

TIMSS 2011 Assessment Design

Chapter

4



Chapter 4

TIMSS 2011 Assessment Design

Overview

The TIMSS 2011 international assessment of student achievement at fourth and eighth grades comprises written tests in mathematics and science together with sets 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. Conducted on a four-year cycle, with each assessment linked to the one that preceded it, TIMSS provides regular and timely data for educators and policy makers on trends in students' mathematics and science achievement.

In addition to measuring trends in achievement at fourth and eighth grades, administering TIMSS at fourth and eighth grades every four years provides the opportunity to monitor achievement changes within a grade cohort, as the fourth grade students in one TIMSS cycle become the eighth grade students in the next cycle. Further, the TIMSS fourth grade assessment provides data that complement PIRLS, IEA's Progress in International Reading Literacy Study, which assesses reading comprehension at fourth grade every five years. The fifth in the TIMSS series of assessments, TIMSS 2011 will be the first TIMSS assessment to have data collection

in the same school year as PIRLS, providing a rare opportunity for countries to collect internationally comparable information on mathematics, science, and reading at the fourth grade in the same year and on the same students.¹

Student Populations Assessed

TIMSS assesses the mathematics and science achievement of students in their fourth and eighth years of formal schooling. Participating countries may choose to assess one or both populations, according to their policy priorities and resource availability. Because in TIMSS the number of years of formal schooling (four or eight) is the basis for comparison among participating countries, the TIMSS assessment is targeted at the grade levels that correspond to these. The TIMSS target populations are defined as follows.

At the fourth grade, the TIMSS target grade should be the grade that represents four years of schooling, counting from the first year of ISCED Level 1.

At the eighth grade, the TIMSS target grade should be the grade that represents eight years of schooling, counting from the first year of ISCED Level 1.

ISCED is the International Standard Classification of Education developed by the UNESCO Institute for Statistics and provides an international standard for describing levels of schooling across countries. The ISCED system describes the full range of schooling, from pre-primary (Level 0) to the second level of tertiary education (Level 6). ISCED Level 1 corresponds to primary education or the first stage of basic education. The first year of Level 1 should mark the beginning of “systematic apprenticeship of reading, writing and mathematics” (UNESCO, 1999). Four years after this would be the target grade for fourth grade TIMSS, and is the fourth grade in most countries. Similarly, eight years after the first year of ISCED Level 1

¹ Countries participating in PIRLS and TIMSS at the fourth grade in 2011 have the option of administering the assessments to the same students or to separate student samples. Most countries are planning to administer the two assessments to the same students.

is the target grade for eighth grade TIMSS, and is the eighth grade in most countries. However, given the cognitive demands of the assessments, TIMSS wants to avoid assessing very young students. Thus TIMSS recommends that countries assess the next higher grade (i.e., fifth grade for fourth grade TIMSS and ninth grade for eighth grade TIMSS) if, for fourth-grade students, the average age at the time of testing would be less than 9.5 years, and, for eighth-grade students, less than 13.5 years.

Reporting Student Achievement

TIMSS 2011 will provide a comprehensive picture of the mathematics and science achievement of fourth- and eighth-grade students in each participating country. This will include achievement in each of the content and cognitive domains (as defined in Chapters 1 and 2) as well as overall mathematics and science achievement. Consistent with the goal of a comprehensive description of mathematics and science achievement, the complete TIMSS 2011 assessment consists of a large pool of mathematics and science questions (known as items) at each grade level. However, to keep the assessment burden on any one student to a minimum, each student is presented with only a sample of the items, as described in the next section. Following data collection, student responses are placed on common mathematics and science scales at each grade level to provide an overall picture of the assessment results for each country.

One of the major strengths of TIMSS is its measurement of trends over time in mathematics and science achievement. The TIMSS achievement scales² provide a common metric on which countries can compare their fourth and eighth grade students' progress in mathematics and science from assessment to assessment. The TIMSS mathematics and science achievement scales were established in 1995 to have a scale average of 500 and a standard deviation of 100,

2 TIMSS provides separate scales for mathematics and for science at each grade level, as well as for each content and cognitive domain.

corresponding to the international means and standard deviations across all of the countries that participated in TIMSS 1995 at the fourth and eighth grades. Using items that were administered in both 1995 and 1999 assessments as a basis for linking the two sets of assessment results, and working separately for mathematics and science, the TIMSS 1999 data also were placed on the scale so that countries could gauge changes in students' mathematics and science achievement since 1995. Using similar procedures, again separately for mathematics and science, the TIMSS 2003 data³ and the TIMSS 2007 data were placed on the TIMSS scale, as will be the data from TIMSS 2011. This will enable TIMSS 2011 countries that have participated in TIMSS since its inception to have comparable achievement data from 1995, 1999, 2003, 2007, and 2011, and to plot changes in performance over this period.⁴

As previously mentioned, in addition to the achievement scales for mathematics and science overall, TIMSS 2011 will construct scales for reporting relative student performance in each of the mathematics and science content and cognitive domains. More specifically, in mathematics at the fourth grade there will be three content scales, corresponding to the three content domains of number, geometric shapes and measures, and data display and four at the eighth grade—number, algebra, geometry, and data and chance. In science, there also will be three content scales at fourth grade: life science, physical science, and earth science, and four at the eighth grade: biology, chemistry, physics, and earth science. The *TIMSS 2011 Assessment Frameworks* specify three cognitive domains, knowing, applying, and reasoning, which span the mathematics and science content at both grades. Reporting scales will be constructed for each cognitive domain in mathematics and science at each grade level.

3 Because TIMSS 1999 was conducted only at the eighth grade, the TIMSS 2003 fourth grade data were linked directly to the TIMSS 1995 data, omitting the 1999 linking step.

4 Countries that have participated in TIMSS at the fourth grade since the beginning will have comparable data from 1995, 2003, 2007, and 2011.

TIMSS 2011 Student Booklet Design

A major consequence of TIMSS' ambitious reporting goals is that many more questions are required for the assessment than can be answered by any one student in the amount of testing time available. Accordingly, TIMSS 2011 uses a matrix-sampling approach that involves packaging the entire assessment pool of mathematics and science items at each grade level into a set of 14 student achievement booklets, with each student completing just one booklet. Each 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. TIMSS uses item-response theory scaling methods to assemble a comprehensive picture of the achievement of the entire student population from the combined responses of individual students to the booklets that they are assigned.⁵ 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.

To facilitate the process of creating the student achievement booklets, TIMSS groups the assessment items into a series of item blocks, with approximately 10-14 items in each block at the fourth grade and 12-18 at the eighth grade. As far as possible, within each block the distribution of items across content and cognitive domains matches the distribution across the item pool overall. As in the TIMSS 2007 assessment, TIMSS 2011 has a total of 28 blocks, 14 containing mathematics items and 14 containing science items. Student booklets were assembled from various combinations of these item blocks.

Following the 2007 assessment, 8 of the 14 mathematics blocks and 8 of the 14 science blocks were secured for use in measuring

5 The TIMSS scaling methodology is described in detail in Foy, Galia, & Li (2008).

trends in 2011. The remaining 12 blocks (6 mathematics and 6 science) were released into the public domain for use in publications, research, and teaching, to be replaced by newly-developed items for the TIMSS 2011 assessment. Accordingly, the 28 blocks in the TIMSS 2011 assessment comprise 16 blocks of trend items (8 mathematics and 8 science) and 12 blocks of new items developed for 2011. As shown in Exhibit 10, the TIMSS 2011 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 2007 assessment, as do blocks ending in 06. The remaining blocks with labels ending in even numbers contain the items developed for use for the first time in TIMSS 2011.

Exhibit 10: TIMSS 2011 Item Blocks – Fourth and Eighth Grades

Mathematics Blocks	Source of Items	Science Blocks	Source of Items
M01	Block M13 from TIMSS 2007	S01	Block S13 from TIMSS 2007
M02	New items for TIMSS 2011	S02	New items for TIMSS 2011
M03	Block M06 from TIMSS 2007	S03	Block S06 from TIMSS 2007
M04	New items for TIMSS 2011	S04	New items for TIMSS 2011
M05	Block M09 from TIMSS 2007	S05	Block S09 from TIMSS 2007
M06	Block M10 from TIMSS 2007	S06	Block S10 from TIMSS 2007
M07	Block M11 from TIMSS 2007	S07	Block S11 from TIMSS 2007
M08	New items for TIMSS 2011	S08	New items for TIMSS 2011
M09	Block M08 from TIMSS 2007	S09	Block S08 from TIMSS 2007
M10	New items for TIMSS 2011	S10	New items for TIMSS 2011
M11	Block M12 from TIMSS 2007	S11	Block S12 from TIMSS 2007
M12	New items for TIMSS 2011	S12	New items for TIMSS 2011
M13	Block M14 from TIMSS 2007	S13	Block S14 from TIMSS 2007
M14	New items for TIMSS 2011	S14	New items for TIMSS 2011

Fourth-grade students are expected to spend 18 minutes on each item block, and eighth-grade students 22½ minutes, on average. Consequently, the 28 blocks of fourth-grade items are estimated to contain almost 8½ hours of testing time and the eighth-grade blocks about 10½ 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 must fit into 72 minutes for fourth grade and 90 minutes for eighth grade. An additional 30 minutes for a student questionnaire also was planned at each grade level.

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 achievement in the mathematics and science content and cognitive domains 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 2011 booklet design, the 28 assessment blocks are distributed across 14 student achievement booklets (see Exhibit 11). The fourth- and eighth-grade booklet designs are identical, although the fourth-grade blocks contain 18 minutes of assessment items and the eighth grade blocks 22½ 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 two mathematics blocks come first, and then the two science blocks, and in the other half the order is reversed. Additionally, in most booklets two of the blocks contain trend items from 2007 and two contain items newly developed for TIMSS 2011. 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 2007, while those in M02 and S02 are items new for TIMSS 2011. 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 2011 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 consisting of two parts, followed by a student questionnaire. The individual student response burden for the TIMSS 2011 assessment is the same as in 2007, 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 2011 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. As described in the *TIMSS 2011 Item Writing Guidelines* (Mullis & Martin, 2009), 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. 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. In TIMSS, 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. Multiple-choice questions allow valid, reliable, and economical measurement of a wide range of content in a relatively short testing time. However, because they do not allow for students' explanations or supporting statements, these 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. Because they allow students to provide explanations, support an answer with reasons or numerical evidence, draw diagrams, or display data, 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 2011 assessment, IEA has archived samples of student responses to the TIMSS 2007 assessments from each country; these are used to train scorers in 2011 and to monitor consistent application for those items appearing in both assessments.

Score Points. In developing the assessment, the aim is to create blocks of items that each provide, on average, about 15 score points at fourth grade and about 18 score points at eighth 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.

Releasing Assessment Material to the Public

The TIMSS assessment in 2011 is the fifth in the TIMSS series of regular four-year studies, and provides data on trends in mathematics and science achievement since 1995, 1999, 2003, and 2007. TIMSS will be administered again in 2015, 2019, and so on into the future. With each assessment, as the international reports are published, many items are released to provide the public with as much information as possible about the nature and contents of the assessment. At the same time, the measurement of trends is safeguarded 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 2011 design, 6 of the 14 assessment blocks in each subject will be released when the assessment results for 2011 are published, and the remaining 8 will be kept secure for use in later assessments. The released blocks will include three blocks containing trend items from 2003, two blocks of trend items from 2007, and one block of items used for the first time in 2011. The released items will be replaced with new items before the next survey cycle, in 2015.

Background Questionnaires

An important purpose of TIMSS is to identify the procedures and practices that are effective in improving students' learning in mathematics and science. To better understand the contextual factors detailed in Chapter 3 that affect students' learning, TIMSS administers background questionnaires to students, their teachers, and their school principals. TIMSS also administers curriculum questionnaires to specialists to collect information about educational policies and the national contexts that shape the content and implementation of the mathematics and science curricula across countries. Finally,

the TIMSS Encyclopedia provides a more qualitative description of mathematics and science education in the participating countries. For countries participating in both TIMSS and PIRLS at the fourth grade, the *Learning to Read Survey* provides a special opportunity to collect information from students' parents and caregivers on their home backgrounds as well as quantitative readiness.

Student Questionnaire

A questionnaire is completed by each student who takes the TIMSS assessment. This questionnaire asks about aspects of students' home and school lives, including basic demographic information, their home environment, school climate for learning, and self-perception and attitudes toward mathematics and science. While some questions are identical in the fourth- and eighth-grade versions, the language is simplified in the fourth-grade version and specific content is altered to be appropriate for the respective grade level. The student questionnaire requires 15-30 minutes to complete.

Teacher Questionnaires

A teacher questionnaire is completed by the teachers of mathematics and science to the students sampled to take part in the TIMSS testing. This questionnaire is designed to gather information on teacher characteristics as well as the classroom contexts for teaching and learning mathematics and science, and the topics taught in these subjects.

In particular, the teacher questionnaire asks about teachers' backgrounds, their views on opportunities for collaboration with other teachers, their job satisfaction, and their education and training as well as professional development. The questionnaire also collects information on 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, assessment practices, and homework.

The fourth- and eighth-grade versions of the questionnaire are similar, with specific content targeted to teachers at the specific grade level. Although the general background questions are parallel across versions, questions pertaining to instructional and assessment practices, content coverage, and teachers' views about teaching the subject matter are tailored toward mathematics or science. Many questions, such as those related to classroom activities are specific to the classes sampled for TIMSS. This questionnaire requires about 30 minutes of teachers' time to complete.

School Questionnaire

The principal of each school participating in TIMSS is asked to respond to this questionnaire. It asks about school characteristics, instructional time, resources and technology, parental involvement, school climate for learning, teaching staff, the role of the principal, and students' school readiness. It is designed to take about 30 minutes.

Curriculum Questionnaires

The National Research Coordinator in each country is responsible for completing the mathematics and science curriculum questionnaire, drawing on the expertise of curriculum specialists and educators. The questionnaire is 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. It also includes questions on attrition and retention policies, the local or national examination system as well as goals and standards for mathematics and science instruction.

TIMSS 2011 Encyclopedia

The *TIMSS 2011 Encyclopedia* provides context for mathematics and science instruction in the participating countries. Information from the curriculum questionnaire is reported along with information about countries' education systems and policies, including emphasis placed on mathematics and science. In addition, the mathematics and science curriculum in each country is summarized as is information about instructional time and use of instructional materials, equipment, and technology. Teachers' educational training and professional development is also described as well as information about examinations and assessments.

