

## Overview

The TIMSS 2007 international assessment of student achievement comprises written tests in mathematics and science together with a set of questionnaires that gather information on the educational and social contexts for achievement. Central to the TIMSS mission is the measurement of student achievement in mathematics and science in a way that does justice to the breadth and richness of these subjects as they are taught in the participating countries, and that monitors countries' improvement or decline by tracking trends in student performance from one assessment cycle to the next. This requires an assessment that is wide ranging in its coverage of mathematics and science and innovative in its measurement approach.

To maximize the information provided to educators and policy makers, TIMSS reports student performance in the major content domains of the subjects (i.e., number, algebra, geometry, and data and chance in mathematics at the eighth grade and biology, chemistry, physics, and earth science in eighth-grade science) as well as performance in mathematics and science overall. In addition, TIMSS in 2007 will for the first time report student achievement by cognitive domain – i.e., knowing, applying, and reasoning. These ambitious coverage and reporting goals make for enormously valuable information for policy decisions, teacher education, and instructional practices, but pose significant challenges for data collection, analysis, and reporting.

A major consequence of TIMSS' ambitious assessment goals is that there are many more questions on the assessment than can be answered by a student in the amount of testing time available. Accordingly, TIMSS 2007 uses a matrix-sampling approach that involves packaging the entire assessment pool of mathematics and science questions into a set of 14 student achievement booklets, with each student completing just one booklet. Each question, or item, appears in two booklets, providing a mechanism for linking together the student responses from the various booklets. Booklets are distributed among students in participating classrooms so that the groups of students completing each booklet are approximately equivalent in terms of student ability. Using Item-Response Theory (IRT) scaling techniques, a comprehensive picture of the achievement of the entire student population is

assembled from the combined responses of individual students to the booklets they are assigned. This approach reduces to manageable proportions what otherwise would be an impossible student burden, albeit at the cost of greater complexity in booklet assembly, data collection, and data analysis.

### **TIMSS 2007 Item Blocks**

To facilitate the process of creating the student achievement booklets, TIMSS groups the assessment items into a series of item blocks, with approximately 10-15 items in each block. As in the TIMSS 2003 assessment, TIMSS 2007 has 28 blocks in total, 14 containing mathematics items and 14 containing science items. Student booklets were assembled from various combinations of these item blocks.

Following the 2003 assessment, half of the mathematics blocks and half of the science blocks were secured for use in measuring trends in 2007. The remaining blocks were released into the public domain for use in publications, research, and teaching, to be replaced by newly-developed items in the TIMSS 2007 assessment. Accordingly, the 28 blocks in the TIMSS 2007 assessment comprise 14 blocks of trend items (seven mathematics and seven science) and 14 blocks of new items developed for 2007. As shown in Exhibit 10, the TIMSS 2007 mathematics blocks are labeled M01 through M14 and the science blocks S01 through S14. Blocks with labels ending in odd numbers (01, 03, 05, etc.) contain the trend items from the 2003 assessment, and those ending in even numbers the new items developed for use for the first time in TIMSS 2007.

## Exhibit 10: TIMSS 2007 Student Booklet Design – Fourth and Eighth Grades

Mathematics Blocks	Source of Items	Science Blocks	Source of Items
M01	Block M05 from TIMSS 2003	S01	Block S14 from TIMSS 2003
M02	New items for TIMSS 2007	S02	New items for TIMSS 2007
M03	Block M06 from TIMSS 2003	S03	Block S05 from TIMSS 2003
M04	New items for TIMSS 2007	S04	New items for TIMSS 2007
M05	Block M07 from TIMSS 2003	S05	Block S06 from TIMSS 2003
M06	New items for TIMSS 2007	S06	New items for TIMSS 2007
M07	Block M08 from TIMSS 2003	S07	Block S07 from TIMSS 2003
M08	New items for TIMSS 2007	S08	New items for TIMSS 2007
M09	Block M11 from TIMSS 2003	S09	Block S08 from TIMSS 2003
M10	New items for TIMSS 2007	S10	New items for TIMSS 2007
M11	Block M12 from TIMSS 2003	S11	Block S11 from TIMSS 2003
M12	New items for TIMSS 2007	S12	New items for TIMSS 2007
M13	Block M14 from TIMSS 2003	S13	Block S12 from TIMSS 2003
M14	New items for TIMSS 2007	S14	New items for TIMSS 2007

Although the general arrangement of items into blocks follows the 2003 approach, based on experience from that assessment the estimated amount of time needed by students to complete a block was increased – from 15 to 22.5 minutes at eighth grade and from 12 to 18 at fourth grade. Consequently, the 28 blocks of eighth-grade items contain an estimated 10½ hours of testing time and the fourth-grade blocks, almost 8½ hours. From past experience with TIMSS, National Research Coordinators from participating countries agreed that the testing time for any one student should not be increased from previous assessments. Thus, as in the past, the assessment time for each student booklet had to fit into 90 minutes for eighth grade and 72 minutes for fourth grade. An additional 30 minutes for a student questionnaire also was planned at each grade level.

## TIMSS 2007 Block Design for Student Achievement Booklets

In choosing how to distribute assessment blocks across student achievement booklets, the major goal was to maximize coverage of the framework while ensuring that every student responded to sufficient items to provide reliable measurement of trends in both mathematics and science. A further goal was to ensure that trends in the mathematics and science content areas could be measured reliably. To enable linking among booklets while keeping the number of booklets to a minimum, each block appears in two booklets.

In the TIMSS 2007 design, the 28 assessment blocks are distributed across 14 student achievement booklets (see Exhibit 11). The same booklet design is used at both fourth and eighth grade, although the eighth-grade blocks contains 22.5 minutes of assessment items and the fourth grade blocks 18 minutes. Each student booklet consists of four blocks of items, two blocks of mathematics items and two of science items. In half of the booklets, the mathematics blocks come first, followed by the science blocks, and in the other half the order is reversed. Additionally, two of the blocks in each booklet contain trend items from 2003 and two contain items newly developed for TIMSS 2007. For example, as may be seen from Exhibit 11, students assigned Booklet 1 complete two blocks of mathematics items, M01 and M02, and two blocks of science items, S01 and S02. The items in blocks M01 and S01 are trend items from TIMSS 2003, while those in M02 and S02 are items new for TIMSS 2007. Similarly, students assigned Booklet 2 complete two science blocks, S02 and S03, followed by two mathematics blocks, M02 and M03. S02 and M02 contain the new items and S03 and M03 the trend items.

**Exhibit 11: TIMSS 2007 Student Achievement Booklet Design –  
Fourth and Eighth Grades**

Student Achievement Booklet	Assessment Blocks			
	Part 1		Part 2	
Booklet 1	M01	M02	S01	S02
Booklet 2	S02	S03	M02	M03
Booklet 3	M03	M04	S03	S04
Booklet 4	S04	S05	M04	M05
Booklet 5	M05	M06	S05	S06
Booklet 6	S06	S07	M06	M07
Booklet 7	M07	M08	S07	S08
Booklet 8	S08	S09	M08	M09
Booklet 9	M09	M10	S09	S10
Booklet 10	S10	S11	M10	M11
Booklet 11	M11	M12	S11	S12
Booklet 12	S12	S13	M12	M13
Booklet 13	M13	M14	S13	S14
Booklet 14	S14	S01	M14	M01

As summarized in Exhibit 12, each student completes one student achievement booklet in two parts, followed by a student questionnaire. The individual student response burden for the assessment is similar to TIMSS in 2003, i.e., 72 minutes for the assessment and 30 minutes for the questionnaire at fourth grade, and 90 minutes and 30 minutes, respectively, at eighth grade.

**Exhibit 12: TIMSS 2007 Student Testing Time – Fourth and Eighth Grades**

Activity	Fourth Grade	Eighth Grade
Student Achievement Booklet – Part 1	36 minutes	45 minutes
Break		
Student Achievement Booklet – Part 2	36 minutes	45 minutes
Break		
Student Questionnaire	30 minutes	30 minutes

Countries participating in TIMSS aim for a sample of at least 4,500 students to ensure that there are enough respondents for each item. The 14 student booklets are distributed among the students in each sampled class according to a predetermined order, so that approximately equal proportions of students respond to each booklet.

### **Question Types and Scoring Procedures**

Students' knowledge and understanding of mathematics and science are assessed through a range of questions in each subject. Two question formats are used in the TIMSS assessment – multiple-choice and constructed-response. At least half of the total number of points represented by all the questions will come from multiple-choice questions. Each multiple-choice question is worth one score point.

Constructed-response questions generally are worth one or two score points, depending on the nature of the task and the skills required to complete it. However, building on the TIMSS 2003 special initiative to develop problem solving and inquiry tasks (known as PSIs) funded by the US National Science Foundation, TIMSS 2007 will have extended reasoning tasks or item sets of related problems. In particular, these reasoning tasks address the thinking skills described in the reasoning cognitive domain in the mathematics and science frameworks, respectively. For the reasoning tasks or item sets, the number of possible points, typically three to six points, depends on the requirements for students to successfully complete the task. In developing assessment questions, the choice of item format depends on the mathematics or science being assessed, and the format that best enables students to demonstrate their proficiency.

*Multiple-Choice Questions.* Multiple-choice questions provide students with four response options, of which only one is correct. These questions can be used to assess any of the behaviors in the cognitive domains. However, because they do not allow for students' explanations or supporting statements, multiple-choice questions may be less suitable for assessing students' ability to make more complex interpretations or evaluations.

In assessing fourth- and eighth-grade students, it is important that linguistic features of the questions be developmentally appropriate. Therefore, the questions are written clearly and concisely. The response options also are written succinctly in order to minimize the reading load of the question. The options that are incorrect are written to be plausible, but not deceptive. For students who may be unfamiliar with this test question format, the instructions given at the beginning of the test include a sample multiple-choice item that illustrates how to select and mark an answer.

*Constructed-Response Questions.* For this type of test item students are required to construct a written response, rather than select a response from a set of options. Constructed-response questions are particularly well-suited for assessing aspects of knowledge and skills that require students to explain phenomena or interpret data based on their background knowledge and experience.

The scoring guide for each constructed-response question describes the essential features of appropriate and complete responses. The guides focus on evidence of the type of behavior the question assesses. They describe evidence of partially correct and completely correct responses. In addition, sample student responses at each level of understanding provide important guidance to those who will be rating the students' responses. In scoring students' responses to constructed-response questions, the focus is solely on students' achievement with respect to the topic being assessed, not on their ability to write well. However, students need to communicate in a manner that will be clear to those scoring their responses.

In addition, scoring guides are designed to enable, for each item, identification of the various successful, partially successful, and

unsuccessful approaches. Diagnosis of common learning difficulties in mathematics and science as evidenced by misconceptions and errors is an important aim of the study.

Since constructed-response questions constitute an important part of the assessment and are an integral part of the measurement of trends, it is very important for scoring guides to be implemented consistently in all countries and in each data collection year. To ensure consistent application of the scoring guides for trend items in the 2007 assessment, IEA has archived samples of student responses to previous assessments from each country; these are used to train scorers in 2007 and to monitor consistent application.

*Score Points.* In developing the assessment, the aim is to create blocks of items that each provide, on average, about 15 score points at eighth grade and about 12 score points at fourth grade. Item blocks contain a variety of item types, including multiple-choice items (1 point each) and constructed-response items (1,2, or more points) that allow for partial as well as full credit. The exact number of score points and the exact distribution of question types per block varies somewhat.

## Scales for Reporting Student Achievement

TIMSS reports trends in student achievement in both mathematics and science and, to provide additional information for improving curriculum and instruction, reports also on student achievement in the major content domains of each subject. To complement the information provided by the content domain scales, TIMSS intends to report how students perform on the cognitive domains of mathematics and science as well as by content.

As each student's achievement booklet contains only a sample of items from the assessment, student responses are combined for an overall picture of the assessment results for each country. Using item response theory (IRT) methods,<sup>1</sup> individual student responses to the items related to mathematics and science are placed on common scales that link to TIMSS results from 1995, 1999, and 2003. At the eighth grade, the TIMSS 2007 mathematics scale allows countries that participated in

<sup>1</sup> For a description of the TIMSS scaling techniques as applied to the 2003 data, see Gonzalez, E., Galia, J., & Li, I. (2004). Scaling methods and procedures for the TIMSS 2003 mathematics and science scales. In M.O. Martin, I.V.S. Mullis, & S.J. Chrostowski (Eds.), *TIMSS 2003 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

TIMSS in 1995, 1999, or 2003 to track their progress in mathematics achievement, and a comparable science scale provides the same information for science. At the fourth grade, the TIMSS 2007 mathematics and science scales link to 1995 and 2003 but not 1999, since the TIMSS 1999 assessment did not include fourth grade.

In addition to the achievement scales for mathematics and science, TIMSS constructs scales for reporting student performance in each of the mathematics and science content domains. In mathematics at the eighth grade there are four content scales, corresponding to the four content domains in TIMSS 2007:

- Number
- Algebra
- Geometry
- Data and Chance

Eighth-grade science in TIMSS 2007 also has four content domain scales:

- Biology
- Chemistry
- Physics
- Earth Science

At fourth grade, TIMSS 2007 has three content domain scales in mathematics:

- Number
- Geometric Shapes and Measures
- Data Display

Fourth-grade science in TIMSS 2007 also has three content domain scales:

- Life Science
- Physical Science
- Earth Science

As described earlier, the TIMSS 2007 frameworks specify three cognitive domains covering mathematics and science at both fourth and eighth grades. The three domains, knowing, applying, and reasoning, describe a range of cognitive processes involved in working in mathematics and science through the primary and middle school years. Reporting scales are constructed for each of the three domains in mathematics and science at both fourth and eighth grades.

### **Releasing Assessment Material to the Public**

The data collection in 2007 is the fourth in the TIMSS series of regular four-year studies, and provides data on trends in mathematics and science achievement since 1995, 1999, and 2003. TIMSS will be administered again in 2011, 2015, and so on into the future. The design provides for releasing many of the items into the public domain as the international reports are published, while safeguarding the trend data by keeping secure a substantial proportion of the items. As items are released, new items will be developed to take their place.

According to the TIMSS 2007 design, six of the 14 assessment blocks in each subject will be released when the assessment results for 2007 are published, and the remaining eight will be kept secure for use in later assessments. The released blocks will include the two blocks containing trend items from 1999,<sup>2</sup> two of the blocks of trend items from 2003, and two blocks of items used for the first time in 2007. The released items will be replaced with new items before the next survey cycle, in 2011.

<sup>2</sup> Because TIMSS did not assess fourth-grade students in 1999, the TIMSS 2007 fourth-grade released blocks comprise four blocks of items from 2003 and two from 2007.

## Background Questionnaires

An important purpose of TIMSS is to study the educational context in which students learn mathematics and science. To that end, TIMSS will administer questionnaires to curriculum specialists, and to the students in participating schools, their mathematics and science teachers, and their school principals. The questions are designed to measure key elements of the curriculum as it is intended, as it is implemented, and as it is learned.

### CURRICULUM QUESTIONNAIRES

The mathematics and science curriculum questionnaires are designed to collect basic information about the organization of the mathematics and science curriculum in each country, and about the content of these subjects intended to be covered up to the fourth and eighth grades. The National Research Coordinator in each country is responsible for completing the questionnaires, drawing upon the knowledge and expertise of curriculum specialists and educators as necessary.

### STUDENT QUESTIONNAIRE

This questionnaire is completed by each student who takes the TIMSS assessment. It asks about aspects of students' home and school lives, including classroom experiences; self-perception and attitudes about mathematics and science; homework and out-of-school activities; computer use; home educational supports; and basic demographic information. The questionnaire requires about 30 minutes to complete.

### TEACHER QUESTIONNAIRES

In each school participating at the eighth grade, one or more eighth-grade mathematics classes are sampled to take part in the TIMSS testing. The mathematics teachers of these classes complete a mathematics teacher questionnaire, providing information on the teachers' background, beliefs, attitudes, educational preparation, and teaching load, as well as details of the pedagogic approach used in those classes. The science teacher (or teachers) of the students in those classes complete

a science teacher questionnaire, which in many respects parallels the mathematics teacher questionnaire. Both questionnaires ask about characteristics of the classes tested in TIMSS; instructional time, materials, and activities for teaching mathematics and science and promoting students' interest in the subjects; use of computers and the internet; assessment practices; and home-school connections. They also ask teachers their views on their opportunities for collaboration with other teachers and professional development, and for information about themselves and their education and training.

At the fourth grade, a single teacher questionnaire containing questions about mathematics and science instruction and about the teachers' background is completed by the classroom teacher of each sampled fourth grade class. The teacher questionnaires require 30-45 minutes of the teachers' time.

#### SCHOOL QUESTIONNAIRE

The principal of each school in TIMSS completes this questionnaire. It asks about enrollment and staffing; resources available to support mathematics and science instruction; school goals and the role of the principal; instructional time; home-school connections; and school climate. It is designed to take about 30 minutes.