2

Chapter 2

Developing the TIMSS 2003 Mathematics and Science Assessment and Scoring Guides

Teresa Smith Neidorf and Robert Garden

2.1 Overview

The development of the TIMSS 2003 mathematics and science assessment was a collaborative process spanning a two-and-a-half-year period, from September 2000 to March 2003, and involving mathematics and science educators and development specialists from all over the world. The work began with a major updating and revision of the existing TIMSS assessment frameworks to address changes during the last decade in curricula and the way mathematics and science are taught (Mullis, Martin, Smith, Garden, Gregory, Gonzalez, Chrostowski, & O'Connor, 2003). The assessment development work was based firmly on the new assessment frameworks and specifications.

Meeting the specifications of the TIMSS 2003 assessment frameworks required a large number of new mathematics and science items to be developed at both fourth and eighth grades. With support and training from the TIMSS International Study Center, National Research Coordinators (NRCs), from participating countries contributed a large pool of items for review and field testing. The International Study Center established two task forces, one in mathematics and one in science,[1] to manage the item development process. To help review, select, and revise items for the assessment and to ensure their mathematical and scientific accuracy, the International Study Center convened the Science and Mathematics Item Review Committee (SMIRC), an international committee of prominent mathematics and science experts

1 The mathematics task force consisted of Robert Garden, TIMSS Mathematics Coordinator, Chancey Jones of Educational Testing Service in the United States, and Graham Ruddock of the National Foundation for Educational Research in England. The science task force consisted of Teresa Smith Neidorf, TIMSS Science Coordinator, Christine O'Sullivan, formerly the Science Coordinator for the U.S. National Assessment for Educational Progress (NAEP), and Svein Lie, University of Oslo, formerly the chair of the TIMSS 1995 Subject Matter Advisory Committee.

nominated by participating countries and representing a range of nations and cultures.[2]

Since the test items were developed in English and translated into 34 languages by the participating countries, both the SMIRC and NRCs were important in identifying any items that might prove difficult to translate consistently.

To ensure that TIMSS 2003 reflects an international perspective, both the framework and test development procedures included substantial contributions from the international community. Exhibit 2.1 provides an overview of the process. This chapter describes the steps taken in developing the TIMSS 2003 mathematics and science assessment, with sections covering frameworks development, the mathematics and science assessment specifications, development of mathematics and science items and scoring guides, and the assessment booklet design.

2.2 Developing the TIMSS 2003 Assessment Frameworks

For the TIMSS 2003 assessment, the curriculum frameworks used as the basis for the 1995 and 1999 TIMSS assessments (Robitaille, McKnight, Schmidt, Britton, Raizen, & Nicol, 1993) were extensively revised and updated. This effort was conducted by the TIMSS International Study Center at Boston College in collaboration with the National Research Coordinators of the TIMSS countries and with guidance from an international Expert Panel. The Expert Panel was made up of 29 internationally recognized experts and included mathematicians and scientists, curriculum experts, and educational practitioners, researchers, and assessment specialists.[3]

The framework development process took approximately one year, beginning in September 2000. Work to update the frameworks began with a review of the TIMSS 1999 curriculum data to identify mathematics and science topics emphasized in the curricula of the TIMSS countries. In addition, a survey of NRCs of more than 20 countries planning to participate in TIMSS 2003, administered in September 2000, provided recommendations for the percentage of the TIMSS 2003 assessment to be devoted to each mathematics and science content area at fourth and eighth grades and to identify any curriculum areas that should receive greater or less emphasis than in the TIMSS 1999 assessment. The TIMSS International Study Center used the results of the review and survey to prepare initial framework discussion documents for the First Expert Panel Meeting.

2 See Appendix A for a list of the members of the Science and Mathematics Item Review Committee.

3 See Appendix A for a list of the members of the Expert Panel.

Exhibit 2.1 **Overview of the TIMSS 2003 Framework and Test Development Process**

Date(s)		Group and Activity
September	2000	National Research Coordinators Complete preliminary survey on recommended coverage of mathematics and science content in the TIMSS 2003 assessment.
September – October	2000	TIMSS International Study Center Review results from TIMSS 1999 curriculum questionnaires and TIMSS 2003 NRC survey; prepare initial framework discussion documents for First Expert Panel Meeting.
November	2000	First Expert Panel Meeting (Boston) Make recommendations for coverage of major content and cognitive domains, updated assessment topics within content areas, and initial draft of TIMSS mathematics and science assessment frameworks.
December	2000	TIMSS International Study Center Develop detailed specifications for assessment topics at fourth and eighth grades and prepare first draft of TIMSS assessment frameworks.
February	2001	First National Research Coordinators Meeting (Hamburg) Review first draft of TIMSS assessment frameworks.
March – April	2001	National Research Coordinators Complete survey of Mathematics and Science Curriculum Topics.
April – May	2001	TIMSS International Study Center Compile TIMSS 2003 Mathematics and Science Curriculum Topics survey results and prepare second draft of frameworks.
May	2001	Second Expert Panel Meeting (Amsterdam) Review and approve second draft of TIMSS 2003 Assessment Frameworks incorporating revisions from the first NRC meeting and results of mathematics and science frameworks topic survey completed by National Research Coordinators; generate preliminary ideas for problem-solving and inquiry tasks.
June	2001	Second National Research Coordinators Meeting (Montreal) Review and approve final draft of TIMSS 2003 Assessment Framework. TIMSS Item-Writing Workshop
June – July	2001	National Research Coordinators Develop and submit items to the International Study Center.
August – September	2001	Mathematics and Science Task Forces Assemble, review and revise international item pool; develop additional items to cover framework.
September	2001	TIMSS International Study Center Publish first edition of the TIMSS Assessment Frameworks and Specifications 2003.
September	2001	First Science and Mathematics Item Review Committee Meeting (Boston) Review/refine international item pool; generate prototype ideas for problem-solving and inquiry tasks.

Exhibit 2.1 **Overview of the TIMSS 2003 Framework and Test Development Process**
(…Continued)

Date(s)		Group and Activity
October	2001	Second Science and Mathematics Item Review Committee Meeting (Portsmouth) Review, revise and select preferred and alternate items for field test; develop problem-solving and inquiry tasks.
October – December	2001	Mathematics and Science Task Forces Assemble draft field test item blocks and problem-solving and inquiry tasks.
December	2001	Third National Research Coordinators Meeting (Madrid) Review and approve field test item blocks and problem-solving and inquiry tasks.
December – January	2002	TIMSS International Study Center Conduct teacher review of problem-solving and inquiry tasks; incorporate final revisions to items and tasks based on NRC and teacher reviews.
January – February	2002	TIMSS International Study Center Conduct small-scale item trial of constructed-response items and problem-solving and inquiry tasks; distribute field test instruments; update scoring guides; prepare field test scoring training materials.
February – April	2002	National Research Coordinators: Translate field test instruments. IEA: Verify field test translations.
March	2002	Fourth National Research Coordinators Meeting (Ghent) Field test scoring training
April	2002	U.S. National Center for Education Statistics Cognitive Laboratory Investigation of Problem-Solving and Inquiry Tasks
April – June	2002	Field test administration
June – July	2002	TIMSS International Study Center Review field test item statistics; revise problem-solving and inquiry tasks; assemble draft main survey item blocks.
July	2002	Third Science and Mathematics Item Review Committee Meeting (Oslo) Review field test results and draft item blocks and scoring guides for main survey.
August	2002	Fifth National Research Coordinators Meeting (Tunis) Review and approve item blocks and scoring guides for main survey.
September	2002	TIMSS International Study Center Conduct small-scale trial of final problem-solving and inquiry tasks; distribute main survey instruments; update main survey scoring guides; prepare main survey scoring training materials.
September – October	2002	National Research Coordinators (southern hemisphere): Translate main survey test instruments. IEA: Verify main survey translations for southern hemisphere countries.
October – December	2002	Main survey administration in southern hemisphere countries
November	2002	Southern hemisphere scoring training for the main survey (Wellington)

Exhibit 2.1 **Overview of the TIMSS 2003 Framework and Test Development Process**
(...Continued)

Date(s)		Group and Activity
December	2002	TIMSS International Study Center Update TIMSS Assessment Frameworks and Specifications document to include example items and from the field test and a revised test booklet design; distribute final version of main survey scoring guides.
December – March	2003	National Research Coordinators (northern hemisphere): Translate main survey test instruments. IEA: Verify main survey translations.
February	2003	Publish second edition of the TIMSS Assessment Frameworks and Specifications 2003.
March	2003	Sixth National Research Coordinators Meeting (Bucharest) Northern hemisphere scoring training for the main survey
March – June	2003	Main survey administration in northern hemisphere countries

At its first meeting, in November 2000, the Expert Panel made recommendations concerning how the assessment time in TIMSS 2003 should be distributed across the mathematics and science content areas at each grade level; made suggestions for the major assessment topics that should be included; and discussed calculator usage and the inclusion of scientific inquiry in the 2003 assessment. They also discussed how the "performance expectations" aspect of the original frameworks might be reformulated as a set of broadly-defined cognitive domains for mathematics and science. Maintaining alignment of the TIMSS 2003 content domains with the reporting categories in TIMSS 1995 and 1999 and the measurement of trend were important considerations. Following the meeting, the International Study Center prepared a first draft of the mathematics and science assessment frameworks incorporating the recommendations of the Expert Panel for review by the NRCs.

The first draft of the frameworks contained initial recommendations for the distribution of the assessment across content and cognitive domains and specific assessment objectives for a broad range of mathematics and science topics at the fourth and eighth grade. The draft frameworks document was reviewed and discussed at the first meeting of TIMSS 2003 National Research Coordinators in February 2001, with representatives from more than 40 countries. Some adjustments to the distributions of assessment time across content and cognitive domains were made in response to NRC suggestions. NRCs also gave input on the appropriateness of the assessment topics for the populations of students being assessed.

Following the meeting, four extensive assessment topic questionnaires (mathematics and science at fourth and eighth grades) were distributed to each NRC to be completed with the assistance of experts in mathematics and science curriculum in each country. The questionnaires asked countries to indicate for every mathematics and science assessment topic in the draft frameworks: i) if the topic is addressed by their curriculum at the appropriate grade level, and ii) whether the topic should be included in the TIMSS 2003 international assessment (even if the topic has not been included in their curriculum by that grade level). Results were obtained from 36 countries and were used to refine the set of topics in the frameworks, focusing on those that were included in the curricula or recommended for inclusion in TIMSS by a significant number of participating countries. In the retained set, nearly all topics were included in the curricula of the majority of countries, and for many topics in more than 90 percent of countries.

A second draft of the frameworks document, incorporating the results of the NRC survey, was further refined and improved at the Second Expert Panel Meeting in May 2001. In July 2001, the International Study Center distributed a third draft of the frameworks to the Expert Panel members and the NRCs for review. Comments and suggestions on this draft were incorporated into the final version, *TIMSS Assessment Frameworks and Specifications 2003* (Mullis et al., 2001), published September 2001. A second edition of the frameworks, incorporating example mathematics and science items from the field test and a revised test booklet design, was published in February 2003 (Mullis et al., 2003). During the process of updating the TIMSS assessment frameworks for 2003, the expert panelists and national representatives reaffirmed the importance of emphasizing problem solving, reasoning and inquiry in the outcomes to be assessed, and this is reflected in the final version of the frameworks.

2.3 Mathematics Assessment Framework and Specifications

The mathematics assessment framework for TIMSS 2003 is framed by two organizing dimensions, a content dimension and a cognitive dimension, analogous to those used in the earlier TIMSS assessments. There are five content domains: number, algebra, measurement, geometry, and data. There are four cognitive domains: knowing facts and procedures, using concepts, solving routine problems, and reasoning. The two dimensions and their domains are the foundation of the mathematics assessment. The content domains define the specific mathematics subject matter covered by the assessment, and the cognitive domains define the sets of behaviors expected of students as they engage with the mathematics content. Exhibit 2.2 shows the target percentages of the total mathematics assessment time to be devoted to each of the content and cognitive domains at fourth and eighth grades.

Exhibit 2.2 **Target Percentages of TIMSS 2003 Mathematics Assessment Devoted to Content and Cognitive Domains by Grade Level**

	Grade 4	Grade 8
Mathematics Content Domains		
Number	40%	30%
Algebra*	15%	25%
Measurement	20%	15%
Geometry	15%	15%
Data	10%	15%
Mathematics Cognitive Domains		
Knowing Facts and Procedures	20%	15%
Using Concepts	20%	20%
Solving Routine Problems	40%	40%
Reasoning	20%	25%

* At fourth grade, the algebra content domain is called patterns and relationships.

2.3.1 Content Domains

For each of the five content domains, the mathematics framework identifies several topic areas to be included in the assessment, as shown in Exhibit 2.3. For example, *number* is further categorized by *whole numbers, fractions and decimals, integers,* and *ratio, proportion, and percent.* Each topic area is presented as a list of objectives covered in a majority of participating countries, at either fourth or eighth grade. The organization of topics across the content domains reflects some minor revision in the reporting categories used in the 1995 and 1999 assessments. However, each of the trend items from 1995 and 1999 may be mapped directly into the content domains defined for 2003.

2.3.2 Cognitive Domains

To respond correctly to TIMSS test items, students need to be familiar with the mathematics content of the items. Just as important, however, items were designed to elicit the use of particular cognitive skills. The assessment framework presents detailed descriptions of the skills and abilities that make up the cognitive domains and that will be assessed in conjunction with the content. These skills and abilities should play a central role in developing items and achieving a balance in learning outcomes assessed by then items in fourth and eighth grades. The student behaviors used to define the mathematics framework have been classified into four cognitive domains, as follows:

Exhibit 2.3 Main Topics Included in the Mathematics Content Domains

Content Domains	Main Topics
Number	Whole numbers Fractions and decimals Integers (grade 8 only) Ratio, proportion, and percent
Algebra	Patterns Algebraic expressions (grade 8 only) Equations and formulas Relationships
Measurement	Attributes and units Tools, techniques, and formulas
Geometry	Lines and angles Two- and three-dimensional shapes Congruence and similarity Locations and spatial relationships Symmetry and transformations
Data	Data collection and organization Data representation Data interpretation Uncertainty and probability (grade 8 only)

Knowing Facts and Procedures: *Facts* encompass the factual knowledge that provide the basic language of mathematics and the essential mathematical facts and properties that form the foundation for mathematical thought. *Procedures* form a bridge between more basic knowledge and the use of mathematics for solving routine problems, especially those encountered by people in their daily lives. Students need to be efficient and accurate in using a variety of computational procedures and tools.

Using Concepts: Familiarity with mathematical concepts is essential for the effective use of mathematics for problem solving, for reasoning, and thus for developing mathematical understanding. Knowledge of concepts enables students to make connections between elements of knowledge, make extensions beyond their existing knowledge, and create mathematical representations.

Solving Routine Problems: Problem solving is a central aim of teaching school mathematics and features prominently in school mathematics textbooks. Routine problems may be standard in classroom exercises designed to provide practice in particular methods or techniques. Some of these problems may be set in a quasi-real context, and may involve extended knowledge of mathematical properties (e.g., solving equations). Though they range in difficulty, routine problems are expected to be sufficiently familiar to students that they essentially involve selecting and applying learned procedures.

Reasoning: Mathematical reasoning involves the capacity for logical, systematic thinking. It includes intuitive and inductive reasoning based on patterns and regularities that can be used to arrive at solutions to non-routine problems, i.e., problems very likely to be unfamiliar to students. Such problems may be purely mathematical or may have real-life settings, and involve application of knowledge and skills to new situations, with interactions among reasoning skills usually a feature.

Examples of the behaviors associated with each of the cognitive domains may be found in Mullis et al. (2003).

2.3.3 Communicating Mathematically

Communicating mathematical ideas and processes is important for many aspects of living and fundamental to the teaching and learning of mathematics. In the TIMSS framework, communication is not a separate cognitive domain but rather an overarching dimension across all mathematics content areas and processes. Communication is fundamental to each of the four TIMSS cognitive domains (*knowing facts and procedures, using concepts, solving routine problems, and reasoning*), and students' communication in and about mathematics should be regarded as assessable in each of these areas. Students in TIMSS may demonstrate communication skills through description and explanation, such as describing or discussing a mathematical object, concept, or model. Communication also occurs in using mathematical terminology and notation, demonstrating the procedure used in solving an equation, or using particular representational modes to present mathematical ideas.

2.3.4 Calculator Policy

The TIMSS policy on calculator use at the eighth grade is to give students the best opportunity to operate in settings that mirror their classroom experience. Beginning with 2003, calculators were permitted but not required for newly-developed eighth-grade assessment materials. Participating countries could decide whether or not their students were allowed to use calculators for the new items. Since calculators were not permitted at the eighth grade in the 1995 or 1999 assessments, the 2003 eighth-grade test booklets were designed so that items from these assessments were placed in the first half and items new in 2003 placed in the second half. Where countries chose to permit eighth-grade students to use calculators, they could use them for the second half of the booklet only. For the fourth-grade assessment, TIMSS 2003 continued the 1995 policy of not permitting calculator use.

2.4 Science Assessment Framework and Specifications

The science assessment framework for TIMSS 2003, like the mathematics framework, is framed by two organizing dimensions, a content dimension and a cognitive dimension. There are five content domains: life science, chemistry, physics, earth science, and environmental science, and three cognitive domains: factual knowledge, conceptual understanding, and reasoning and analysis. Exhibit 2.4 shows the target percentages of the total science assessment time to be devoted to each of the science content and cognitive domains for fourth and eighth grades. In contrast to TIMSS 1999, where a separate reporting category of "Scientific Inquiry and the Nature of Science" was included, the TIMSS 2003 framework treats scientific inquiry as a separate assessment strand that overlaps all of the fields of science and has both content- and skills-based components. Although scientific inquiry is not treated as a separate reporting category in TIMSS 2003, the framework specifies that outcomes related to scientific inquiry will represent up to 15 percent of the total science assessment time at each grade level to permit some level of reporting student performance in this area. Further descriptions of the assessment specifications for the content domains, cognitive domains, and scientific inquiry assessment strand are provided in the following sections.

Exhibit 2.4 **Target Percentages of TIMSS 2003 Science Assessment Devoted to Content and Cognitive Domains by Grade Level**

	Grade 4	Grade 8
Science Content Domains		
Life Science	45%	30%
Physical Science	35%	*
Chemistry	*	15%
Physics	*	25%
Earth Science	20%	15%
Environmental Science	*	15%
Science Cognitive Domains		
Factual Knowledge	40%	30%
Conceptual Understanding	35%	35%
Reasoning and Analysis	25%	35%

* At fourth grade, Physical Science included Physics and Chemistry topics. Also, a few Environmental Science topics that addressed the use of conservation of natural resources and changes in environments were included in Earth Science and Life Science.

2.4.1 Content Domains

For each of the science content domains, the framework identifies several main topic areas that are to be included in the assessment as shown in Exhibit 2.5. Most of the main topics are appropriate for both grades, but some topics are included at the eighth grade only, as indicated. For each main topic area, the frameworks document includes a list of specific subtopics or assessment objectives appropriate for each grade level. This structure of the frameworks highlights the development of knowledge and abilities across the grades.

Exhibit 2.5 **Main Topics Included in the Science Content Domains**

Content Domain	Main Topics
Life Science	Types, characteristics, and classification of living things
	Structure, function, and life processes in organisms
	Cells and their functions (grade 8 only)
	Development and life cycles of organisms
	Reproduction and heredity
	Diversity, adaptation, and natural selection
	Ecosystems
	Human health
Chemistry	Classification and composition of matter
	Particulate structure of matter (grade 8 only)
	Properties and uses of water
	Acids and bases (grade 8 only)
	Chemical change
Physics	Physical states and changes in matter
	Energy types, sources and conversions
	Heat and temperature
	Light
	Sound and vibration (grade 8 only)
	Electricity and magnetism
	Forces and motion
Earth Science	Earth’s structure and physical features
	Earth’s processes, cycles and history
	Earth in the solar system and the universe
Environmental Science	Changes in population (grade 8 only)
	Use and conservation of natural resources
	Changes in environments

2.4.2 Cognitive Domains

The set of skills and abilities to be demonstrated by students in responding to items across the science topics is organized into the three broad cognitive domains specified in the framework – factual knowledge, conceptual understanding, and reasoning and analysis. The exact nature of behaviors elicited by the TIMSS items in each of these categories varies between fourth and eighth grade in accordance with the increased cognitive ability, maturity, instruction, experience, and conceptual understanding of students at the higher grade level. A brief description of each cognitive domain and the set of skills and abilities required by TIMSS items corresponding to each are listed below.

Factual Knowledge: This refers to students' knowledge base of relevant science facts, information, tools, and procedures. Items may require students to recall/recognize accurate statements about science facts and concepts; demonstrate knowledge/use of correct scientific terms; describe scientific processes, properties, characteristics, structure, function, and relationships; and demonstrate knowledge about the use of scientific tools and procedures.

Conceptual Understanding: Students should be able to demonstrate a grasp of the relationships that explain the physical world and relate the observable to more abstract or general concepts. Items may require students to provide examples to illustrate general concepts; compare/contrast and classify objects, materials and organisms; use diagrams/models; relate underlying concepts to observed or inferred properties/behaviors; extract/apply textual, tabular or graphical information; find solutions to problems involving the direct application of concepts; and provide explanations.

Reasoning and Analysis: This includes problem-solving and scientific reasoning processes involved in the more complex tasks related to science. Items may require students to analyze/interpret problems; integrate/synthesize a number of factors or related concepts across mathematics and science; hypothesize/predict; design investigations and procedures; analyze/interpret data; draw conclusions; generalize; evaluate; and justify explanations and problem solutions.

2.4.3 Scientific Inquiry

The scientific inquiry strand is assessed through longer problem-solving and inquiry tasks as well as some individual items that require students to apply scientific inquiry skills in a practical context. While not full scientific investigations, the tasks are designed to require a basic understanding of the nature of science and investigation and elicit some of the skills essential to the scientific inquiry process. Tasks may include some portion of the following major phases in the scientific inquiry process:

- Formulating questions and hypotheses
- Designing investigations
- Collecting, representing, analyzing, and interpreting data
- Drawing conclusions and developing explanations based on evidence

The same general assessment outcomes related to scientific inquiry are appropriate for both fourth and eighth grades, but the specific understandings and abilities to be demonstrated increase in complexity across grades. The items and tasks developed to measure scientific inquiry skills are set in content-based contexts. These items are, therefore, classified with respect to content and cognitive categories as well as scientific inquiry and will contribute to the appropriate content reporting scale.

2.5 Developing Mathematics and Science Items and Scoring Guides

Test development for TIMSS 2003 involved developing a set of items aligned with the *TIMSS Assessment Frameworks and Specifications* in each mathematics and science content and cognitive domain. In addition to the target percentages of assessment time to be devoted to the mathematics and science content and cognitive domains, the frameworks give guidelines for the distribution of testing time across item formats, specifying that at least one-third of the assessment should come from constructed-response items. Since approximately half of the eighth-grade items from TIMSS 1999 and one-third of the fourth-grade items from TIMSS 1995 had been kept secure and were to be included, these trend items were taken into account in allocating the test development effort to the different assessment areas for TIMSS 2003. Item development blueprints, specifying the approximate number of mathematics and science items to be developed in each content area in the frameworks, formed the basis for test development for TIMSS 2003. These blueprints were created by:

- estimating the number of items needed in the final test based on the total score points and percentage of score points in each content domain specified in the frameworks,
- distributing this number of items across the mathematics and science main topic areas in accordance with their breadth of content,
- accounting for the number of trend items already included in each topic area,
- ensuring coverage of the cognitive domains and appropriate numbers of multiple-choice and constructed-response items, and
- scaling up the number of items to be developed to allow for attrition during the item selection and field-testing process.

This section describes the test development procedure, including the consideration of trend items, development of the international item pool including problem-solving and inquiry tasks, item review and revision, field testing, item selection for the main survey, and the development of scoring guides for the constructed-response items.

2.5.1 Trend Items

In developing the TIMSS 2003 test blueprints, the trend items from 1995 and 1999 were mapped into the content and cognitive categories in the new 2003 frameworks. As shown in Exhibits 2.6 and 2.7, the mathematics and science trend items cover a range of content domains at both grades. Eighth-grade trend items include both multiple-choice and constructed-response items, while fourth-grade trend items are nearly all multiple-choice. Therefore, a larger proportion of constructed-response items needed to be developed for grade 4.

Exhibit 2.6 **Mathematics Trend Items at Grade 4 and Grade 8 by Content Domain and Item Format**

	Grade 4 Trend Items			Grade 8 Trend Items		
Content Domain	**Multiple Choice**	**Constructed Response**	**Total**	**Multiple Choice**	**Constructed Response**	**Total**
Number	19	0	19	19	6	25
Algebra*	2	0	2	11	5	16
Measurement	8	0	8	8	8	16
Geometry	4	0	4	11	1	12
Data	4	0	4	10	0	10
Total	**37**	**0**	**37**	**59**	**20**	**79**

* Called Patterns and Relationships at Grade 4.

Exhibit 2.7 **Science Trend Items at Grade 4 and Grade 8 by Content Domain and Item Format**

	Grade 4 Trend Items			Grade 8 Trend Items		
Content Domain	**Multiple Choice**	**Constructed Response**	**Total**	**Multiple Choice**	**Constructed Response**	**Total**
Life Science	11	1	12	12	5	17
Physical Science	9	0	9	--	--	--
Chemistry	--	--	--	13	1	14
Physics	--	--	--	14	8	22
Earth Science	11	1	12	10	2	12
Environmental Science	--	--	--	4	5	9
Total	**31**	**2**	**33**	**53**	**21**	**74**

2.5.2 Developing the International Item Pool for TIMSS 2003

Test development for TIMSS 2003 was an international collaborative process, involving participants from more than 30 countries. To maximize the effectiveness of the contributions from national centers, the International Study Center developed a detailed item-writing manual and conducted a workshop for countries that wished to provide items for the international item pool. At this workshop, the mathematics and science task forces reviewed general item-writing guidelines for multiple-choice and constructed-response items and provided specific training in writing mathematics and science items in accordance with the *TIMSS Assessment Frameworks and Specifications 2003*. After the training sessions, participants were organized into item-writing subgroups by mathematics and science content domains for the development and review of items. Nearly 200 draft items were developed at the item-writing workshop.

Following the workshop, national centers developed additional items in mathematics and/or science for the fourth or eighth grade in accordance with their interest and capacity. To maximize contributions from international item writers and ensure adequate item development in the appropriate mathematics and science content areas, some specifications were given by the International Study Center to focus item development in areas not already covered by the trend items. Draft items were submitted by the national centers to the International Study Center, which coordinated the contributions from participating countries and managed the overall test development and review process to ensure that the TIMSS tests were aligned with the assessment frameworks.

Each item from the national centers was submitted with an item-writing form that identified the portion of the framework that the item was designed to assess – content domain, main topic and specific assessment objective, and the primary cognitive domain. Science items also were designated as to whether or not they were intended to measure knowledge and skills associated with the scientific inquiry strand. This development process resulted in an initial item pool of more than 1300 items across both grades, with contributions from 35 countries covering a broad range of mathematics and science topics.

2.5.3 Item Review and Revision

The mathematics and science task forces assembled, reviewed, and revised the draft items submitted by participating countries and confirmed the classification of items with respect to the frameworks. They also developed additional items for areas of the frameworks not well covered by the country submissions. The resultant item pool of more than 2000 items covered a wide array of topics in the mathematics and science content domains at each grade level and reflected the range of cognitive domains and item types specified in the frameworks. The task forces then made a preliminary selection from among these draft items for review by the Science and Mathematics Item Review Committee (SMIRC).

The SMIRC conducted its initial item review work in two meetings, the first in September and the second in October 2001. Working from test development blueprints identifying the number of items needed in each content domain, the SMIRC made much progress in choosing among alternative items, refining the most promising items, and supplementing this set of items in content areas lacking coverage.

Between the second SMIRC meeting and the third NRC meeting in December 2001, the mathematics and science task forces continued the work of developing, reviewing and revising the items for the field test. The draft field test items were organized into a set of "preferred" and "alternate" item blocks based on input from the SMIRC (see section 2.5.5). At the third NRC meeting, the "preferred" item blocks were reviewed in plenary with all NRCs. The "alternate" item blocks were made available for review in separate review sessions, and NRCs provided feedback on these items in comment sheets. Both "preferred" and "alternate" items were subsequently revised in line with suggestions received from NRCs. In general, the items for the field test were well received, and NRCs were satisfied that the items consituted a very satisfactory field test item pool.

2.5.4 Developing the Problem-Solving and Inquiry Tasks

To address the importance placed in the frameworks on the assessment of problem-solving, reasoning and scientific inquiry, a set of tasks were developed to assess how well students can draw on and integrate a variety of processes and understandings in mathematics and science to conduct investigations and solve problems. At the first NRC meeting, it was decided that from an operational perspective, it was important that the tasks developed for TIMSS 2003 be less demanding to administer than the performance assessment conducted in TIMSS 1995. Specifically, the tasks needed to be self-contained, involve minimal equipment, and be integrated into the main test administration without any special accommodations or additional testing sessions. Thus, a major challenge for TIMSS 2003 was to develop a set of relevant problem-solving and inquiry tasks that would satisfy the requirements set forth by the Expert Panel and the national representatives.

The development of tasks was an evolutionary process, starting with a "brainstorming" session of international mathematics and science experts at the second Expert Panel meeting in May 2001. The expert panel developed a number of innovative prototype ideas for investigative or "real-world" tasks, many of which integrated ideas across mathematics and science. At the TIMSS item-writing workshop for participating countries, the approach to developing problem-solving and inquiry tasks was discussed, and some ideas were submitted by national centers as part of the international item development process. Much of the development for the problem-solving and inquiry tasks occurred at meetings of the Science and Mathematics Item Review Committee. At the first SMIRC meeting in September 2001, the initial set of ideas for tasks put forth by the Expert Panel and submitted from national centers were discussed for their appropriateness and feasibility, and considerable progress was made in drafting tasks using these ideas as a starting point. The ISC staff and a subset of SMIRC members further refined the initial drafts in the following few weeks, and these first drafts were presented to the full SMIRC at their second meeting in October 2001. The drafts were reviewed, and a subset was selected and substantially elaborated, at the second SMIRC meeting. Additional tasks were developed to ensure that the set of tasks covered a range of content areas in mathematics and science.

Following the second meeting of the SMIRC, ISC staff and the mathematics and science task force members continued to work on the tasks and prepare them for the presentation and review of field test materials at the second NRC meeting. A number of modifications recommended by the NRCs were incorporated following the meeting. It was suggested at the NRC meeting that the accessibility, reading level, and appropriateness of content and terminology for fourth- and eighth-grade students be further evaluated, particularly for the science tasks. To address this concern, the ISC recruited two experienced fourth-grade and eighth-grade science teachers in the Boston area to review the revised science tasks. The feedback from the teachers was very positive overall, indicating that most of the content was now grade appropriate and the tasks were interesting and engaging. After some revisions in layout, content, and language based on the results of the teacher review, a small-scale item pilot of the problem-solving and inquiry tasks and other constructed-response items was conducted in February 2002 in seven countries that tested in English. This pilot yielded a total of approximately 4500 student responses at each grade level, with 30 to 40 responses to each task. The results of this international pilot provided valuable information about how the tasks functioned internationally and were used primarily to refine the scoring guides and obtain student responses for use in preparing scoring training materials for the field test.

A total of 19 tasks (9 at fourth grade and 10 at eighth grade) were selected for the field test. Each task included a series of related test items, mostly constructed-response, that were linked by a common theme and involved an investigation or extended problem-solving situation. Some of the mathematics tasks involved manipulatives such as cardboard rulers or geometric tiles; no equipment or manipulatives were required for the science tasks. Although some of the initial ideas for science tasks involved the use of equipment, during further development stages it was decided that the type of equipment required was not feasible in the test administration setting. Each of the tasks in the field test was designed to take up to 12 minutes at the fourth grade and up to 15 minutes at the eighth grade, the length of one assessment block (see section 2.6.1 on booklet/block design). Exhibits 2.8 and 2.9 describe the problem-solving and inquiry tasks selected for the field test and the main content covered in each for the fourth grade and eighth grade, respectively.

Results from the international field test (section 2.5.5) were used to select the problem-solving and inquiry tasks that performed best internationally for the main survey. In addition, a cognitive laboratory investigation of the field-test version of the problem-solving and inquiry tasks was conducted by the United States National Center for Education Statistics. This involved working with a small group of students to probe their understanding of the demands of the tasks and to uncover any conceptual difficulties encountered in them. The results of the international field test as well as the experiences from the cognitive laboratory investigation were used to inform the selection process and to make revisions to improve the clarity of directions, layout, reading level, use of manipulatives, and scoring guides for the main survey. In general, the problem-solving and inquiry tasks selected for the main survey were shortened from the field test version. In some cases, an entire task or large portions of a task were selected. In other cases, individual items within tasks were selected and adapted to function as stand-alone items. As shown in Exhibit 2.10, a total of 13 problem-solving and inquiry tasks were selected for the main survey (6 at fourth grade and 7 at eighth grade).

Exhibit 2.8 **Problem-Solving and Inquiry Tasks Selected for the Field Test – Grade 4**

Name of Task	Description	Main Content
	Mathematics Tasks	
Geometry Tiles	Students are given three types of square tiles (black, white, and triangle tiles half black and half white) that can be placed together to form patterns. Students create two-dimensional shapes; compute fraction of pattern that is black; create patterns satisfying given conditions.	Geometry and Number
Number Tiles	Students are given number tiles marked from 0 to 9 that can be used to create addition, subtraction and multiplication problems. By choosing the place value of the numbers (units or tens), students combine their tiles to create problems that give a total closest to a given number, and to create the largest possible answer.	Number
Trading Cards	Three types of trading cards can be exchanged according to equivalency rules. Students compute how many cards they would get if they trade n cards of a certain type by another type of cards; explain how to maximize the number of cards they could get by trading; and infer conversion rules.	Number
Reversible Numbers	Presents examples of reversible numbers (e.g., 66, 121, 3003) and a general rule to make reversible numbers starting from two-digit numbers. Students provide examples of reversible numbers meeting certain conditions; create reversible numbers following one- or two-step rules; justify why reversible numbers cannot have three different digits; evaluate rules to create reversible numbers.	Number
Map It!	Students are shown maps drawn to scale indicating several locations. Using a cardboard ruler, students measure distance in centimetres between towns; estimate distance in kilometres; infer which towns are closer; compute time required to travel from one town to another; mark new plausible locations in the map so that they satisfy given conditions.	Measurement and Number
	Science Tasks	
Oceans and Tidepools	Presents textual and graphical information about the oceans and tidepools and a series of exploratory questions involving food chains, features of organisms, and ocean resources; students make predictions, provide explanations; select set-ups to investigate the effect of salt level on seaweed.	Life Science
Garden	Presents a practical situation involving a plan for a garden and a series of questions about plant growth and dispersal, light conditions, and importance/control of insects; students make predictions; provide explanations; interpret diagram; extract tabular information; relate position of sun and light conditions to complete table of plants in each area.	Life Science and Earth Science
Patterns on Earth	Presents historical information about measuring time using observed patterns (phases of the moon, daily cycle of the sun, appearance of shadows, periodic motion of pendulums); students evaluate graphical representations; complete diagrams; extend and relate patterns to time measurements; relate periodic motion of a pendulum to gravity.	Earth Science and Physical Science
Light and Color	Presents a practical situation involving an investigation of the effect of the light source on the color of materials; students describe and interpret results of the investigation; draw conclusions; make predictions and generalize results to new situations; compare with situations where color changes are due to changes in materials.	Physical Science

Exhibit 2.9 **Problem-Solving and Inquiry Tasks Selected for the Field Test – Grade 8**

Name of Task	Description	Main Content
	Mathematics Tasks	
Geometry Tiling	Provides four identical geometry tiles and several grids showing how tiles can be placed to form patterns. Students place tiles on a grid to make a pattern symmetrical about a given line; extend geometric patterns using symbols to represent the position of the tiles; and create whole new symmetrical patterns using symbols.	Geometry, Number, and Algebra
Class Trip	Students are given a map, bus timetables, trip rates per student, and a series of conditions that must be met in planning a class trip. Students estimate distances; compute costs for different trip options; evaluate if conditions can be met; decide upon which trip to make; and justify their choice.	Measurement, Number, and Data
Red and Black Tiles	Presents red and black tiles that can be combined to form square shapes with a given pattern but having different sizes. Students extend numeric and geometric patterns; identify number of tiles of each type required to form a shape of a given size; and infer the general algebraic expression to find out the number of tiles needed for any shape.	Algebra
Phone Plans	Presents two telephone payment plans involving fixed and variable costs. Students read and interpret data from a table to decide which plan would be cheapest under a range of conditions and justify their selection of a plan.	Data
Bird House	Students are given plans for making a wooden birdhouse. Working from the scale drawings in the plans, and using a ruler, students determine the actual size of the wood pieces required to build the birdhouse. They also infer the size of a missing piece, and draw it.	Measurement, Number, and Geometry
Number Triangles	Presents number triangles with some numbers missing, and an adding rule to combine the existing numbers to determine the missing numbers. Students determine how to create different combinations of odd and even numbers, and how to get positive and negative integers; they also identify ranges of values that satisfy given conditions.	Number
	Science Tasks	
Oceans	Presents textual and graphical information about the oceans and a series of exploratory questions involving food webs, adaptations of organisms, resources, and human exploration using sonar technology; students make predictions; provide explanations; interpret graphical information; describe procedures.	Life Science and Earth Science
Galapagos Islands	Presents textual and graphical information about the Galapagos Islands and a series of exploratory questions involving formation, arrival of organisms, impact of humans, adaptations and competition among species; students make predictions; interpret graphical data; draw conclusions; provide explanations.	Life Science and Earth Science
Metal Crown	Presents an investigation of a crown of unknown composition; students predict observable properties; describe a procedure to determine volume and density; evaluate results from repeated measures; draw conclusion by comparing measurements to density/cost data for various metals.	Physical Science
Light Filters	Presents practical situations involving color change due to the light source or to changes in materials; students interpret and explain results of an investigation of the effect of light sources/filters on color; apply knowledge of chemical change to a new situation (dye fade); design an investigation of the effect of light source on dye fade.	Physical Science

Exhibit 2.10 **Problem-Solving and Inquiry Tasks Selected for the Main Survey at Grade 4 and Grade 8**

Grade 4		Grade 8	
Name of Task	**Content Domains***	**Name of Task**	**Content Domains***
		Mathematics Tasks	
Geometry Tiles	Geometry (2) Number (4)	Geometry Tiling	Geometry (4) Algebra (1)
Number Tiles	Number (7)	Class Trip	Measurement (2) Number (2) Data (6)
Trading Cards	Number (6)	Red and Black Tiles	Algebra (8)
Marytown (portions of original Map It task)	Measurement (3)	Phone Plans	Data (6)
		Science Tasks	
Garden	Life Science (7) Earth Science (1)	Life in the Oceans (portions of original Oceans task)	Life Science (7)
Light and Color	Physical Science (7)	Galapagos Islands	Life Science (7)
		Metal Crown	Physics (4) Chemistry (3)

* The number of score points in each content domain is indicated in parentheses. The tasks range from three to ten score points.

2.5.5 Field Test

To evaluate the international performance of the new items developed for TIMSS 2003, a full-scale field test was conducted at both the fourth and eighth grades during the period April to June 2002. In total, 41 countries participated in the eighth-grade field test and 20 countries in the fourth grade. The field test in each country was administered to a random sample of a minimum of 25 schools, with two classrooms per school. To ensure that an adequate number of items were available for selection, substantially more items were field tested (1-1/2 to 2 times) than were needed in the assessment, particularly constructed-response items and items in content areas not already covered by trend items from 1995 and 1999.

Including the problem-solving and inquiry tasks, a total of 435 items were included in the fourth-grade field test, 229 in mathematics and 206 in science. At the eighth grade, a total of 386 items were included in the field test, 190 in mathematics and 196 in science. Since some constructed-response items contribute two score points, this corresponds to a total number of score points of 242 in mathematics and 248 in science at the fourth grade, and 211 in mathematics and 239 in science at the eighth grade.

2.5.6 Item Selection for the Main Survey

International item analysis of the results from the field test was used to inform the review and selection of items and tasks for the main survey. Data almanacs were produced containing basic item statistics for each country and internationally to evaluate the item difficulty, how well items discriminated between high-and low-performing students, the effectiveness of distracters in multiple-choice items, scoring reliability for constructed-response items, the frequency of occurrence of diagnostic codes used in the scoring guides, and whether there were any biases towards or against individual countries or in favor of boys or girls.

The TIMSS International Study Center conducted an initial review of the field-test results in early July 2002, using data from 36 countries at the eighth grade and 19 countries at the fourth grade that were available for analysis at that time. This review included NRC input from field test survey activities reports, feedback on items and scoring guides, and translation verification reports to identify any items with translation or cultural issues affecting international item performance. On the basis of this review, the mathematics and science coordinators identified the set of test items they felt would be most appropriate for use in the assessment, taking into account individual item statistics as well as alignment with the frameworks. Draft blocks of items for the assessment were then assembled for review by the Science and Mathematics Item Review Committee.

At its third meeting on July 15 - 18, 2002, the SMIRC reviewed the proposed item blocks, examining the field test item statistics to identify any anomalies. Items that did not work well were replaced with alternate items from the same content area. The problem-solving and inquiry tasks received particular attention and improvements were made where necessary. Revisions to items included improving graphics and item layout, clarifying stems, and revising distracters selected by very low percentages of students. In a few instances, item format was changed from multiple-choice to constructed-response or vice-versa. The final set of items selected was chosen to provide an appropriate balance in content coverage, level of difficulty, and item types.

Based on the recommendations of the SMIRC, the International Study Center prepared draft instruments for the assessment to be reviewed by the National Research Coordinators at their fifth meeting in August 2002. The draft instruments were well received and widely discussed by NRCs, who recommended a number of additional improvements that were incorporated into the final instruments distributed in September 2002. A total of 243 new items at the fourth grade and 230 items at the eighth grade were selected for the main survey. Including both trend and new items, the final tests include 313 items at the fourth grade and 383 items at the eighth grade. Exhibits 2.11 and 2.12 show the distribution of new and trend items in the main survey by subject and item format for fourth and eighth grades, respectively, and reflect the individual items and all item subparts included in multi-part items and problem-solving and inquiry tasks. Between 40 and 50 percent of the total score points are contributed by constructed-response items at both grades, which exceeds the minimum proportion of one-third specified in the frameworks.

Exhibit 2.11 Distribution of New and Trend Items in the TIMSS 2003 Main Survey by Subject and Item Format – Grade 4

	Number of Items				
Item Format	**New Items**	**Trend Items**	**Total (New + Trend)**	**Total Score Points**	**Percentage of Score Points**
		Mathematics Items			
Multiple Choice	55	37	92	92	54%
Constructed Response	69	0	69	77	46%
Total Mathematics Items	**124**	**37**	**161**	**169**	
		Science Items			
Multiple Choice	60	31	91	91	54%
Constructed Response	59	2	61	77	46%
Total Science Items	**119**	**33**	**152**	**168**	
		All Items			
Multiple Choice	115	68	183	183	54%
Constructed Response	128	2	130	154	46%
Total Items	**243**	**70**	**313**	**337**	

Exhibit 2.12 **Distribution of New and Trend Items in the TIMSS 2003 Main Survey by Subject and Item Format – Grade 8**

Item Format	Number of Items			Total Score Points	Percentage of Score Points
	New Items	Trend Items	Total (New + Trend)		
Mathematics Items					
Multiple Choice	69	59	128	128	60%
Constructed Response	46	20	66	87	40%
Total Mathematics Items	**115**	**79**	**194**	**215**	
Science Items					
Multiple Choice	56	53	109	109	52%
Constructed Response	59	21	80	102	48%
Total Science Items	**115**	**74**	**189**	**211**	
All Items					
Multiple Choice	125	112	237	237	56%
Constructed Response	105	41	146	189	44%
Total Items	**230**	**153**	**383**	**426**	

2.5.7 Scoring of Constructed-Response Items

In the TIMSS 2003 assessment, constructed-response items made up more than 40 percent of the total assessment time, and a large number of constructed-response items were developed and field tested. Scoring guide development for the constructed-response items was a considerable effort and an integral part of the test development process for TIMSS 2003. This section describes the TIMSS general scoring method, the scoring guide development process, and the scoring training materials and procedures.

2.5.7.1 The TIMSS General Scoring Method

TIMSS 2003 used the same approach to scoring as the previous TIMSS assessments. As in TIMSS 1995 and 1999, both short-answer items and extended-response items were included in the assessment. Short-answer items typically are worth one score point and require a numerical response in mathematics or a brief descriptive response in science. Extended-response items are worth a maximum of two score points and require students to show their work or provide explanations using words and/or diagrams to demonstrate their conceptual and procedural knowledge. The generalized scoring guides for mathematics and science items developed for TIMSS 1999 (Exhibit 2.13) also were applied in TIMSS 2003.

Exhibit 2.13 **TIMSS Generalized Scoring Guide for Mathematics and Science Items**

Mathematics	Science
Extended-Response Items	
2 Points A two-point response is complete and correct. The response demonstrates a thorough understanding of the mathematical concepts and/or procedures embodied in the task. • Indicates that the student has completed the task, showing mathematically sound procedures • Contains clear, complete explanations and/or adequate work when required	**2 Points** A two-point response is complete and correct. The response demonstrates a thorough understanding of the science concepts and/or procedures embodied in the task. • Indicates that the student has completed all aspects of the task, showing the correct application of scientific concepts and/or procedures • Contains clear, complete explanations and/or adequate work when required
1 Point A one-point response is only partially correct. The response demonstrates only a partial understanding of the mathematical concepts and/or procedures embodied in the task. • Addresses some elements of the task correctly but may be incomplete or contain some procedural or conceptual flaws • May contain a correct solution with incorrect, unrelated, or no work and/or explanation when required • May contain an incorrect solution but applies a mathematically appropriate process	**1 Point** A one-point response is only partially correct. The response demonstrates only a partial understanding of the science concepts and/or procedures embodied in the task. • Addresses some elements of the task correctly but may be incomplete or contain some procedural or conceptual flaws • May contain a correct answer but with an incomplete explanation when required • May contain an incorrect answer but with an explanation indicating a correct understanding of some of the scientific concepts
0 Points A zero-point response is completely incorrect, irrelevant, or incoherent.	**0 Points** A zero-point response is seriously inaccurate or inadequate, irrelevant, or incoherent.
Short-Answer Items	
1 Point A one-point response is correct. The response indicates that the student has completed the task correctly.	**1 Point** A one-point response is correct. The response indicates that the student has completed the task correctly.
0 Points A zero-point response is completely incorrect, irrelevant, or incoherent.	**0 Points** A zero-point response is completely incorrect, irrelevant, or incoherent.

Each constructed-response item has its own scoring guide that utilizes a two-digit scoring scheme to provide diagnostic information. The first digit designates the correctness level of the response: 2 for a two-point response, 1 for a 1-point response, and 7 for an incorrect response. The second digit, combined with the first, represents a diagnostic code used to identify specific types of approaches, strategies, or common errors and misconceptions. A second digit of 0-5 may be used for pre-defined international codes at each correctness level, while a second digit of 9 corresponds to "other" types of responses that fall within the appropriate correctness level but do not fit any of the pre-defined international codes. A special code (99) is given for completely blank responses. In general, only a few diagnostic codes are used to track high-frequency correct or partial approaches or common misconceptions and errors, and a particular effort was made in TIMSS 2003 to minimize the number of diagnostic codes used. In addition to the international codes, second digit codes of 7 and 8 may be used by national centers to monitor specific responses not already captured by the internationally-defined codes. The general TIMSS two-digit scoring scheme is summarized in Exhibit 2.14.

Exhibit 2.14 **TIMSS Two-Digit Scoring Scheme for Constructed-Response Items**

Two- Point Items			One-Point Items		
Correctness Level	**International Code(s)**		**Correctness Level**	**International Code(s)**	
Correct Responses	20 – 25:	category/method #1 - #5	Correct Responses	10 – 15:	category/method #1- #5
	29:	other correct method		19:	other correct method
Partial Responses	10 – 15:	category/method #1- #5	Incorrect Responses	70 – 75:	misconception/error #1- #5
	19:	other partial method		79:	other error
Incorrect Responses	70 – 75:	misconception/error #1 - #5	Blank	99	
	79:	other error			
Blank	99				

2.5.7.2 Developing the TIMSS 2003 Scoring Guides

Items and scoring guides were developed in parallel, with draft scoring guides provided by item writers along with their item submissions. Scoring guides were further developed during item review and revision by the mathematics and science task forces and at the first two meetings of the Science and Mathematics Item Review Committee. Draft field-test versions of the scoring guides were reviewed by National Research Coordinators at their third NRC meeting. In February 2002, prior to the field test, a small-scale pilot of fourth- and eighth-grade constructed-response items was conducted in seven countries that tested in English. This pilot included all of the problem-solving and inquiry tasks as well as other items with more challenging scoring guides. Results from the pilot were used to finalize scoring guides for the field test by identifying common responses and clarifying the threshold for correct versus partial or incorrect responses. Selected student responses from the pilot were included as examples in the scoring guides and materials for scoring training for the field test.

In general, the scoring reliability from the field test was quite high, with an average percent agreement correctness of more than 90 percent for nearly all items. However, scoring relibility data did suggest some scoring guide revisions. NRC feedback on their scoring experiences during the field test also was used to make improvements in the scoring guides. In addition, sets of student booklets from the field test were collected from all of the English-test countries as sources of example student responses to clarify codes and prepare scoring training materials for the assessment.

During the review of the main survey test instruments at the fifth NRC meeting in August 2002, the changes recommended at the SMIRC meeting were discussed and NRCs made some additional suggestions for revisions to the scoring guides. Because so many changes were made to the problem-solving and inquiry tasks after the field test, these were included in a second small-scale item trial conducted in September 2002 in five countries that test in English. Student responses from this trial provided examples for the final scoring guides and for scoring training materials. The scoring guides and training materials were used during the first international scoring training session in November 2002 for southern hemisphere countries. A few additional revisions and clarifications were suggested by the national representatives at this training session. These were incorporated into the guides prior to their general distribution in December 2002.

Scoring guides for the trend constructed-response items (35 items from the eighth-grade 1999 assessment and 2 items from the fourth-grade 1995 assessment) were essentially unchanged from the versions used in the previous assessments, except for some modifications made to be consistent with the TIMSS 2003 format.[4]

2.5.7.3 Scoring Training Materials and Procedures

As in previous assessments, the International Study Center used a "train-the-trainers" approach to provide training on the international procedures for scoring the TIMSS 2003 constructed-response items. National Research Coordinators and/or other personnel responsible for training scorers in each country participated in training sessions for the field test and the main survey. In each of these sessions, the general TIMSS scoring approach was reviewed, and participants were then trained on a subset of constructed-response items. The subset of items was selected to reflect a range of scoring guide types and situations encountered across the TIMSS mathematics and science items and included some of the most complicated scoring guides.

Training was organized into four sessions by subject and grade (mathematics fourth and eighth grades and science fourth and eighth grades) conducted by the mathematics and science coordinators and task force members. Participants received the international version of the scoring guides and a binder for each subject/grade combination containing a set of prescored

4 Scoring guides for a few eighth-grade science items from 1999 were simplified to reduce the number of diagnostic codes. In all cases, the overall scoring strategy was retained to ensure score-level reliability from 1999 to 2003.

example student responses illustrating the diagnostic codes and the rationale used to score the responses and a set of 10-20 unscored practice responses for each item. The student responses were selected from the international small-scale item pilot and field-test booklets.

The purpose of the international scoring training was to present a model for use in each country and an opportunity to practice and resolve scoring issues with the most difficult items. The training teams discussed the need for NRCs to prepare comparable materials for training in their own country for all constructed-reponse items and a larger number of practice responses for the more challenging scoring guides during the national training sessions. The following general procedures were followed in the scoring training for each item:

- Participants read the item and its scoring guide.
- Trainers discussed the rationale and methodology of the scoring guide.
- Trainers presented and discussed the set of prescored example student responses.
- Participants scored the set of practice student responses.
- Trainers led a group discussion of the scores given to the practice responses to reach a common understanding of the interpretation and application of the scoring guide.

Scoring training for the field test was conducted at the fourth NRC meeting in March 2002. Two full days of scoring training were devoted to the science items, with one day for each grade. For mathematics, training for both grades was done over a total of one and one-half days.

Scoring training for the assessment was conducted in the same fashion as for the field test, with separate sessions devoted for each subject/grade combination. For the assessment scoring training, 40 total items were included for eighth grade – 20 mathematics items and 20 science items. This set of items represents nearly 30 percent of the constructed-response items in the eighth-grade assessment. For fourth grade, 14 items were included for mathematics, and 16 items were included for science. This represents more than 25 percent of the constructed-response items in the fourth-grade assessment. For each grade, at least one item from each of the problem-solving and inquiry tasks was selected for training.

Two main scoring training sessions were conducted for the 2003 assessment, one for countries on a southern hemisphere schedule and one for countries on a northern hemisphere schedule.[5] The first was held in November 2002 in Wellington, New Zealand, for southern-hemisphere countries. The second, held in March 2003 in conjunction with the sixth NRC meeting

5 An extra scoring training session was organized in May 2003 for northern hemisphere countries that were unable to attend the main training session.

in Bucharest, Romania, was for the remaining countries. At each session, a full day of training was devoted to each subject for eighth grade and a little less for fourth grade (about a half day for mathematics and three-quarters for science). After the completion of scoring training, code sheets for the example and practice papers were distributed to NRCs for use in organizing scoring training materials in their own countries.

2.6 Assessment Booklet Design

In order to cover the frameworks, the pool of items and tasks included in the TIMSS asssessment is extensive and would require much more testing time than could be alloted for individual students (about seven hours at grade 8 and five and one-half hours at grade 4). Therefore, as in the 1995 and 1999 assessments, TIMSS 2003 uses a matrix-sampling technique that involves dividing the entire assessment pool into a set of unique item blocks, distributing these blocks across a set of booklets, and rotating the booklets among the students. Each student takes one booklet containing both mathematics and science items.[6]

2.6.1 Block and Booklet Design

The TIMSS design for 2003 divides the 313 items at fourth grade and 383 items at eighth grade into 28 item blocks at each grade, 14 mathematics blocks labeled M01 through M14, and 14 science blocks labeled S01 through S14. Each block contains either mathematics items only or science items only. This general block design, shown in Exhibit 2.15, is the same for both grades, although for the assessment time is 12 minutes for fourth-grade blocks and 15 minutes for eighth-grade blocks. At the eighth grade, six blocks in each subject (blocks 01 – 06) contain secure items from 1995 and 1999 to measure trends and eight blocks (07 – 14) contain new items developed for TIMSS 2003. Since fourth grade was not included in the 1999 assessment, trend items from 1995 only were available, and these were placed in the first three blocks. The remaining 11 blocks contain items new in 2003.

In the TIMSS 2003 design, the 28 blocks of items are distributed across 12 student booklets, as shown in Exhibit 2.16. Each booklet consists of six blocks of items. To enable linking between booklets, each block appears in two, three, or four different booklets. The assessment time for individual students is 72 minutes at fourth grade and 90 minutes at eighth grade, which is comparable to that in the 1995 and 1999 assessments.

6 See Mullis et al. (2003) for more information on the assessment booklet design.

Exhibit 2.15 **General Design of the TIMSS 2003 Matrix-Sampling Blocks**

Source of Items	Mathematics Blocks	Science Blocks
Trend Items (TIMSS 1995 or 1999)	M01	S01
Trend Items (TIMSS 1995 or 1999)	M02	S02
Trend Items (TIMSS 1995 or 1999)	M03	S03
Trend Items (TIMSS 1999)	M04	S04
Trend Items (TIMSS 1999)	M05	S05
Trend Items (TIMSS 1999)	M06	S06
New Replacement Items (TIMSS 2003)	M07	S07
New Replacement Items (TIMSS 2003)	M08	S08
New Replacement Items (TIMSS 2003)	M09	S09
New Replacement Items (TIMSS 2003)	M10	S10
New Replacement Items (TIMSS 2003)	M11	S11
New Replacement Items (TIMSS 2003)	M12	S12
New Replacement Items (TIMSS 2003)	M13	S13
New Replacement Items (TIMSS 2003)	M14	S14

The booklets are organized into two three-block sessions (Parts I and II), with a break in between each part. Since the use of calculators was introduced for the first time in TIMSS 2003 at the eighth grade, this had an impact on the booklet design. To ensure that calculators could be used for the new items but not for the trend items from 1995 and 1999, the trend items (blocks 01 – 06) were placed in Part I of the test booklets to be completed without calculators before the break. After the break, calculators were allowed for the new items (blocks 07 – 12) at eighth grade but not fourth grade. To provide a more balanced design, however, two mathematics trend blocks (M05 and M06) and two science trend blocks (S05 and S06) also were placed in Part II of one booklet each.

Exhibit 2.16 **Booklet Design for TIMSS 2003 – Grade 4 and Grade 8**

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

2.6.2 Assembling Item Blocks

The assessment blocks were assembled to create a balance across blocks and booklets with respect to content domain, cognitive domain, and item format. Although a balance was achieved at the overall assessment level, the distribution of item types varies across blocks. The trend blocks from 1995 (blocks 01 – 03) contain mostly multiple-choice items, while the blocks containing the problem-solving and inquiry tasks have a higher proportion of constructed-response items. Each block contains an average of 12 score points at fourth grade and 15 score points at eighth grade, and the percentage of score points from constructed-response items in each block ranges from 0 to about 80 percent. On average, there are 6-7 multiple-choice items, 4-5 short-answer items, and 0-1 extended-response items per block at the fourth grade. At the eighth grade, there are 8-9 multiple-choice items, 3-4 short-answer items, and 1-2 extended-response items per block, on average. Depending on the exact number of multiple-choice, short-answer, and extended-response items in each block, the total number of items in a block ranges from 10 to 13 at fourth grade and from 11 to 16 at eighth grade.

2.6.3 Incorporating Trend Items

In TIMSS 1995 and 1999, items were organized into 26 item clusters (labeled A through Z). Clusters A-R contained sets of both mathematics and science items, clusters S-V only mathematics items, and clusters W-Z only science items. After the 1995 assessment, clusters A-H (containing nearly all multiple-choice items) were held secure for future assessments; clusters I-Z were released and replaced with new items in the 1999 assessment. Since the fourth grade was not included in the 1999 assessment, only clusters A-H from 1995 are available as trend items for the 2003 assessment, and these clusters contain nearly all multiple-choice items.

At the eighth grade, clusters I-Z contained items developed for the 1999 assessment. At the end of TIMSS 1999, the “even” clusters (B, D, F, etc.) were released and the “odd” clusters (A, C, E, etc.) were held secure as trend items for the 2003 assessment. Therefore, the following clusters of trend items at the eighth grade are available for the 2003 assessment:

- 1995 items: A, C, E, G (mathematics and science)
- 1999 items: I, K, M, O, Q (mathematics and scence); S, U (mathematics); W, Y (science)

Because of the new booklet and block design specified in the TIMSS 2003 frameworks, the trend item clusters from 1995 and 1999 were reorganized for the TIMSS 2003 assessment. In accordance with the TIMSS 2003 test design, mathematics and science items from 1995 were assigned to blocks M01-M03 or S01-S03. Most items from 1999 were assigned to blocks M04-M06 or S04-S06, although some were assigned to blocks M01-M03 or S01-S03 where there were insufficient 1995 items to fill these blocks. In addition, some new items were added to fill the trend blocks; in particular, blocks M04-M06 and S04-S06 contain all new items at the fourth grade. The assignment of 1995 and 1999 trend item clusters to the TIMSS 2003 item blocks and the resulting distribution of score points across assessment years is summarized for the fourth and eighth grades in Exhibits 2.17 and 2.18, respectively.

2.6.4 Alignment with the Mathematics and Science Frameworks

The test development process for TIMSS 2003 successfully produced fourth- and eighth-grade assessments aligned with the *TIMSS Assessment Frameworks and Specifications 2003*. Details about the coverage of the frameworks are given separately for the fourth- and eighth-grades in the following sections.

Exhibit 2.17 **TIMSS 2003 Mathematics and Science Blocks – Grade 4: Number of Items from 1995 Trend Clusters and Score Points by Assessment Year**

Block	Number of Items from Trend Clusters*	Score Points by Assessment Year		
		1995	2003	Total
	Mathematics Blocks			
M01	C(4), E(4), G(4)	12	0	12
M02	A(3), D(5), F(5)	13	0	13
M03	A(2), B(5), H(5)	12	0	12
M04 – M14	–	0	132	132
Mathematics Total		**37**	**132**	**169**
	Science Blocks			
S01	A(4), B(4), F(3)	11	0	11
S02	D(2), G(5), H(4)	11	0	11
S03	C(5), D(2), E(4)	11	0	11
S04 – S14	–	0	135	135
Science Total		**33**	**135**	**168**
Overall Total Score Points		**70**	**267**	**337**

* The number of items from each trend cluster is indicated in parentheses. Items in clusters A-H were developed for the 1995 assessment; grade 4 was not included in the 1999 assessment. Blocks M04-M14 and S04-S14 contain only new items developed for TIMSS 2003.

2.6.4.1 Fourth-Grade Assessment

Exhibit 2.19 shows the distribution of score points across content and cognitive domains in the fourth-grade mathematics assessment. The percentage of score points across both content and cognitive categories is very close to the target percentages specified in the frameworks (Exhibit 2.2). Exhibit 2.20 shows the score-point distribution for the fourth-grade science assessment, as well as the score points in the scientific inquiry assessment strand (see Exhibit 2.4 for the science framework target percentages). For both mathematics and science, items reflecting the full range of cognitive domains are included in each content domain. About 10 percent of the score points in science, covering a wide range of science content, also contribute to the scientific inquiry strand.

Exhibit 2.18 **TIMSS 2003 Mathematics and Science Blocks – Grade 8: Number of Items from 1995/1999 Trend Clusters and Score Points by Assessment Year**

Block	Number of Items from Trend Clusters*	Score Points by Assessment Year			
		1995	1999	2003	Total
		Mathematics Blocks			
M01	A(6), G(6)	12	0	3	15
M02	C(5), Q(10)	5	10	0	15
M03	E(6), O(9)	6	9	0	15
M04	I(9), S(7)	0	17	0	17
M05	K(9), U(4)	0	16	0	16
M06	M (8)	0	8	7	15
M07 – M14	–	0	0	122	122
Mathematics Total		**23**	**60**	**132**	**215**
		Science Blocks			
S01	E(6), K(10)	6	10	0	16
S02	A(6), C(6)	12	0	3	15
S03	G(6), O(8)	6	8	0	14
S04	M(6), W(4)	0	11	4	15
S05	I(11), Y(3)	0	15	0	15
S06	Q (8)	0	8	7	15
S07 – S14	–	0	0	121	121
Science Total		**24**	**52**	**135**	**211**
Overall Total Score Points		**47**	**112**	**267**	**426**

* The number of items from each trend cluster is indicated in parentheses. Items in clusters A-H were developed for the 1995 assessment; items in clusters I-Z were developed for the 1999 assessment. Blocks M07-M14 and S07-S14 contain only new items developed for TIMSS 2003.

Exhibit 2.19 **Distribution of Score Points in the TIMSS 2003 Mathematics Assessment by Content and Cognitive Domains – Grade 4**

Content Domain	Cognitive Domain				Total Score Points	Percentage of Score Points
	Knowing Facts and Procedures	Using Concepts	Solving Routine Problems	Reasoning		
Number	15	17	27	9	68	40%
Patterns and Relationships	3	5	9	8	25	15%
Measurement	9	3	12	9	33	20%
Geometry	12	8	4	1	25	15%
Data	0	6	9	3	18	11%
Total Score Points	**39**	**39**	**61**	**30**	**169**	
Percentage of Score Points	**23%**	**23%**	**36%**	**18%**		

Exhibit 2.20 **Distribution of Score Points in the TIMSS 2003 Science Assessment by Content and Cognitive Domains, and Scientific Inquiry Strand – Grade 4**

Content Domain	Cognitive Domain			Total Score Points	Percentage of Score Points	Scientific Inquiry Score Points
	Factual Knowledge	Conceptual Understanding	Reasoning and Analysis			
Life Science	28	28	16	72	43%	4
Physical Science	16	26	17	59	35%	12
Earth Science	15	16	6	37	22%	1
Total Score Points	**59**	**70**	**39**	**168**		**17**
Percentage of Score Points	**35%**	**42%**	**23%**			

In accordance with the frameworks, a range of item types is reflected in the TIMSS 2003 assessment, including multiple-choice, short-answer, and extended-response items. Exhibit 2.21 shows the breakdown of the fourth-grade mathematics and science items by item type and cognitive domain, indicating that each content domain covers a range of item types.

Exhibit 2.21 **Number of Mathematics and Science Items in TIMSS 2003 by Item Type and Content Domain – Grade 4**

Content Domain	Item Type			Total Number of Items
	Multiple Choice	Short Answer	Extended Response	
Mathematics Items				
Number	30	31	2	63
Patterns and Relationships	16	7	1	24
Measurement	23	10	0	33
Geometry	12	11	1	24
Data	11	5	1	17
Total Mathematics Items	**92**	**64**	**5**	**161**
Science Items				
Life Science	41	23	1	65
Physical Science	29	20	4	53
Earth Science	21	13	0	34
Total Science Items	**91**	**56**	**5**	**152**
Total Overall Items	**183**	**120**	**10**	**313**

TIMSS reports trends in student achievement in mathematics in the major content domains of each subject. To facilitate linking to previous assessments, TIMSS 2003 includes items from 1995 in the fourth grade and from 1995 and 1999 in the eighth grade in each content domain. Exhibit 2.22 shows, for the fourth-grade assessment, the number of score points in mathematics and science contributed by items used previously in 1995 and by those used for the first time in 2003. In mathematics, the number of score points in the five content domains ranges from a maximum of 19 (Number) to a minimum of 2 (Patterns and Relationships). In science, there are between 9 and 12 score points from the 1995 assessment in the content domains. Because there are relatively few items and score points from the 1995 assessment in most content domains, TIMSS 2003 developed achievement scales linking 1995 and 2003 for mathematics and science overall, but not for individual content domains. However, the TIMSS 2003 design makes provision for sufficient trend items to develop achievement scales linking the content domains from 2003 onwards, i.e., to 2007, 2011, and so on.

Exhibit 2.22 **Number of Score Points in TIMSS 2003 from Each Assessment Year by Mathematics and Science Content Domain – Grade 4**

	Assessment Year			
Content Domain	**From 1995**	**From 1999**	**New in 2003**	**Total 2003**
		Mathematics		
Number	19	N/A	49	68
Patterns and Relationships	2	N/A	23	25
Measurement	8	N/A	25	33
Geometry	4	N/A	21	25
Data	4	N/A	14	18
Total in Mathematics	**37**	**N/A**	**132**	**169**
		Science		
Life Science	12	N/A	60	72
Physical Science	9	N/A	50	59
Earth Science	12	N/A	25	37
Total in Science	**33**	**N/A**	**135**	**168**
Total Overall	**70**	**N/A**	**267**	**337**

N/A: Not Applicable – TIMSS was not administered at fourth grade in 1999.

The block and booklet design for TIMSS 2003 ensures that the student booklets contain an appropriate balance of mathematics and science content. Exhibit 2.23 shows the number of mathematics and science score points available in each fourth-grade booklet. The number of score points per booklets ranges from 71 to 80, with an average of 75. In accordance with the frameworks, in booklets 1-6 about two-thirds of the score points come from mathematics items and one-third from science. Conversely, in booklets 7-12 about two-thirds of the score points come from science items and one-third from mathematics. All student booklets contain items from each of the mathematics and science content domains.

Exhibit 2.23 **Maximum Number of Score Points in TIMSS 2003 in Each Booklet by Mathematics and Science Content Domain – Grade 4**

	Booklet											
Content Domain	1	2	3	4	5	6	7	8	9	10	11	12
Mathematics												
Number	19	18	23	17	21	20	5	7	10	11	11	8
Patterns and Relationships	6	6	6	8	6	5	7	3	4	2	3	6
Measurement	12	13	7	9	11	12	6	8	2	3	5	6
Geometry	4	7	7	9	4	7	2	3	5	6	4	3
Data	8	5	6	5	4	3	4	3	2	2	4	2
Total in Mathematics	**49**	**49**	**49**	**48**	**46**	**47**	**24**	**24**	**23**	**24**	**27**	**25**
Science												
Life Science	13	10	13	13	9	11	18	22	19	21	29	24
Physical Science	6	9	10	8	7	10	12	14	20	25	14	17
Earth Science	8	10	4	4	9	7	18	16	11	7	10	9
Total in Science	**27**	**29**	**27**	**25**	**25**	**28**	**48**	**52**	**50**	**53**	**53**	**50**
Total Overall	**76**	**78**	**76**	**73**	**71**	**75**	**72**	**76**	**73**	**77**	**80**	**75**

2.6.4.2 Eighth-Grade Assessment

Exhibit 2.24 shows the distribution of score points across content and cognitive domains in the TIMSS 2003 eighth-grade mathematics assessment. The percentage of score points is close to the target percentages (Exhibit 2.2) for nearly all content and cognitive categories, although the assessment has a somewhat higher percentage of items in *knowing facts and procedures* and a lower percentage in *solving routine problems*. Exhibit 2.25 shows the distribution of score points across content and cognitive domains in the eighth-grade science assessment, as well as the number of score points in each content domain that also pertain to the scientific inquiry assessment strand. The percentages of score points in the content and cognitive domains of the science assessment also are close to their targets (see Exhibit 2.4). As with the fourth-grade assessment, items reflecting a range of cognitive domains are included in each of the mathematics and science content domains at the eighth grade. About 14 percent of the score points in science, covering a wide range of science content, also contribute to the scientific inquiry strand.

Exhibit 2.24 **Distribution of Score Points in the TIMSS 2003 Mathematics Assessment by Content and Cognitive Domains – Grade 8**

Content Domain	Cognitive Domain				Total Score Points	Percentage of Score Points
	Knowing Facts and Procedures	Using Concepts	Solving Routine Problems	Reasoning		
Number	15	11	27	7	60	28%
Algebra	13	12	10	18	53	25%
Measurement	9	2	15	8	34	16%
Geometry	7	8	10	9	34	16%
Data	1	6	14	13	34	16%
Total Score Points	**45**	**39**	**76**	**55**	**215**	
Percentage of Score Points	**21%**	**18%**	**35%**	**26%**		

Exhibit 2.25 **Distribution of Score Points in the TIMSS 2003 Science Assessment by Content and Cognitive Domains and Scientific Inquiry Strand – Grade 8**

Content Domain	Cognitive Domain			Total Score Points	Percentage of Score Points	Scientific Inquiry Score Points
	Factual Knowledge	Conceptual Understanding	Reasoning and Analysis			
Life Science	24	24	17	65	31%	8
Chemistry	7	16	11	34	16%	6
Physics	7	23	19	49	23%	9
Earth Science	12	13	8	33	16%	1
Environmental Science	9	4	17	30	14%	6
Total Score Points	**59**	**80**	**72**	**211**		**30**
Percentage of Score Points	**28%**	**38%**	**34%**			

Exhibit 2.26 shows the number of multiple-choice, short-answer, and extended-response items in each content domain for the eighth-grade assessment. As in the fourth grade, each of the content domains at eighth grade includes a range of item types.

Exhibit 2.26 **Number of Mathematics and Science Items in TIMSS 2003 by Item Type and Content Domain – Grade 8**

Content Domain	Item Type: Multiple Choice	Item Type: Short Answer	Item Type: Extended Response	Total Number of Items
Mathematics Items				
Number	43	11	3	57
Algebra	29	13	5	47
Measurement	19	9	3	31
Geometry	22	6	3	31
Data	15	8	5	28
Total Mathematics	**128**	**47**	**19**	**194**
Science Items				
Life Science	29	17	8	54
Chemistry	20	10	1	31
Physics	28	15	3	46
Earth Science	22	9	0	31
Environmental Science	10	8	9	27
Total Science	**109**	**59**	**21**	**189**
Total Items	**237**	**106**	**40**	**383**

To study trends in eighth-grade student mathematics and science achievement, TIMSS 2003 included items from the TIMSS 1995, 1999, and 2003 assessments. Exhibit 2.27 shows the number of score points in mathematics and science contributed by items used previously in 1995 and in 1999 as well as by those used for the first time in 2003. Among items from 1995, the number of score points in each content domain ranges from 3 to 6, and among 1999 items, from 6 to 20. TIMSS 2003 developed achievement scales linking 1995, 1999, and 2003 for mathematics and science overall, but because there are relatively few items and score points from the 1995 and 1999 assessments in content domains, TIMSS did not develop scales for measuring trends in individual content domains. However, the TIMSS 2003 design makes provision for sufficient trend items to develop achievement scales linking the content domains from 2003 onwards, i.e., to 2007, 2011, and so on. TIMSS used average percents correct to show changes in performance in the content domains from 1999 to 2003.

Exhibit 2.27 **Number of Score Points in TIMSS 2003 from Each Assessment Year by Mathematics and Science Content Domain – Grade 8**

Content Domain	Assessment Year			Total 2003
	From 1995	From 1999	New in 2003	
Mathematics				
Number	6	20	34	60
Algebra	6	11	36	53
Measurement	4	14	16	34
Geometry	4	8	22	34
Data	3	7	24	34
Total in Mathematics	**23**	**60**	**132**	**215**
Science				
Life Science	6	12	47	65
Chemistry	4	11	19	34
Physics	5	17	27	49
Earth Science	6	6	21	33
Environmental Science	3	6	21	30
Total in Science	**24**	**52**	**135**	**211**
Total Overall	**47**	**112**	**267**	**426**

Exhibit 2.28 shows the maximum number of score points in mathematics, science, and overall and the distribution of score points across the mathematics and science content domains for each booklet in the eighth-grade assessment. The total score points in each booklet ranges from 90 to 97, with an average of 94. As in the fourth grade, about two-thirds of score points are from mathematics items in booklets 1-6, and about two-thirds of score points are from science items in booklets 7-12. Each booklet covers the full range of mathematics and science content domains.

Exhibit 2.28 **Maximum Number of Score Points in TIMSS 2003 in Each Booklet by Mathematics and Science Content Domain – Grade 8**

	Booklet											
Content Domain	1	2	3	4	5	6	7	8	9	10	11	12
					Mathematics							
Number	19	22	19	14	25	17	11	10	8	7	8	7
Algebra	12	12	15	22	7	15	4	6	9	6	7	9
Measurement	13	7	11	13	11	5	4	8	8	4	5	4
Geometry	8	11	11	7	14	10	5	4	5	4	5	5
Data	11	8	6	8	4	12	8	3	2	8	5	6
Total Mathematics	**63**	**60**	**62**	**64**	**61**	**59**	**32**	**31**	**32**	**29**	**30**	**31**
					Science							
Life Science	10	12	7	7	9	10	20	19	18	23	21	23
Chemistry	6	4	5	6	4	5	10	10	11	8	10	10
Physics	6	8	9	6	7	9	17	11	13	19	14	13
Earth Science	4	5	3	5	8	5	9	12	11	7	7	10
Environmental Science	6	4	9	4	4	3	7	10	11	11	8	8
Total Science	**32**	**33**	**33**	**28**	**32**	**32**	**63**	**62**	**64**	**68**	**60**	**64**
Total Overall	**95**	**93**	**95**	**92**	**93**	**91**	**95**	**93**	**96**	**97**	**90**	**95**

2.6.5 Item Release Policy

TIMSS 2003 is the third assessment in a series of regular four-year studies, providing trend data from 1995 and 1999. As in previous assessments, the design for TIMSS 2003 and beyond (2007, 2011, etc.) provides for retaining some of the items for the measurement of trend and releasing some items into the public domain. In TIMSS 2003, half of the 14 assessment blocks in each subject will be released after the assessment results for 2003 are published. The released blocks will include all three mathematics and three science blocks containing trend items from 1995 (blocks M01 – M03, S01 – S03), one mathematics and one science block of trend items from 1999 (blocks M04 and S04)[7] and three blocks of new mathematics and science items and tasks developed for 2003 (blocks M09, M10, and M13; S09, S10, and S13). As item blocks are released, new items will be developed to take their place, and the release policy for future assessments will ensure that item blocks are cycled out after three assessments.

In the assignment of items to blocks in TIMSS 2003, particular attention was paid to balancing the blocks with respect to content domain to

7 At fourth grade, these blocks contain new 2003 items.

ensure that adequate numbers of items are held secure in each area for the purposes of measuring trend in future studies. In addition, the placement of the problem-solving and inquiry tasks results in about half of the tasks being retained and half being released after 2003. The released item set provides valuable information for interpreting the international and national reports and for use in secondary analyses. Therefore, it is also important that the released set be representative of the overall test to provide as much information as possible about the nature and scope of the test. Exhibits 2.29 and 2.30 show the number of secure and released items from the TIMSS 2003 assessment for fourth and eighth grades broken down by content domain. Approximately half of the items overall and in each content domain are released and half are kept secure. At the fourth grade, however, more than half of the items in the Number content domain and less than half of the items in Patterns and Relationships are released.

Exhibit 2.29 **Number of Items in Each Mathematics and Science Content Domain by Release Status in TIMSS 2003 – Grade 4**

Content Domain	Secure	Released	Total
Mathematics			
Number	24	39	63
Patterns and Relationships	17	7	24
Measurement	19	14	33
Geometry	12	12	24
Data	9	8	17
Total Mathematics	**81**	**80**	**161**
Science			
Life Science	32	33	65
Physical Science	29	24	53
Earth Science	15	19	34
Total Science	**76**	**76**	**152**
Total Overall	**157**	**156**	**313**

Exhibit 2.30 **Number of Items in Each Mathematics and Science Content Domain by Release Status in TIMSS 2003 – Grade 8**

Content Domain	Secure	Released	Total
Mathematics			
Number	26	31	57
Algebra	23	24	47
Measurement	14	17	31
Geometry	15	16	31
Data	17	11	28
Total Mathematics	**95**	**99**	**194**
Science			
Life Science	27	27	54
Chemistry	14	17	31
Physics	23	23	46
Earth Science	15	16	31
Environmental Science	15	12	27
Total Science	**94**	**95**	**189**
Total Overall	**189**	**194**	**383**

References

Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzalez, E.J., Chrostowski, S.J., & O'Connor, K.M. (2001), *TIMSS Assessment Frameworks and Specifications 2003*, Chestnut Hill, MA: TIMSS International Study Center, Boston College.

Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzalez, E.J., Chrostowski, S.J., & O'Connor, K.M. (2003), *TIMSS Assessment Frameworks and Specifications 2003 (2nd ed.)*, Chestnut Hill, MA: TIMSS International Study Center, Boston College.

Robitaille, D.F., McKnight, C.C., Schmidt, W.H., Britton, E., Raizen, S., & Nicol, C. (1993), *Curriculum Frameworks for Mathematics and Science: TIMSS Monograph No. 1*, Vancouver, BC: Pacific Educational Press.