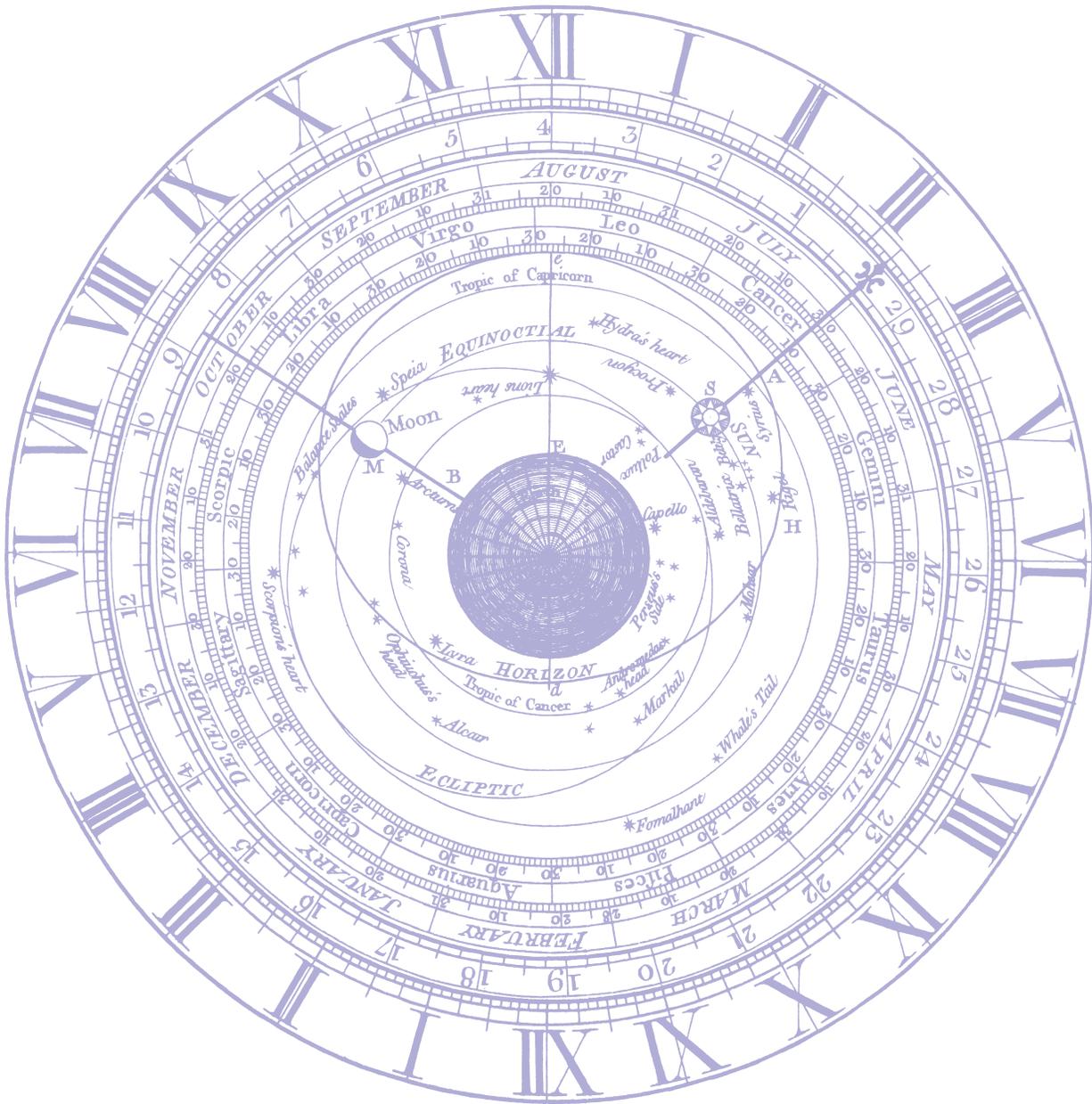


Assessment Design



Assessment Design

Scope of the Assessment

To measure students' achievement in mathematics and science at the fourth and eighth grades and gather information about the contexts for achievement, the TIMSS 2003 assessment includes written tests of mathematics and science and a series of questionnaires focusing on contexts for student learning in those subjects. This chapter describes the design of the assessment and specifications for operationalizing the components of the study. A characteristic of TIMSS is that it includes both mathematics and science, with each student completing parts of the assessment in both subjects.

The TIMSS frameworks have broad coverage goals, and consequently the TIMSS Expert Panel found that a valid assessment of the mathematics and science described in the frameworks would require a substantial pool of assessment items and extensive testing time – at least seven hours at eighth grade (mathematics and science together) and more than five and one half hours at fourth grade. While the assessment material that can be presented in that time should provide good coverage of the mathematics and science students encounter at school and in their everyday lives, it is not reasonable to expect each student to answer the entire pool of test items.

Dividing up the Item Pool

Since the testing time required by the entire assessment item pool greatly exceeds the time available for testing individual students, TIMSS divides the assessment material among students. The TIMSS approach, based on matrix sampling techniques, involves dividing the item pool among a set of student booklets, with each student completing just one booklet. Items are assigned to booklets in such a way

that a comprehensive picture of the achievement of the entire student population can be assembled from the combined responses of individual students to individual booklets.

Based on 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 needed to fit into approximately 90 minutes for eighth grade and approximately 65 minutes for fourth grade. An additional 15-30 minutes for a student questionnaire is also planned at each grade level.

To facilitate the creation of the student booklets, the items in the assessment pool are first grouped into clusters or blocks of items. These then become the building blocks from which the student booklets will be assembled. In TIMSS 2003, the total item pool at each grade level (seven hours of testing at eighth grade and more than five and one half hours at fourth grade) will be divided into 28 blocks, 14 in mathematics and 14 in science, as shown in Exhibit 6. Each block will contain mathematics items only or science items only. Eighth-grade

Exhibit 6: General Design of the TIMSS 2003 Matrix-Sampling Blocks

Source of Items	Mathematics Blocks	Science Blocks
Trend Items (TIMSS 1995 or 1999)	M1	S1
Trend Items (TIMSS 1995 or 1999)	M2	S2
Trend Items (TIMSS 1995 or 1999)	M3	S3
Trend Items (TIMSS 1999)	M4	S4
Trend Items (TIMSS 1999)	M5	S5
Trend Items (TIMSS 1999)	M6	S6
New Replacement Items	M7	S7
New Replacement Items	M8	S8
New Replacement Items	M9	S9
New Replacement Items	M10	S10
New Replacement Items	M11	S11
New Replacement Items	M12	S12
New Replacement Items	M13	S13
New Replacement Items	M14	S14

blocks will contain 15 minutes of assessment items and fourth-grade blocks 12 minutes; otherwise the general design is identical at both grade levels. The blocks containing mathematics items will be labeled M1 through M14 and the science items S1 through S14.

Since TIMSS in 2003 and in later cycles plans to provide an up-to-date assessment of student achievement in mathematics and science while also measuring trends in achievement since 1995 and 1999,¹ the TIMSS design for 2003 includes items from earlier assessments to measure trends as well as innovative new problem-solving and inquiry items and replacement items for those released into the public domain. Of the 14 item blocks in each subject, six (blocks 1 through 6) contain secure items from earlier TIMSS assessments to measure trends,² and eight (blocks 7 through 14) contain new replacement items.

Although calculators were not permitted in 1995 or 1999, calculators may be used in the eighth-grade assessment at the discretion of each participating country.³ Calculators will not be permitted in the fourth-grade assessment.

Block Design for Student Booklets

In choosing how to distribute assessment blocks across student 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, at least some blocks had to be paired with others. Since the number of booklets can become very large if each block is to be paired with all other blocks, it was necessary to choose judiciously among possible block combinations to keep the number of student booklets to a minimum.

The decision to allow calculator use at the eighth grade for the first time in 2003 also had an impact on the booklet design. Since calculators were not allowed in 1995 or 1999 but will be permitted in the eighth-grade assessment in 2003, it was necessary in order to safeguard the measurement of trend to arrange the item blocks in the booklets so that calculators could be used for the new assessment items, but not for the trend items. Accordingly, the trend blocks were placed in the first part of each booklet, to be completed without calculators before the break. However, two mathematics trend blocks (M5 and M6) and two science trend blocks (S5 and S6) also were placed in the second part of one booklet each.

In the TIMSS 2003 design, the 28 assessment blocks will be distributed across 12 student booklets (see Exhibit 7). The same booklet design will be used at both fourth and eighth grade, although the eighth-grade blocks will contain 15 minutes of assessment items and the fourth grade blocks 12 minutes. Each student booklet will consist of six blocks of items. Half the booklets will contain four mathematics

1 TIMSS will measure trends at the eighth grade from 1995 and 1999, but at the fourth grade from 1995 only, since the TIMSS 1999 assessment was conducted at the eighth grade but not the fourth grade.

2 The six trend blocks for the TIMSS 2003 eighth-grade assessment will contain secure items from both the 1995 and 1999 TIMSS assessments, with all trend items from 1995 being placed in blocks 1 through 3. Because there are no fourth-grade items from 1999, trend blocks 4 through 6 for the fourth-grade assessment will consist of new items. In order to create balanced blocks, some new items are also included in blocks 1 through 6 at the eighth grade.

3 To avoid introducing bias into the measurement of trends, calculators will be used only with items new in 2003, and not with items measuring trends.

blocks and two science blocks, and the other half will contain four science blocks and two mathematics blocks. All student booklets will contain at least two blocks of mathematics items and two blocks of science items, so that all students will respond to enough items to provide reliable measurement of trends in both subjects. In addition, the students that are assigned booklets with four mathematics blocks (half of the student sample) will provide sufficient data to measure trends in the mathematics content areas, and the students assigned the booklets with four science blocks will provide data on trends in the science content areas.

As may be seen from Exhibit 7, students assigned Booklet 1 will complete four blocks of mathematics items, M1, M2, M5, and M7, and two blocks of science items, S6 and S7. The items in blocks M1, M2, M5, and S6 will be trend items from earlier TIMSS assessments, while those in M7 and S7 will be new items. Students assigned Booklet 2, 3, 4, 5, or 6 will also have four mathematics blocks and two science blocks, although the blocks will vary from booklet to booklet, as shown in Exhibit 7. Students assigned Booklets 7 through 12 each will complete four blocks of science and two blocks of mathematics items. To enable linking between booklets, all blocks will appear in at least two of the 12 booklets, with the trend blocks appearing in three or four of the booklets. Countries participating in TIMSS will aim for a sample of at least 4,500 students to ensure that there are enough respondents for each item. The 12 student booklets will be rotated among the students in each sampled class, so

Exhibit 7: TIMSS 2003 Booklet Design – Fourth and Eighth Grade

Student Booklet	Assessment Blocks					
Booklet 1	M1	M2	S6	S7	M5	M7
Booklet 2	M2	M3	S5	S8	M6	M8
Booklet 3	M3	M4	S4	S9	M13	M11
Booklet 4	M4	M5	S3	S10	M14	M12
Booklet 5	M5	M6	S2	S11	M9	M13
Booklet 6	M6	M1	S1	S12	M10	M14
Booklet 7	S1	S2	M6	M7	S5	S7
Booklet 8	S2	S3	M5	M8	S6	S8
Booklet 9	S3	S4	M4	M9	S13	S11
Booklet 10	S4	S5	M3	M10	S14	S12
Booklet 11	S5	S6	M2	M11	S9	S13
Booklet 12	S6	S1	M1	M12	S10	S14

that approximately equal proportions of students respond to each booklet.

As summarized in Exhibit 8, each student will complete just one of the 12 student booklets and a student questionnaire. The individual student response burden for the assessment is similar to TIMSS in 1995 and 1999, i.e., 72 minutes for the assessment and 30 minutes for the questionnaire at fourth grade (slightly more than the 65 minutes for the assessment in 1995), and 90 minutes and 30 minutes, respectively, at eighth grade. Students may not use calculators while working on Part 1 of the student booklet, as this

Exhibit 8: TIMSS 2003 Student Testing Time

Activity	Fourth Grade	Eighth Grade
Student Booklet – Part 1 (Calculators Not Permitted)	36 minutes	45 minutes
Break		
Student Booklet – Part 2 (Calculators Permitted – 8th grade only)	36 minutes	45 minutes
Break		
Student Questionnaire	30 minutes	30 minutes

contains the trend items. Calculators may be used for Part 2 of the student booklet, but only by eighth-grade students. Calculators are not permitted in the fourth-grade assessment.

Question Types and Scoring Procedures

Students' knowledge and understanding of mathematics and science will be assessed through a range of questions in each subject. Two question formats will be used in the TIMSS assessment – multiple-choice and constructed-response. Each multiple-choice question will be worth one point. Constructed-response questions generally will be worth one, two, or three points, depending on the nature of the task and the skills required to complete it. Extended problem-solving and inquiry tasks may require students to work with materials and manipulatives and to provide one or more extended constructed responses. Up to two-thirds of the total number of points represented by all the questions will come from multiple-choice questions. In developing assessment questions, the choice of item format will depend on the mathematics or science being assessed, and the format that will best enable students to demonstrate their proficiency.

Multiple-Choice Questions. Multiple-choice questions provide students with four or five 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 are also 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.

In the TIMSS assessment, constructed-response questions will be worth one, two, or three points, depending on the nature of the task and the extent of the explanation the question requires. In these questions, it is important to provide enough information to help students understand clearly the nature of the response expected.

Scoring guides for each constructed-response question describe 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 2003 assessment, IEA has archived samples of student responses from each country; these will be used to train scorers in 2003 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. For example, at eighth grade blocks 1 through 14 in each subject could be made up of approximately 8 multiple-choice items (1 point each), 2 or 3 short constructed-response items (1 or 2 points each), and 1 extended constructed-response item (3 points). The exact number of score points and the exact distribution of question types per block will vary somewhat. Since the blocks for the fourth-grade assessment will be designed to yield 12 rather than 15 score points, there will be fewer items but the relative proportions of different item types will be approximately the same.

Scales for Reporting Student Achievement

TIMSS will report trends in student achievement in both the general areas of mathematics and science and in the major subject matter content areas. As each student will respond to only part of the assessment, these parts must be combined for an overall picture of the assessment results for each country. Using item response theory (IRT) methods,⁴ individual student responses to the items related to mathematics and science will be placed on common scales that link to TIMSS results from 1995 and 1999. At the eighth grade, there will be an overall mathematics scale that will allow countries that participated in TIMSS in 1995 or 1999 to track their progress in mathematics achievement since then, and a similar scale in science overall that will provide the same information for science. At the fourth grade, the overall mathematics and science scales will link to 1995 only, since the TIMSS 1999 assessment did not include fourth grade. All students will have overall mathematics and science scores.

All student responses will contribute to the measurement of achievement in each of the mathematics and science content areas. In addition, those students assigned booklets with four blocks of mathematics items (half of the student sample) will provide the data to report on trends in mathematics content areas, while those assigned booklets with four blocks of science items (the other half) will provide data on trends in science content areas.

4 For a description of the TIMSS scaling techniques as applied to the 1999 data, see Yamamoto, K. and Kulick, E. (2000), "Scaling Methods and Procedures for the TIMSS Mathematics and Science Scales" in M.O. Martin, K.D. Gregory, and S.E. Stemler (eds.), *TIMSS 1999 Technical Report*, Chestnut Hill, MA: Boston College.

In mathematics at the eighth grade there will be five content reporting categories in 2003:

- Number
- Algebra
- Measurement
- Geometry
- Data

At fourth grade there also will be five content reporting categories in mathematics:

- Number
- Patterns, Equations, and Relationships
- Measurement
- Geometry
- Data

Eighth-grade science will have five content reporting categories:

- Life Science
- Chemistry
- Physics
- Earth Science
- Environmental Science

At fourth grade, science will have just three content reporting categories:

- Life Science
- Physical Science
- Earth Science

Results will be reported separately for each content area and grade level.

In addition to the IRT scales that will be used to summarize achievement in mathematics and science content areas and in these subjects overall, TIMSS will report on performance in

each of the cognitive domains in terms of the average percentage of students answering items correctly in each domain. This approach may also be used to report student performance in scientific inquiry.

Releasing Assessment Material to the Public

The data collection in 2003 will be the third in the TIMSS series of regular four-year studies, and will provide data on trends in mathematics and science achievement since 1995 and 1999. TIMSS will be administered again in 2007, 2011, 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 design, half of the 14 assessment blocks in each subject will be released when the assessment results for 2003 are published, and half will be kept secure for use in later assessments. The released blocks will include the three blocks containing trend items from 1995, one block of trend items from 1999, and three blocks of items used for the first time in 2003.⁵ The released items will be replaced with new items before the next survey cycle, in 2007.

Background Questionnaires

An important purpose of TIMSS is to study the educational context within 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,

⁵ Since there are no 1999 trend items at fourth grade, four blocks of 2003 items will be released at this grade level.

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

Student Questionnaire. This questionnaire will be 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 15-30 minutes to complete.

Teacher Questionnaires. In each school participating at the eighth grade, a single eighth-grade mathematics class will be sampled to take part in the TIMSS testing. The mathematics teacher of that class will be asked to complete a mathematics teacher questionnaire, providing information on the teacher's background, beliefs, attitudes, educational preparation, and teaching load, as well as details of the pedagogic approach used in that class. The science teacher (or teachers) of the students in that class will be asked to complete a science teacher questionnaire, which in many respects will parallel the mathematics questionnaire. Both questionnaires ask about characteristics of the class 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, there will be a single teacher questionnaire containing questions about mathematics and science instruction and about the teachers' background that will be completed by the classroom teacher of the sampled fourth-grade class. The teacher questionnaires require 30-45 minutes of the teacher's time.

School Questionnaire. The principal of each school in TIMSS will be asked to respond to this questionnaire. It asks about enrollment and staffing; resources available to support mathematics and science instruction, such as the availability of instructional materials and staff; school goals and the role of the principal; instructional time; home-school connections; and school climate. It is designed to take about 30 minutes.

