

# Chapter 1



## *Overview of TIMSS 2007*

Ina V.S. Mullis and Michael O. Martin

### **1.1 Introduction**

IEA is a recognized pioneer of international assessments, having conducted comparative studies of students' academic achievement for 50 years. However, even for IEA and its TIMSS & PIRLS International Study Center at Boston College, TIMSS 2007 was a very complex and ambitious undertaking, involving 67 participants (59 countries and 8 benchmarking entities) in a cooperative, global endeavor to develop and implement a wide-ranging assessment of mathematics and science achievement at fourth and eighth grades, providing a wealth of information about the educational context and current achievement of students in 2007, while measuring trends from earlier cycles of TIMSS in 1995, 1999, and 2003.

The design, development, and implementation of TIMSS 2007 are documented in a series of publications produced at various stages of the project. The *TIMSS 2007 Assessment Frameworks* (Mullis, Martin, Ruddock, O'Sullivan, Arora, & Erberber, 2005) contains the mathematics and science frameworks underlying the assessments at the fourth and eighth grades as well as the contextual framework for the questionnaires, and describes the design of the assessment.

Implementing TIMSS 2007 involved widespread participation from countries around the world, many of whom were collecting the third or fourth cycle of trend data. The *TIMSS 2007 International Mathematics Report* (Mullis, Martin, & Foy, 2008) and the *TIMSS 2007 International Science Report* (Martin, Mullis, & Foy, 2008) summarize fourth- and eighth-grade students' mathematics and science achievement in each of the 59 participating countries and 8 benchmarking participants. The complete TIMSS 2007 database is available on DVD accompanied by the *TIMSS 2007 User Guide for the International Database* (Foy & Olson, 2009).

The *TIMSS 2007 Encyclopedia* (Mullis, Martin, Olson, Berger, Milne, & Stanco, 2008) contains the countries' and benchmarking participants' descriptions of their national contexts for mathematics and science education as well as their mathematics and science curricula. The more qualitative information provided in the *TIMSS 2007 Encyclopedia* is intended to complement both the *TIMSS 2007 International Mathematics Report* and the *TIMSS 2007 International Science Report*.

The purpose of the *TIMSS 2007 Technical Report* is to provide further detail and documentation about the processes underlying the development of the TIMSS 2007 instruments and the methods used in sampling, data collection, scaling, and data analysis. In particular, the *TIMSS 2007 Technical 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 and benchmarking entities.

## 1.2 Participants in TIMSS 2007

Exhibit 1.1 shows a map of the world identifying the TIMSS 2007 countries and benchmarking participants (regional entities that follow the same assessment procedures as the countries). Exhibit 1.2 lists the TIMSS 2007 participants, and indicates the grade(s) at which they participated and the previous cycles of TIMSS they participated in at that grade. It can be seen that many of the TIMSS 2007 countries and benchmarking participants have data for both the fourth and eighth grades. Exhibit 2 also shows that most TIMSS 2007 participants have trend data and, for each participant, whether it is for two, three, or four points in time—1995, 1999, 2003, and 2007.

TIMSS 2007 was administered near the end of the school year in each country. In countries in the Southern Hemisphere (where the school year typically ends in November or December) the assessment was conducted in October or November 2006. In the Northern Hemisphere, the school year typically ends in June; so in these countries the assessment was conducted in April, May, or June 2007.

### 1.3 TIMSS 2007 Instruments

The TIMSS 2007 assessment contained 353 items at the fourth grade, including 179 in mathematics and 174 in science. At the eighth grade there were 429 items, 215 in mathematics and 214 items in science. At both grades, the TIMSS 2007 assessment involved assembling the items into 14 blocks of mathematics items and 14 blocks of science items, and then assembling the blocks into 14 booklets, each one including 2 blocks of mathematics items and 2 blocks of science items assembled according to a very careful rotated design. Each student was administered a single booklet. Details about the development process and types of items can be found in Chapter 2.

Chapter 3 contains information about developing the four different types of background questionnaires. In brief, students answered questions pertaining to their home and school environments. The teachers of the sampled students responded to questions about characteristics of the class tested, instructional activities for teaching mathematics or science, the topics covered in students' lessons, and about their education, training, and opportunities for professional development. The principals of schools responded to questions about enrolment and school characteristics, school organization, staffing and resources, and the school environment. The *Curriculum Questionnaire*, a responsibility of the National Research Coordinators, provided data about participants' mathematics and science curricula. As an innovation for TIMSS 2007, the *Curriculum Questionnaire* was administered online.

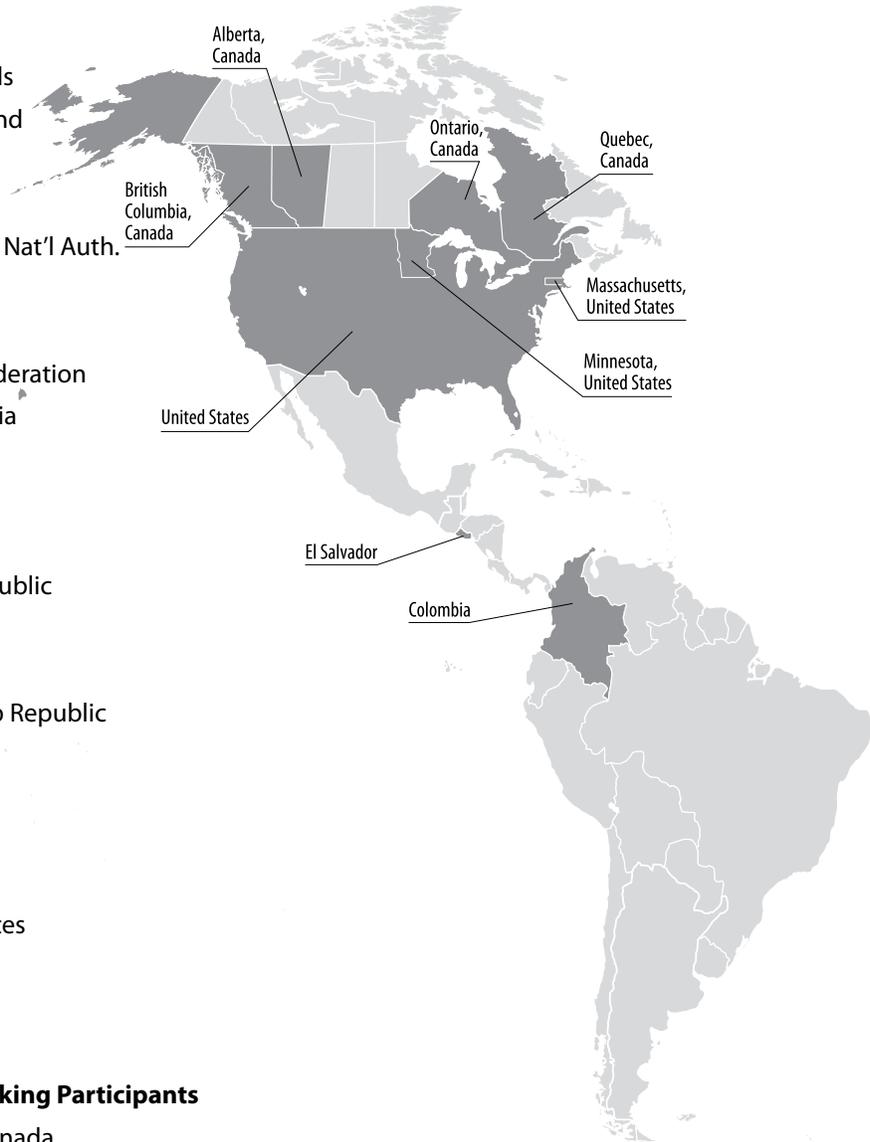
To increase reliability in reporting background data, the questions in the background questionnaires form a number of scales. These scales and other sets of background questions are used to create background indices for reporting. The methods used to create the TIMSS 2007 background indices are discussed in Chapter 12.

**Exhibit 1.1 Countries Participating in TIMSS 2007**

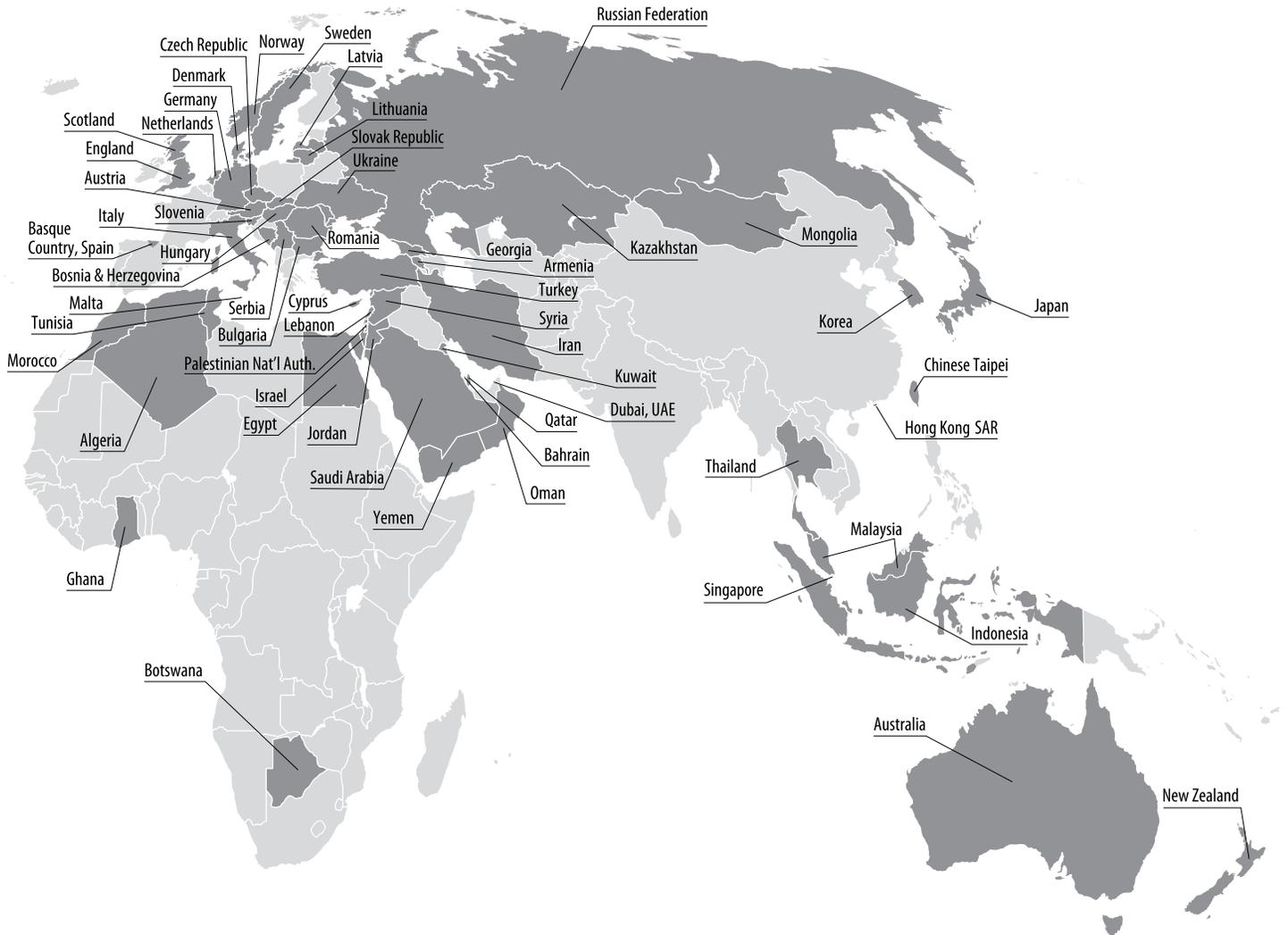
Algeria	Mongolia
Armenia	Morocco
Australia	Netherlands
Austria	New Zealand
Bahrain	Norway
Bosnia and Herzegovina	Oman
Botswana	Palestinian Nat'l Auth.
Bulgaria	Qatar
Chinese Taipei	Romania
Colombia	Russian Federation
Cyprus	Saudi Arabia
Czech Republic	Scotland
Denmark	Serbia
Egypt	Singapore
El Salvador	Slovak Republic
England	Slovenia
Georgia	Sweden
Germany	Syrian Arab Republic
Ghana	Thailand
Hong Kong SAR	Tunisia
Hungary	Turkey
Indonesia	Ukraine
Iran, Islamic Rep. of	United States
Israel	Yemen
Italy	
Japan	
Jordan	
Kazakhstan	
Korea, Rep. of	
Kuwait	
Latvia	
Lebanon	
Lithuania	
Malaysia	
Malta	

**Benchmarking Participants**

Alberta, Canada
Basque Country, Spain
British Columbia, Canada
Dubai, UAE
Massachusetts, US
Minnesota, US
Ontario, Canada
Quebec, Canada



**Exhibit 1.1 Countries Participating in TIMSS 2007 (Continued)**



**Exhibit 1.2 Countries Participating in TIMSS 1995 Through 2007**

Country	Grade 4			Grade 8			
	2007	2003	1995	2007	2003	1999	1995
Algeria	●			●			
Armenia	●	●		●	●		
Australia	●	●	●	●	●	●	●
Austria	●		●				●
Bahrain				●	●		
Bosnia and Herzegovina				●			
Botswana				●	●		
Bulgaria				●	●	●	●
Chinese Taipei	●	●		●	●	●	
Colombia	●			●			●
Cyprus		●	●	●	●	●	●
Czech Republic	●		●	●		●	●
Denmark	●						●
Egypt				●	●		
El Salvador	●			●			
England	●	●	●	●	●	●	●
Georgia	●			●			
Germany	●						●
Ghana				●	●		
Hong Kong SAR	●	●	●	●	●	●	●
Hungary	●	●	●	●	●	●	●
Indonesia				●	●	●	
Iran, Islamic Rep. of	●	●	●	●	●	●	●
Israel			●	●	●	●	●
Italy	●	●	●	●	●	●	●
Japan	●	●	●	●	●	●	●
Jordan				●	●	●	
Kazakhstan	●						
Korea, Rep. of			●	●	●	●	●
Kuwait	●		●	●			●
Latvia	●	●	●		●	●	●
Lebanon				●	●		
Lithuania	●	●		●	●	●	●
Malaysia				●	●	●	
Malta				●			
Mongolia	●			●			
Morocco	●	●		●	●	●	
Netherlands	●	●	●		●	●	●
New Zealand	●	●	●		●	●	●
Norway	●	●	●	●	●		●
Oman				●			
Palestinian Nat'l Auth.				●	●		
Qatar	●			●			
Romania				●	●	●	●
Russian Federation	●	●		●	●	●	●
Saudi Arabia				●	●		

SOURCE: IEA's Trends in International Mathematics and Science Study (TIMSS) 2007

**Exhibit 1.2 Countries Participating in TIMSS 1995 Through 2007 (Continued)**

Country	Grade 4			Grade 8			
	2007	2003	1995	2007	2003	1999	1995
Scotland	●	●	●	●	●		●
Serbia				●	●		
Singapore	●	●	●	●	●	●	●
Slovak Republic	●				●	●	●
Slovenia	●	●	●	●	●	●	●
Sweden	●			●	●		●
Syrian Arab Republic				●	●		
Thailand			●	●		●	●
Tunisia	●	●		●	●	●	
Turkey				●		●	
Ukraine	●			●			
United States	●	●	●	●	●	●	●
Yemen	●	●					
<b>Benchmarking Participants</b>							
Alberta, Canada	●		●			●	●
Basque Country, Spain				●	●		
British Columbia, Canada	●			●		●	
Dubai, UAE	●			●			
Massachusetts, US	●			●		●	
Minnesota, US	●		●	●			●
Ontario, Canada	●	●	●	●	●	●	●
Quebec, Canada	●	●	●	●	●	●	●

SOURCE: IEA's Trends in International Mathematics and Science Study (TIMSS) 2007

#### **1.4 Translation Verification**

Chapter 4 describes the steps involved in translating the test instruments and background questionnaires from English into numerous different languages. To ensure comparability among translated instruments, participants are given detailed specifications about the process to use in translating the materials, the IEA Secretariat manages a rigorous translation verification process, and the TIMSS & PIRLS International Study Center conducts a verification of final instrument layout before instruments are printed.

#### **1.5 Sample Design, Implementation, and Participation**

As explained in Chapter 5, the TIMSS 2007 assessment was administered to carefully drawn probability samples of students from the target populations in each country. The target populations were students enrolled in the fourth grade or eighth grade of formal schooling, counting from the first year of primary school defined by UNESCO's International Standard Classification for Education (UNESCO, 1999). Accordingly, the fourth year or the eighth year of formal schooling should be the fourth grade or eighth grade, respectively, in most countries. However, to avoid testing students who are very young, TIMSS has a policy that the average age of children in the grade tested should not be below 9.5 for the fourth year of schooling or 13.5 for the eighth year of schooling. The basic sampling design was a two-stage stratified cluster design. The first stage consisted of sampling schools, and the second stage consisted of sampling intact classrooms from the target grade in the sampled schools. Typically, at each grade, countries sampled 150 schools and one or two intact classrooms.

Information about the sampling weights and documentation of the participation rates is found in Chapter 9. Most 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.

#### **1.6 Survey Operations and Quality Assurance in Data Collection**

Each country and benchmarking participant was responsible for carrying out all aspects of data collection and scoring, using standardized procedures explained in a series of survey operations procedure units and various training manuals. The data collection and scoring procedures are described in Chapter 6. In addition, 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 7 indicated that, in general, the national centers were able to conduct the data collection efficiently, professionally, and in compliance with international procedures.

### **1.7 The TIMSS 2007 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 were forwarded from the participants, the data underwent an exhaustive cleaning process. As described in Chapter 8, 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.

Chapter 10 provides details about the process implemented by the TIMSS & PIRLS International Study Center to review 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. In general, the items exhibited very good psychometric properties in all countries, and the scoring reliability was satisfactory (above 90% agreement in most cases).

### **1.8 Scaling the Achievement Data and the International Benchmarks**

The TIMSS mathematics and science achievement scales were designed to provide reliable measures of student achievement across the trend cycles of the TIMSS assessments, based on the metric established with the 1995 data. As described in Chapter 11, student achievement was summarized using item response theory (IRT) scaling methods. For more accurate estimation of results for subpopulations of students, the TIMSS scaling made use of plausible-value technology. In addition to the overall scales used to estimate student achievement in each assessment including TIMSS 2007 and to measure trends over time, IRT scales were created for each of the content and cognitive domains described in the *TIMSS 2007 Assessment Frameworks*.

Chapter 13 describes the procedures used to report student achievement at the TIMSS International Benchmarks. To describe what performance on the TIMSS achievement scales means in terms of students' mathematics

or science proficiency, TIMSS conducted a scale anchoring analysis to describe and interpret student achievement at the Advanced (625), High (550), Intermediate (475), and Low (400) International Benchmarks.

### 1.9 Ensuring Comparative Validity

In conclusion, a major purpose of the *TIMSS 2007 Technical Report* is to provide detailed documentation about the procedures and methods used by TIMSS 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 updating the assessment frameworks for TIMSS 2007 through release of the international database and User's Guide.

TIMSS is dedicated to addressing the classical concerns of high quality measurement—reliability and validity. TIMSS has procedures to ensure that the tests are reliable, that is, that they are constructed with sufficient items to provide reliable measurement, and are accompanied by detailed administration and scoring procedures and supported by extensive training to ensure that the results are not impacted by extraneous factors. Because reliability is not sufficient for good measurement, considerable effort also is dedicated to the validity of the tests, that is, the extent to which inferences drawn from the results can be supported by evidence. For example, does a student with high achievement on the TIMSS mathematics achievement scale actually have a high degree of proficiency in mathematics for an eighth grade student? Validity involves unified agreement in conceptualizing and articulating the constructs of mathematics and science as they apply to fourth- and eighth-grade students, and unified agreement that the items included in the assessments measure those articulations of mathematics and science, respectively.

In addition, as an international study, TIMSS must have comparative validity. For comparative validity, the classical 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.

The various chapters of this report describe the TIMSS quality assurance program to ensure comparative validity. Chapter 2 describes how the *TIMSS 2007 Frameworks* were updated through widespread collaboration among the participating countries, and modified to align with

current perspectives. The updates were based on surveys of the participating countries and 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 2007 involved a full-scale field test that was essential for confirming the appropriateness and comparability of the items. The translation verification process for the TIMSS 2007 assessment is described in Chapter 4. Every effort was made to ensure that the translations were comparable across countries. The data collection and scoring methods are described in Chapter 6, including the complete documentation of the survey operations procedures in manuals and specific training in aspects of data collection. The results of the TIMSS Quality Control Monitoring program conducted as part of test administration are included in Chapter 7.

Chapter 5 describes the efforts taken to ensure sample comparability across countries. Chapter 9 describes the implementation of the sampling procedures. With very few exceptions, countries assessed the correct grade(s), included all of the students in their definition of the target population, kept exclusions to a minimum (lower than 5%), and implemented accurate classroom sampling using the WinW3S software developed by IEA for this purpose. These chapters also describe how each country's sampling procedures must be fully documented, and that the participation rates must be high standards (at least 85% of students and 85% of schools) or be annotated in the international reports.

Chapter 8 addresses the issue of the comparability of the data. It explains how the IEA Data Processing and Research Center (DPC) provides data entry software and variable codebooks to standardize data preparation and conducts extensive training seminars. The IEA DPC checks within each country's data files and across countries for internal consistency and accuracy, and interacts with the country's to resolve data issues.

Subsequent to the field test, and then again, prior to scaling, a thorough review of item statistics was conducted. For every item for every country, the results are scrutinized for discrimination and scoring reliability. Also, the data are reviewed for item-by-country interactions. As described in Chapter 10, for each item TIMSS examined each country's performance on the item

in light of its overall performance in 2007, and for trend countries this was done to compare interaction patterns for both assessments.

Chapter 11 describes the scaling methodology for the TIMSS 2007 achievement data, 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. The scaling was implemented separately for each country and separately for the 30 different scales (overall achievement, content domains, and cognitive domains for the fourth and eighth grades) and all of the results were plotted and checked. Again, with a few rare exceptions for the more difficult domains, the achievement score distributions were very satisfactory and provided an excellent basis for further analysis and reporting the results.

TIMSS devotes considerable effort to ensure that the data can be trusted for important decision making based on comparisons between countries and much of that information is documented in this technical report.

## References

---

- Foy, P. & Olson, J.F. (2009). *TIMSS 2007 user guide for the international database*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M.O., Mullis, I.V.S., & Foy, P. (with Olson, J.F., Erberber, E., Preuschoff, C., & Galia, J.). (2008). *TIMSS 2007 international science report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I.V.S., Martin, M.O., & Foy, P. (with Olson, J.F., Preuschoff, C., Erberber, E., Arora, A., & Galia, J.). (2008). *TIMSS 2007 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I.V.S., Martin, M.O., Olson, J.F., Berger, D.R., Milne, D., & Stanco, G.M. (Eds.). (2008). *TIMSS 2007 encyclopedia: A guide to mathematics and science education around the world*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I.V.S., Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y., Arora, A., & Erberber, E. (2005). *TIMSS 2007 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- UNESCO Institute for Statistics. (1999). *Operational manual for ISCED-1997 (international standard classification of education)*. Paris: Author.