

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.

