)			
MA13007	M1_07	0.83 (0.05)	0.22 (0.06)	0.18 (0.02)			
MA13008	M1_08	0.79 (0.04)	–0.74 (0.09)	0.16 (0.04)			
MA13009	M1_09	0.63 (0.03)	–0.17 (0.08)	0.09 (0.03)			
MA13011	M2_01	0.96 (0.08)	0.05 (0.07)	0.15 (0.03)			
MA13012	M2_02	1.29 (0.13)	0.94 (0.04)	0.18 (0.02)			

Exhibit C.3 IRT Parameters for the TIMSS Advanced 1995–2008 Concurrent Scaling of Advanced Mathematics (Continued)

Item		Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
MA13013	M2_03	1.03 (0.08)	–0.02 (0.07)	0.18 (0.03)			
MA13015	M2_05	1.02 (0.08)	0.08 (0.07)	0.23 (0.03)			
MA13016	M2_06	0.69 (0.09)	1.26 (0.07)	0.13 (0.02)			
MA13017	M2_07	0.85 (0.07)	0.64 (0.05)	0.11 (0.02)			
MA13018	M2_08	0.44 (0.07)	–0.16 (0.35)	0.29 (0.08)			
MA13019	M2_09	1.09 (0.14)	1.63 (0.08)	0.16 (0.01)			
MA13020	M2_10	1.92 (0.24)	1.13 (0.04)	0.26 (0.01)			
MA13021	M3_01	0.36 (0.03)	–0.67 (0.08)	0.20 (0.00)			
MA13024	M3_04	1.36 (0.12)	0.45 (0.04)	0.25 (0.02)			
MA13025A	M3_05A	0.67 (0.03)	0.10 (0.03)				
MA13025B	M3_05B	0.93 (0.06)	1.91 (0.08)				
MA13026A	M3_06A	0.63 (0.03)	–0.43 (0.04)				
MA13026B	M3_06B	0.87 (0.04)	1.27 (0.05)				
MA13027	M3_07	0.49 (0.01)	0.74 (0.03)		–1.97 (0.08)	1.97 (0.08)	
MA13028	M3_08	0.71 (0.04)	1.62 (0.08)				
MA13029	M3_09	0.42 (0.01)	0.34 (0.03)		–1.73 (0.08)	1.73 (0.08)	
MA23005	M4_01	0.49 (0.08)	–0.11 (0.29)	0.10 (0.09)			
MA23145	M4_02	0.58 (0.04)	0.57 (0.06)				
MA23187	M4_03	0.32 (0.01)	–0.21 (0.05)		–1.87 (0.12)	1.87 (0.11)	
MA23201	M4_04	0.83 (0.05)	0.04 (0.04)				
MA23154	M4_05	0.49 (0.02)	–0.21 (0.03)		–1.16 (0.08)	1.16 (0.08)	
MA23206	M4_06	1.44 (0.13)	0.04 (0.05)	0.21 (0.03)			
MA23166	M4_07	1.16 (0.08)	1.43 (0.07)				
MA23043	M4_08	0.66 (0.02)	0.80 (0.04)		–0.95 (0.07)	0.95 (0.07)	
MA23076	M4_09	0.81 (0.09)	0.14 (0.10)	0.15 (0.04)			
MA23176	M4_10	0.76 (0.12)	0.41 (0.12)	0.31 (0.04)			
MA23098	M4_11	1.19 (0.19)	0.95 (0.07)	0.30 (0.02)			
MA23144	M5_01	1.05 (0.14)	0.68 (0.06)	0.18 (0.03)			
MA23185	M5_02	0.82 (0.12)	0.59 (0.10)	0.22 (0.04)			
MA23054	M5_03	0.60 (0.02)	0.68 (0.04)		–0.72 (0.06)	0.72 (0.07)	
MA23064	M5_04	0.85 (0.14)	1.19 (0.09)	0.17 (0.03)			
MA23131A	M5_05A	1.05 (0.06)	0.34 (0.03)				
MA23131B	M5_05B	1.01 (0.07)	1.18 (0.06)				
MA23157	M5_06	0.73 (0.04)	0.80 (0.04)		0.56 (0.04)	–0.56 (0.06)	
MA23045	M5_07	1.12 (0.16)	1.24 (0.07)	0.15 (0.02)			
MA23082	M5_08	0.98 (0.16)	0.41 (0.11)	0.38 (0.04)			
MA23020	M5_09	0.96 (0.16)	0.93 (0.08)	0.29 (0.03)			
MA23094	M5_10	0.48 (0.03)	1.33 (0.07)		–0.44 (0.07)	0.44 (0.10)	
MA23069	M6_01	1.02 (0.15)	0.31 (0.08)	0.13 (0.04)			
MA23135	M6_02	0.66 (0.06)	0.01 (0.06)				
MA23208	M6_03	1.04 (0.19)	0.75 (0.09)	0.20 (0.03)			

Exhibit C.3 IRT Parameters for the TIMSS Advanced 1995–2008 Concurrent Scaling of Advanced Mathematics (Continued)

Item		Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
MA23165	M6_04	0.58 (0.05)	0.34 (0.08)				
MA23039	M6_05	0.68 (0.10)	–0.56 (0.22)	0.02 (0.09)			
MA23159	M6_06	0.71 (0.06)	–0.39 (0.06)				
MA23198	M6_07	1.37 (0.11)	0.60 (0.04)				
MA23042	M6_08	1.21 (0.24)	0.53 (0.09)	0.35 (0.03)			
MA23055	M6_09	0.93 (0.13)	0.35 (0.09)	0.10 (0.04)			
MA23080	M6_10	1.24 (0.15)	0.19 (0.07)	0.12 (0.03)			
MA23021	M6_11	2.09 (0.43)	1.06 (0.06)	0.16 (0.02)			
MA23004	M7_01	0.86 (0.21)	0.71 (0.15)	0.29 (0.05)			
MA23063	M7_02	1.32 (0.26)	1.26 (0.09)	0.14 (0.02)			
MA23141	M7_03	1.13 (0.10)	0.80 (0.05)				
MA23133	M7_04	0.91 (0.17)	1.00 (0.09)	0.08 (0.03)			
MA23158	M7_05	1.51 (0.26)	1.11 (0.07)	0.11 (0.02)			
MA23151	M7_06	0.89 (0.18)	0.61 (0.13)	0.23 (0.05)			
MA23035A	M7_07A	1.14 (0.09)	0.34 (0.04)				
MA23035B	M7_07B	1.03 (0.10)	1.06 (0.07)				
MA23050	M7_08	1.13 (0.29)	1.28 (0.12)	0.23 (0.03)			
MA23041	M7_09	0.96 (0.21)	0.89 (0.11)	0.20 (0.04)			
MA23182	M7_10	1.27 (0.22)	0.50 (0.08)	0.20 (0.04)			
MA23170	M7_11	0.98 (0.08)	0.66 (0.06)				

Exhibit C.4 IRT Parameters for the TIMSS Advanced 1995–2008 Concurrent Scaling of Physics

Item		Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})
PA13049	G01	1.02 (0.21)	1.30 (0.10)	0.27 (0.03)		
PA13050	G02	0.99 (0.11)	–0.12 (0.12)	0.18 (0.06)		
PA13051	G03	1.70 (0.59)	1.76 (0.12)	0.38 (0.02)		
PA13053	G05	0.70 (0.16)	1.37 (0.13)	0.21 (0.04)		
PA13054	G06	0.88 (0.15)	0.47 (0.15)	0.30 (0.05)		
PA13055	G07	0.89 (0.17)	1.53 (0.10)	0.16 (0.03)		
PA13056	G08	0.96 (0.15)	1.09 (0.08)	0.16 (0.03)		
PA13057	G09	0.90 (0.23)	2.13 (0.20)	0.13 (0.02)		
PA13058	G10	1.30 (0.25)	1.37 (0.08)	0.21 (0.02)		
PA13059	G11	0.42 (0.05)	2.72 (0.29)			
PA13060	G12	0.50 (0.02)	0.36 (0.04)		–1.00 (0.08)	1.00 (0.08)
PA13062	G14	1.21 (0.08)	0.84 (0.04)			
PA13063	G15	1.15 (0.09)	1.32 (0.06)			
PA13064	G16	0.50 (0.05)	–1.05 (0.11)			
PA13065	G17	0.39 (0.04)	1.42 (0.15)			
PA13066	G18	0.52 (0.03)	1.57 (0.08)		–0.42 (0.08)	0.42 (0.11)
PA13067	G19	0.63 (0.03)	1.20 (0.05)		–1.04 (0.08)	1.04 (0.10)
PA13068	H01	0.91 (0.40)	2.38 (0.35)	0.36 (0.02)		
PA13069	H02	0.54 (0.12)	0.78 (0.26)	0.27 (0.07)		
PA13070	H03	1.11 (0.19)	1.08 (0.08)	0.21 (0.03)		
PA13071	H04	1.36 (0.22)	1.20 (0.07)	0.19 (0.02)		
PA13072	H05	0.64 (0.13)	1.03 (0.17)	0.23 (0.05)		
PA13073	H06	1.42 (0.40)	1.88 (0.13)	0.26 (0.02)		
PA13074	H07	0.84 (0.12)	0.81 (0.09)	0.13 (0.04)		
PA13075	H08	0.98 (0.14)	1.07 (0.07)	0.10 (0.03)		
PA13076	H09	0.86 (0.13)	1.32 (0.08)	0.08 (0.02)		
PA13077	H10	1.32 (0.18)	1.00 (0.06)	0.12 (0.02)		
PA13079	H12	0.95 (0.06)	1.02 (0.06)			
PA13080	H13	1.24 (0.08)	0.58 (0.04)			
PA13081	H14	0.63 (0.06)	2.09 (0.17)			
PA13083	H16	0.47 (0.02)	1.08 (0.05)		–2.10 (0.13)	2.10 (0.14)
PA13084	H17	0.47 (0.02)	1.38 (0.06)		–2.36 (0.14)	2.36 (0.16)
PA13085	H18	0.41 (0.03)	1.20 (0.08)		0.16 (0.08)	–0.16 (0.11)
PA13086A	H19A	0.59 (0.04)	0.76 (0.05)		–0.02 (0.07)	0.02 (0.08)
PA13086B	H19B	0.52 (0.05)	–0.53 (0.09)			
PA13001	P1_01	0.38 (0.04)	–0.75 (0.42)	0.29 (0.09)		
PA13002	P1_02	0.90 (0.07)	0.39 (0.07)	0.31 (0.03)		
PA13003	P1_03	0.86 (0.06)	–0.31 (0.10)	0.22 (0.04)		
PA13004	P1_04	0.59 (0.04)	–1.64 (0.21)	0.20 (0.08)		
PA13005	P1_05	0.53 (0.05)	–0.94 (0.31)	0.33 (0.08)		
PA13006	P1_06	1.23 (0.16)	1.68 (0.06)	0.30 (0.01)		

Exhibit C.4 IRT Parameters for the TIMSS Advanced 1995–2008 Concurrent Scaling of Physics (Continued)

Item		Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})
PA13009	P1_09	0.95 (0.09)	1.31 (0.05)	0.22 (0.01)		
PA13011	P2_01	0.44 (0.09)	1.11 (0.25)	0.30 (0.06)		
PA13012	P2_02	1.32 (0.14)	1.39 (0.05)	0.09 (0.01)		
PA13013	P2_03	0.88 (0.13)	1.31 (0.07)	0.23 (0.02)		
PA13014	P2_04	0.93 (0.09)	0.41 (0.08)	0.20 (0.03)		
PA13015	P2_05	0.56 (0.04)	–0.35 (0.05)	0.25 (0.00)		
PA13016	P2_06	0.99 (0.11)	1.17 (0.05)	0.12 (0.02)		
PA13017	P2_07	0.26 (0.05)	0.41 (0.48)	0.21 (0.08)		
PA13018	P2_08	0.53 (0.08)	0.98 (0.16)	0.21 (0.05)		
PA13019	P2_09	0.57 (0.11)	1.82 (0.13)	0.19 (0.03)		
PA13021	P3_01	0.92 (0.10)	0.71 (0.07)	0.20 (0.03)		
PA13022	P3_02	0.57 (0.02)	1.28 (0.04)		–1.10 (0.06)	1.10 (0.07)
PA13023	P3_03	0.47 (0.03)	–0.33 (0.05)			
PA13024	P3_04	0.59 (0.02)	1.01 (0.03)		–0.54 (0.05)	0.54 (0.06)
PA13025	P3_05	0.60 (0.02)	1.19 (0.04)		–0.89 (0.06)	0.89 (0.07)
PA13026	P3_06	0.90 (0.06)	1.78 (0.08)			
PA13027A	P3_07A	0.62 (0.04)	0.94 (0.05)			
PA23071	P4_01	0.56 (0.17)	1.41 (0.22)	0.43 (0.05)		
PA23146	P4_02	0.52 (0.04)	–0.36 (0.06)			
PA23029	P4_03	1.05 (0.12)	0.35 (0.07)	0.16 (0.03)		
PA23104	P4_04	0.78 (0.17)	0.87 (0.14)	0.38 (0.04)		
PA23038	P4_05	0.64 (0.09)	0.46 (0.14)	0.14 (0.05)		
PA23041	P4_06	0.40 (0.03)	–1.53 (0.12)			
PA23053	P4_07	0.37 (0.02)	0.28 (0.05)		–0.55 (0.09)	0.55 (0.09)
PA23148	P4_08	0.36 (0.11)	–0.34 (0.95)	0.20 (0.21)		
PA23119	P4_09	0.44 (0.02)	1.42 (0.06)		–2.07 (0.12)	2.07 (0.14)
PA23088	P4_10	0.58 (0.04)	–0.45 (0.06)			
PA23066	P4_11	0.58 (0.04)	0.38 (0.05)			
PA23048	P5_01	0.80 (0.14)	–0.20 (0.24)	0.49 (0.06)		
PA23039	P5_02	0.74 (0.14)	1.16 (0.11)	0.23 (0.04)		
PA23035	P5_03	0.65 (0.04)	–0.30 (0.05)			
PA23042	P5_04	0.74 (0.15)	1.07 (0.13)	0.33 (0.04)		
PA23012	P5_05	0.77 (0.05)	1.27 (0.07)			
PA23131	P5_06	0.91 (0.21)	1.57 (0.12)	0.25 (0.03)		
PA23051	P5_07	0.86 (0.05)	0.34 (0.04)			
PA23085	P5_08	0.91 (0.19)	1.39 (0.11)	0.31 (0.03)		
PA23130	P5_09	0.48 (0.02)	0.94 (0.04)		–1.54 (0.09)	1.54 (0.10)
PA23086	P5_10	0.61 (0.04)	1.10 (0.07)			
PA23064	P5_11	0.80 (0.13)	0.64 (0.12)	0.27 (0.04)		
PA23050	P6_01	0.98 (0.27)	1.08 (0.13)	0.35 (0.04)		
PA23056	P6_02	0.46 (0.18)	0.24 (0.70)	0.40 (0.14)		

Exhibit C.4 IRT Parameters for the TIMSS Advanced 1995–2008 Concurrent Scaling of Physics (Continued)

Item		Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})
PA23142	P6_03	1.10 (0.22)	0.58 (0.11)	0.27 (0.04)		
PA23072	P6_04	0.54 (0.05)	0.09 (0.08)	0.00 (0.00)		
PA23022	P6_05	0.39 (0.03)	1.72 (0.12)		–2.04 (0.18)	2.04 (0.21)
PA23030	P6_06	1.49 (0.33)	1.32 (0.09)	0.11 (0.02)		
PA23078	P6_07	0.26 (0.04)	0.72 (0.18)			
PA23113	P6_08	0.82 (0.23)	1.20 (0.15)	0.30 (0.04)		
PA23128	P6_09	0.36 (0.04)	0.73 (0.13)			
PA23058	P6_10	1.05 (0.22)	0.64 (0.12)	0.28 (0.04)		
PA23115	P6_11	1.16 (0.24)	1.07 (0.09)	0.16 (0.03)		
PA23110	P7_01	0.57 (0.15)	0.78 (0.28)	0.22 (0.09)		
PA23014	P7_02	0.49 (0.05)	0.01 (0.08)			
PA23025	P7_03	0.70 (0.04)	1.28 (0.06)		–0.93 (0.10)	0.93 (0.12)
PA23028	P7_04	0.95 (0.17)	0.52 (0.14)	0.25 (0.05)		
PA23034	P7_05	0.39 (0.05)	1.28 (0.18)			
PA23044	P7_06	0.71 (0.07)	1.29 (0.10)			
PA23082	P7_07	0.63 (0.06)	–0.01 (0.07)			
PA23140	P7_08	0.55 (0.16)	1.20 (0.22)	0.16 (0.08)		
PA23084	P7_09	0.67 (0.04)	1.25 (0.06)		–1.83 (0.15)	1.83 (0.16)
PA23059	P7_10	0.66 (0.21)	–0.25 (0.56)	0.45 (0.14)		
PA23138	P7_11	1.46 (0.29)	0.24 (0.13)	0.47 (0.05)		
PA23137	P7_12	0.51 (0.06)	0.56 (0.09)			

Appendix D

Item Descriptions Developed During the TIMSS Advanced 2008 Benchmarking

Advanced Mathematics

Items at Intermediate International Benchmark (475)

Algebra

M2_01	Rationalizes the denominator in an expression
M6_02	Solves a rational inequality with linear numerator and denominator

Calculus

M1_01	Determines the expression of a function of a function in a simple case
M1_04	Determines the limit of a rational function in x where the numerator and denominator are both quadratic as x tends to infinity
M2_03	Determines the sign of a rational function with numerator and denominator in factored form

M3_05	Recognizes from its graph the points where a function is not continuous
M4_05	Finds the second derivative of a simple function
M6_04	Determines the limit of a rational function where the numerator and denominator are both quadratic
M6_05	Differentiates an exponential function with a simple trigonometric exponent
M6_06	Differentiates a rational function where the numerator and denominator are both linear
M6_08	Integrates a function of the form $\frac{ax^2+b}{cx}$

Geometry

M1_08	Uses properties of an isosceles right triangle to determine the length of a given median
M2_08	Calculates the difference between vectors in coordinate form
M3_01	Identifies the three-dimensional figure traced out by a line rotating around another line
M3_06	Draws and labels the image of a triangle under reflection
M5_08	Identifies coordinates of the fourth vertex of a parallelogram when three vertices are given

Items at High International Benchmark (550)

Algebra

M1_02	Analyzes a piecewise-defined function consisting of linear segments to identify its graph
M1_03	Compares two models given in a word problem by solving a quadratic inequality

M1_09	Identifies the points with integer coordinates on a graph of a function of the form $y = \frac{a}{x}$
M4_01	Determines the term in a geometric sequence having a given value
M4_04	Analyzes steps in a given solution of a simple logarithmic equation and identifies an error
M5_02	Identifies two constants in a rational function given two points on its graph
M5_05	Solves a word problem by finding the distance between the points at which a parabola intersects the x-axis
M6_03	Identifies the graph that represents the relationship between the volume of a sphere and its diameter

Calculus

M1_06	Differentiates a function of the form $\frac{a}{\sqrt{bx + c}}$
M2_05	Differentiates an exponential function where the exponent is a simple polynomial
M3_04	Evaluates the definite integral of a function of the form $y = \frac{ax + b}{x^2}$
M4_06	Analyzes the graph of a function to determine the sign of its derivative
M6_07	Justifies a statement about slopes at two points on the graph of a trigonometric function
M7_06	Analyzes properties of a function and its first and second derivatives to identify its graph
M7_07	Determines the points of intersection with the x-axis of a simple function of the fourth degree

Geometry

M1_07	Finds the sum of the slopes of the three sides of an equilateral triangle with one side along the x-axis

M2_07	Identifies the equation of a line through a given point and perpendicular to a given line
M4_09	Evaluates the shortest path between opposite vertices on the surface of a cube
M4_10	Solves a word problem about height given the distance and angle of elevation
M4_11	Uses properties of vectors to analyze equivalence of conditions involving the sum and difference of two vectors
M6_09	Identifies the equation of a circle given its graph
M6_10	Uses basic properties of sine and cosine functions to determine the number of possible solutions of a simple trigonometric equation
M7_10	Identify solutions of a trigonometric equation of the form sin(ax)=b

Items at Advanced International Benchmark (625)

Algebra

M2_02	Calculates the cube of a complex number given in trigonometric form
M3_07	Apply the concept of limit in a word problem about regular polygons
M4_02	Solves a word problem about the number of permutations
M4_03	Solves a word problem comparing dimensions of two cylindrical containers given their volumes
M5_01	Given the first three terms, calculates the sum of an infinite geometric series
M5_03	Solves a straightforward logarithmic equation

M6_01	Given the first and third terms, calculates the sum of an infinite geometric series
M7_01	Solves a word problem by finding a certain term of a geometric sequence
M7_03	Determines the coefficients of a quadratic function given the points of intersection between the graph and the axes
M7_04	Finds the minimum of a function of a function

Calculus

M3_05	Recognizes from its graph the points where a function is not differentiable
M5_06	Given the graph of the derivative of a function, determines the *x*-values of the maximum point and the point of inflection of the function
M7_05	Solves a multi-step word problem involving distance as a function of time
M7_07	Determines the maximum and minimum points of a simple function of the fourth degree
M7_08	Calculates the definite integral given the graph of a function and the areas between the curve and the *x*-axis

Geometry

M2_09	Given two points, identifies an equation that represents the set of all points twice as far from one of the given points as from the other
M2_10	Uses vector sums and differences to express a relationship among three vectors shown in a figure
M3_09	Based on the coordinates of the vertices of a given quadrilateral (which is a parallelogram), proves that the diagonals of that particular quadrilateral bisect each other

M5_09	Given functions of the form y=a sin(bx+c), compares amplitudes and periods
M6_11	Solves a multi-step word problem involving trigonometric ratios to identify the length of a side of a regular polygon inscribed in a circle
M7_11	Given two points on a line and a triangle in a Cartesian plane, uses mathematical properties to determine whether the line is parallel to a side of the triangle

Items above Advanced International Benchmark (625)

Algebra

M3_08	Specifies the essential steps of a proof by induction
M5_04	Given one imaginary root, identifies the constant term of a third-degree polynomial with known coefficients
M7_02	Rationalizes an expression where the denominator is a complex number

Calculus

M2_06	Maximizes the volume of a cylinder given a relationship between its height and diameter
M4_07	Solves a multi-step word problem by maximizing the profit given a quadratic cost function and the unit selling price
M4_08	Calculates the area between the graphs of a linear and a quadratic function
M5_05	Solves a multi-step word problem by calculating the area between two intersecting parabolas whose equations are given

M5_07	Determines the vertical line that divides a specified area between a parabola and the *x*-axis into equal parts
M7_09	Identifies the indefinite integral of an exponential function where the exponent is a linear polynomial

Geometry

M3_06	Draws and labels the image of a triangle under rotation
M5_10	Calculates the two possible lengths of a side of a triangle given an angle and the lengths of two sides that do not include the angle

Physics

Items at Intermediate International Benchmark (475)

Mechanics

P1_05	Calculates falling distance from rest, assuming negligible air resistance
P3_03	Uses the relationship between wave speed and wavelength to calculate the wavelength
P4_02	Identifies a basic property of circular motion, given constant speed
P7_02	Identifies forces acting on a body thrown up into the air

Electricity and Magnetism

P1_04	Recognizes the circuit showing resistances that corresponds to given conditions
P3_01	Orders types of electromagnetic radiation by wavelength
P4_06	Identifies the meaning of the symbols in a formula

P5_01	Identifies a given range of wavelengths

Heat and Temperature

P2_05	Recognizes a process of energy transfer
P4_08	Applies knowledge of the gas and energy laws in a meteorological situation
P6_02	Selects the best explanation of the greenhouse effect
P7_07	Relates specific heat capacities of different materials to observed phenomena

Atomic and Nuclear Physics

P2_01	Identifies a correct statement about black lines in the spectrum of light
P2_07	Recognizes a statement consistent with the photoelectric effect
P4_10	Identifies the number of protons and neutrons in given isotopes
P7_10	Recognizes the number of neutrons in a nucleus, given its atomic notation
P7_11	Selects the best description of an atomic nucleus

Items at High International Benchmark (550)

Mechanics

P1_01	Interprets oscilloscope readings with regard to pitch and loudness of sounds
P1_03	Applies Newton's Laws to recognize the tension in a string connecting hanging objects
P2_04	Derives an expression for the speed of an object moving in a vertical circular path

P4_03	Recognizes a situation where mechanical energy is transformed into heat
P6_04	Applies the energy law to calculate the maximum compression of a spring

Electricity and Magnetism

P4_04	Recognizes the direction of the electric force on a charged object in an electric field
P4_05	Applies understanding of series and parallel connections of resistors to compare total resistances
P5_03	Applies Ohm's Law and the Joule's law to solve a problem about resistance
P5_04	Recognizes paths of particles in a magnetic field
P7_05	Draws an arrow from a certain point showing the direction of an electric field from two point charges

Heat and Temperature

P5_07	Applies knowledge of specific heat to solve a problem of transfer of energy
P6_03	Identifies the type of electromagnetic radiation related to the temperature of a heat-emitting body

Atomic and Nuclear Physics

P1_02	Uses the law of radioactive decay to calculate the half-life of a radioactive element
P6_10	Recognizes that the nucleus of an atom is very small relative to the size of the entire atom

Items at Advanced International Benchmark (625)

Mechanics

P3_07	Uses a graph of experimental data about a falling object to calculate the value of acceleration due to gravity.
P4_01	Selects the graph that best represents variation of potential energy of a moving body
P5_05	Solves a problem by using the characteristics of free fall
P7_01	Applies Newton's third law of motion to compare the size of forces
P7_04	Interprets a graph and applies the definition of momentum to solve a problem

Electricity and Magnetism

P1_06	Applies Coulomb's law to find a point where the net force from two charges on a third charge is zero
P1_09	Analyzes changes in ammeter and voltmeter readings in a complex circuit diagram
P2_06	Identifies the direction of the force on a current-carrying conductor in a given magnetic field
P2_08	Analyzes a complex circuit diagram to solve a power consumption problem
P5_02	Interprets a current-by-resistance graph to calculate the internal resistance of a battery
P6_06	Identifies mutual electric forces acting on two charged particles
P6_09	Recalls that glass absorbs ultraviolet light

Heat and Temperature

P4_07	Applies the gas laws to solve a straightforward problem

P5_08	Applies coefficients of linear expansion to compare the lengths of two rods of different materials
P6_01	Applies knowledge of heat conduction in different materials
P7_08	Identifies the range of temperatures at which electromagnetic radiation is visible

Atomic and Nuclear Physics

P2_03	Applies knowledge of how Rutherford's scattering experiment worked
P2_09	Recognizes the effect of a nuclear reaction on the atomic and mass numbers of an atom
P4_11	Completes the equation for a nuclear reaction
P5_11	Applies knowledge of radioactive decay in the solution of word problems
P6_11	Recognizes a basic explanation of beta decay in a radioactive isotope
P7_12	Writes the symbol for a particular atomic nucleus given the number of its protons and neutrons

Items above Advanced International Benchmark (625)

Mechanics

P2_02	Applies Newton's third law to identify forces on two interacting spring balances
P5_06	Demonstrates knowledge of the most fundamental principle of relativity
P6_05	Uses Newton's second law and the law of gravity to solve a problem involving planetary motion

P7_03	Uses the law of conservation of momentum to formulate and solve a multi-step word problem

Electricity and Magnetism

P3_04	Applies the principle of equilibrium of electrical and gravitational forces acting on a charged object to calculate the electric field strength
P3_06	Shows that the period of revolution of a charged particle in a magnetic field is independent of its speed
P6_07	Demonstrates understanding of the effect of two point charges on a third charge when the positions of the first two charges are interchanged
P6_08	Recognizes that a laser beam can cause damage because the beam stays highly concentrated
P7_06	Describes a procedure to demonstrate electromagnetic induction

Heat and Temperature

P3_02	Calculates final temperature when two materials of different temperatures are brought together
P4_09	Interprets a nonroutine problem situation and explains that an object in temperature equilibrium gains heat at the same rate as it loses it
P5_09	Applies knowledge of light absorption in a problem situation about observed color
P5_10	Interprets a nonroutine problem situation and relates wavelengths of light to the temperature of the emitting body

P7_09 Interprets a complex problem situation and applies the gas laws and Dalton's law of mixtures to calculate pressure

Atomic and Nuclear Physics

P3_05 Applies Einstein's equation for the photoelectric effect to explain whether electrons will be emitted from different metals

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

BOSTON COLLEGE

ISBN 1-889938-56-4

TIMSS Advanced 2008

timssandpirls.bc.edu