Martin, M.O. (1996) "Third International Mathematics and Science Study: An Overview" in M.O. Martin and D.L. Kelly (eds.), *Third International Mathematics and Science Study (TIMSS) Technical Report, Volume I: Design and Development*. Chestnut Hill, MA: Boston College.

1. Third International Mathematics and Science Study: An Overview

Michael O. Martin

1.1 INTRODUCTION

The Third International Mathematics and Science Study (TIMSS) is the largest and most ambitious international comparative study of student achievement to date. Under the auspices of the International Association for the Evaluation of Educational Achievement (IEA), TIMSS brought together educational researchers from more than 50 countries to design and implement a study of the teaching and learning of mathematics and science in each country.

TIMSS is a cross-national survey of student achievement in mathematics and science that was conducted at three levels of the educational system. Forty-five countries took part in the survey (see Figure 1.1). The students, their teachers, and the principals of their schools were asked to respond to questionnaires about their backgrounds and their attitudes, experiences, and practices in the teaching and learning of mathematics and science. This report documents in detail the development and implementation of the TIMSS achievement survey.

A project of the magnitude of TIMSS necessarily has a long life cycle. Planning for TIMSS began in 1989; the first meeting of National Research Coordinators was held in 1990; data collection took place from the latter part of 1994 through 1995; the first international reports are planned for release in late 1996; and further international reports will be issued through 1997. A large number of people contributed to the many strands that made up TIMSS. They came from all areas of educational assessment and included specialists in policy analysis, curriculum design, survey research, test construction, psychometrics, survey sampling, and data analysis.

An achievement survey of the scale of TIMSS not only has a long life cycle, but also passes through several distinct stages. In the development stage, attention focuses on refining the aims of the study, establishing the parameters of the survey, designing and developing the data collection instruments, and developing data collection procedures. In the operational stage, samples are drawn, survey materials are distributed, training is conducted, and data are collected, checked, scored, and entered into databases. In the analysis and reporting stage, the data are processed, summarized, and presented, first in simple descriptive reports and later in more complex analytical volumes.

Figure 1.1 Countries Participating in the TIMSS Achievement Survey

Argentina	Hungary	Norway
Australia	Iceland	Philippines
Austria	Indonesia	Portugal
Belgium*	Iran, Islamic Republic	Romania
Bulgaria	Ireland	Russian Federation
Canada	Israel	Scotland
Colombia	Italy	Singapore
Cyprus	Japan	Slovak Republic
Czech Republic	Korea	Slovenia
Denmark	Kuwait	South Africa
England	Latvia	Spain
France	Lithuania	Sweden
Germany	Mexico	Switzerland
Greece	Netherlands	Thailand
Hong Kong	New Zealand	United States

* The Flemish and French educational systems in Belgium participated separately.

In addition to disseminating its findings as widely as possible, TIMSS aims to document fully the procedures and practices used to achieve the study goals. The *TIMSS Technical Report* is an important part of this effort. While the details of the TIMSS procedures are described in the various procedural manuals, this report presents the technical aspects of the study design, and provides the background to and the rationale for many of the design decisions taken.

Because of the long life cycle of TIMSS, and the involvement of so many individuals at its various stages, it was desired to issue the *TIMSS Technical Report* in two volumes, each documenting a major stage of the project and produced soon after the completion of that stage. Accordingly, the first volume documents the study design and the development of TIMSS up to, but not including, the operational stage of main data collection. The second volume will describe the operational stage, which consisted mainly of collecting and processing the data, and will describe the analytic procedures underlying the analysis and reporting phase of TIMSS.

1.2 THE CONCEPTUAL FRAMEWORK FOR TIMSS

IEA studies have as a central aim the measurement of student achievement in school subjects, with a view to learning more about the nature and extent of student achievement and the context in which it occurs. The ultimate goal is to isolate the factors directly relating to student learning that can be manipulated through policy changes in, for example, curricular emphasis, allocation of resources, or instructional practices. Clearly, an adequate understanding of the influences on student learning can come only from careful study of the nature of student achievement, and the characteristics of the learners themselves, the curriculum they follow, the teaching methods of their teachers, and the resources in their classrooms and their schools. Such school and classroom features are of course embedded in the community and the educational system, which in turn are aspects of society in general.

The designers of TIMSS chose to focus on curriculum as a broad explanatory factor underlying student achievement (Robitaille and Garden, 1996). From that perspective, curriculum was considered to have three manifestations: what society would like to see taught (the intended curriculum), what is actually taught in the classroom (the implemented curriculum), and what the students learn (the attained curriculum). This conceptualization was first developed for the IEA's Second International Mathematics Study (Travers and Westbury, 1989).

The three aspects of the curriculum bring together three major influences on student achievement. The intended curriculum states society's goals for teaching and learning. These expectations reflect the ideals and traditions of the greater society, and are constrained by the resources of the educational system. The implemented curriculum is what is taught in the classroom. Although presumably inspired by the intended curriculum, the actual classroom events are usually determined in large part by the classroom teacher, whose behavior may be greatly influenced by his or her own education, training, and experience, by the nature and organizational structure of the school, by interaction with teaching colleagues, and by the composition of the student body. The attained curriculum is what the students actually learn. Student achievement depends partly on the implemented curriculum and its social and educational context, and to a large extent on the characteristics of individual students, including ability, attitude, interests, and effort.

While the three-strand model of curriculum draws attention to three different aspects of the teaching and learning enterprise, it does have a unifying theme: the provision of educational opportunities to students. The curriculum, both as intended and as implemented, provides and delimits learning opportunities for students–a necessary though not sufficient condition for student learning.

Considering the curriculum in all its aspects as a channel through which learning opportunities are offered to students leads to a number of general questions that can be used to organize inquiry about that process. In TIMSS, four general research questions helped to guide the development of the study:

- What are students expected to learn?
- Who provides the instruction?
- How is instruction organized?

- What have students learned?

The first of these questions concerns the intended curriculum, and is addressed in TIMSS by an extensive comparative analysis of curricular documents and textbooks from each participating country. The second and third questions address major aspects of the implemented curriculum: what are the characteristics of the teaching force in each country (education, experience, attitudes and opinions), and how do teachers go about instructing their students (what teaching approaches do they use, and what curricular areas do they emphasize)? The final question deals with the attained curriculum: what have students learned, how does student achievement vary from country to country, and what factors are associated with student learning?

The study of the intended curriculum was a major part of the initial phase of the project. The TIMSS curriculum analysis consisted of an ambitious content analysis of curriculum guides, textbooks, and questionnaires completed by curriculum experts and educationalists. Its aim was a detailed rendering of the curricular intentions of the participating countries.

Data for the study of the implemented curriculum were collected as part of a large-scale international survey of student achievement. Questionnaires completed by the mathematics and science teachers of the students in the survey, and by the principals of their schools, provided information about the topics in mathematics and science that were taught, the instructional methods adopted in the classroom, the organizational structures that supported teaching, and the factors that were seen to facilitate or inhibit teaching and learning.

The student achievement survey provides data for the study of the attained curriculum. The wide-ranging mathematics and science tests that were administered to nationally representative samples of students at three levels of the educational system provide not only a sound basis for international comparisons of student achievement, but a rich resource for the study of the attained curriculum in each country. Information about students' characteristics, and about their attitudes, beliefs, and experiences, comes from a questionnaire completed by each participating student. This information will help to identify the student characteristics associated with learning and provide a context for the study of the attained curriculum.

1.3 THE TIMSS CURRICULUM FRAMEWORKS

The TIMSS curriculum frameworks (Robitaille et al., 1993) were conceived early in the study as an organizing structure within which the elements of school mathematics and science could be described, categorized, and discussed. In the TIMSS curriculum analysis, the frameworks provided the system of categories by which the contents of textbooks and curriculum guides were coded and analyzed. The same system of categories was used to collect information from teachers about what mathematics and science they have taught. Finally, the system formed a basis for constructing the TIMSS achievement tests.

The TIMSS curriculum frameworks have their antecedents in the content-by-cognitive-behavior grids used in earlier studies (e.g., Travers and Westbury, 1989) to categorize curriculum units or achievement test items. A content-by-cognitive-behavior grid is usually represented as a matrix, or two-dimensional array, where the horizontal dimension represents a hierarchy of behavior levels at which students may perform, while the vertical dimension specifies subject-matter topics or areas. Individual items or curriculum units are assigned to a particular cell of the matrix. These grids facilitate comparisons of curricula and the development of achievement tests by summarizing curriculum composition and test scope.

The TIMSS curriculum frameworks are an ambitious attempt to expand the concept of the content-by-cognitive-behavior grids.

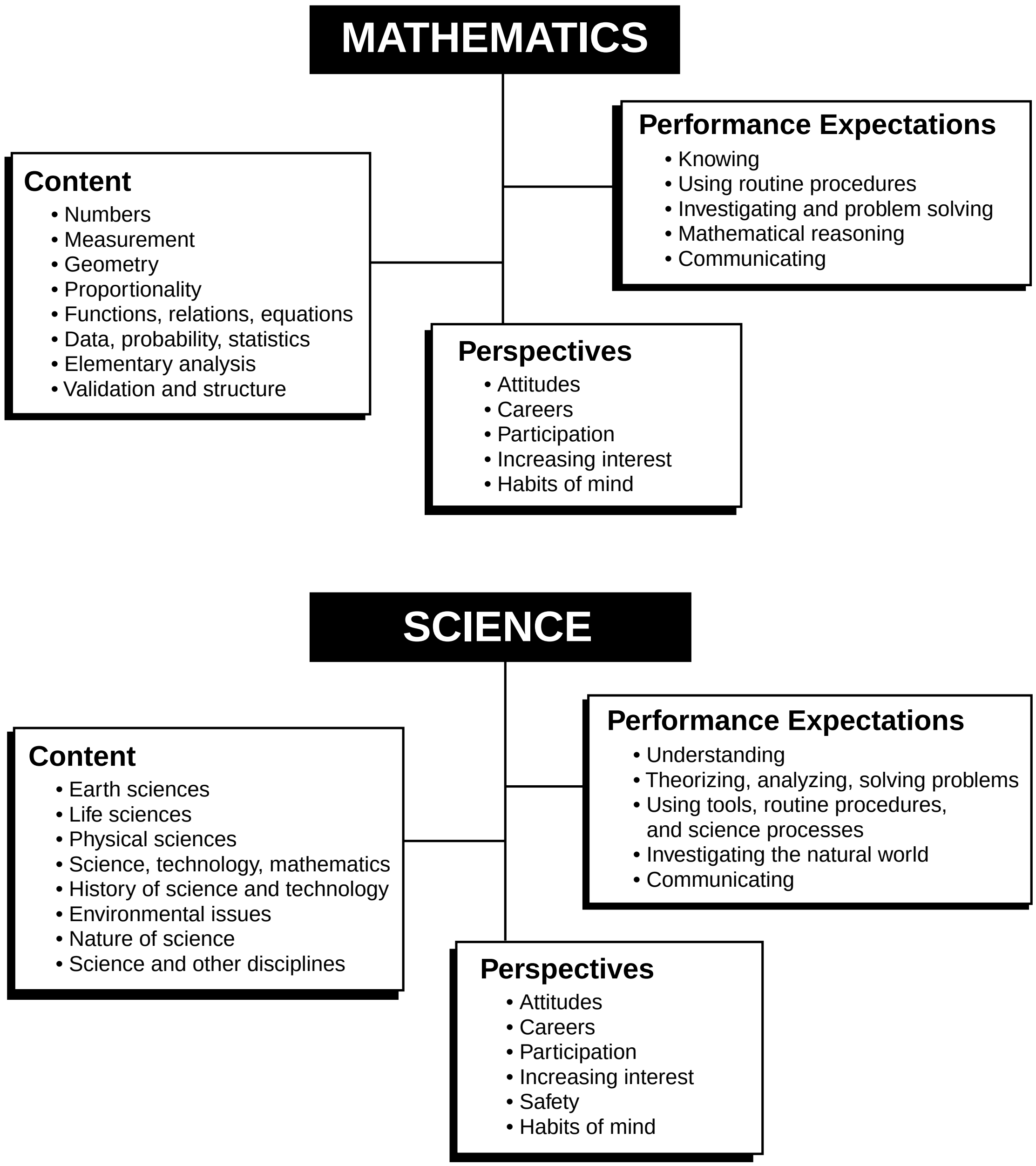
> For the purposes of TIMSS, curriculum consists of the concepts, processes, and attitudes of school mathematics and science that are intended for, implemented in, or attained during students' schooling experiences. Any piece of curriculum so conceived–whether intended, implemented, or attained, whether a test item, a paragraph in an "official" curriculum guide, or a block of material in a student textbook—may be characterized in terms of three parameters: subject-matter content, performance expectations, and perspectives or context (Robitaille et al., 1993, 43).

Subject-matter content, performance expectations, and perspectives constitute the three dimensions, or aspects, of the TIMSS curriculum frameworks. *Subject-matter content* refers simply to the content of the mathematics or science curriculum unit or test item under consideration. *Performance expectations* are a reconceptualization of the earlier cognitive behavior dimension. Their purpose is to describe, in a non-hierarchical way, the many kinds of performance or behavior that a given test item or curriculum unit might elicit from students. The *perspectives* aspect is relevant to analysis of documents such as textbooks, and is intended to permit the categorization of curricular components according to the nature of the discipline as reflected in the material, or the context within which the material is presented.

Each of the three aspects is partitioned into a number of categories, which are themselves partitioned into subcategories, which are further partitioned as necessary. The curriculum frameworks (the major categories are shown in Figure 1.2) were developed separately for mathematics and science. Each framework has the same general structure, and includes the same three aspects: subject-matter content, performance expectations, and perspectives.[1]

[1] The complete TIMSS curriculum frameworks can be found in Robitaille, D., et al. (1993). *TIMSS Monograph No. 1: Curriculum Frameworks for Mathematics and Science*. Vancouver: Pacific Educational Press.

Figure 1.2 The Major Categories of the TIMSS Curriculum Frameworks



1.4 THE TIMSS CURRICULUM ANALYSIS

The TIMSS analysis of the intended curriculum focused on curriculum guides, textbooks, and experts as the sources of information about each country's curricular intentions. The investigation of variations in curricula across countries involved three major data collection efforts: (1) a detailed page-by-page document analysis of curriculum guides and selected textbooks; (2) mapping (or tracing) the coverage of topics in the TIMSS frameworks across textbook series and curriculum guides for all pre-university grades; and (3) collecting questionnaire data designed to characterize the organization of the educational system, the decision-making process regarding learning goals, and the general contexts for learning mathematics and science.

In the document analysis, the participating countries partitioned the curriculum guides and textbooks into homogeneous blocks and coded the substance of each block according to the TIMSS frameworks. The document analysis provided detailed information for the grades studied, but does not allow tracing the full continuum of topic coverage through all the grades in the pre-university system. Information on continuity of coverage was obtained by tracing topics through the curriculum from the beginning of schooling to the end of secondary school. The topic tracing for TIMSS included two procedures. In the first, curriculum experts within each country characterized the points at which instruction is begun, finalized, and concentrated on for all topics in the frameworks. In this effort, each topic was treated discretely even though many of the topics are related in terms of their specification in the learning goals. Therefore, for six topics each within mathematics and the sciences, a second tracing procedure was used, based on the curriculum guides that specified how subtopics fit together in the coverage of a topic as a whole. The twelve topics were selected as being of special interest to the mathematics and science education communities. Taken together, the two tracing procedures offer both breadth, covering all topics across all grades, and depth in terms of covering a limited number of topics across all grades (Beaton, Martin and Mullis, in press).

The TIMSS curriculum analysis was conducted by the Survey of Mathematics and Science Opportunities (SMSO) project of Michigan State University, under the direction of William H. Schmidt. The initial results of this study will be presented in two volumes: *Many Visions, Many Aims: A Cross National Investigation of Curricular Intentions in School Mathematics* (Schmidt, W., McKnight, C., Valverde, G., Houang, R., and Wiley, D., in press) and *Many Visions, Many Aims: A Cross National Investigation of Curricular Intentions in School Science* (Schmidt, W., Raizen, S., Britton, E., Bianchi, L., Wolfe, R., in press).

1.5 THE STUDENT POPULATIONS

TIMSS chose to study student achievement at three points in the educational process: at the earliest point at which most children are considered old enough to respond to written test questions (Population 1); at a point at which students in most countries have finished primary education and are beginning secondary education (Population 2); and at the end of secondary education (Population 3). The question whether student populations should be defined by chronological age or grade level in school is one that faces all comparative surveys of student achievement. TIMSS has addressed this issue by defining (for Populations 1 and 2) the target population as the pair of adjacent grades that contains the largest proportion of a particular one-year age group (9-year-olds for Population 1, and 13-year-olds for Population 2). Most cross-country comparisons in TIMSS will be based on grade levels, since educational systems are organized around grade levels, but it will also be possible to make cross-country comparisons on the basis of student age for countries where the pair of adjacent grades contains a high percentage of the age cohort.

The student populations in TIMSS are defined as follows.

- Population 1: all students enrolled in the two adjacent grades that contain the largest proportion of students of age 9 years at the time of testing.
- Population 2: all students enrolled in the two adjacent grades that contain the largest proportion of students of age 13 years at the time of testing.
- Population 3: all students in their final year of secondary education, including students in vocational education programs.

  Population 3 has two optional subpopulations:

  - *Students taking advanced courses in mathematics*
  - *Students taking advanced courses in physics.*

Population 2 was compulsory for all participating countries. Countries could choose whether or not to participate in Populations 1 and 3 (and the subpopulations of Population 3).

1.6 SURVEY ADMINISTRATION DATES

Since school systems in countries in the Northern and Southern Hemispheres do not have the same school year, TIMSS had to operate two survey administration schedules. The schedules are shown in Table 1.1. These periods were chosen with the aim of testing students as late in the school year as practicable, so as to reflect the knowledge gained throughout the year.

Table 1.1 Survey Administration Dates

	Populations 1 and 2	**Population 3**
Southern Hemisphere	September-November 1994	August 1995
Northern Hemisphere	February-May 1995	February-May 1995

1.7 THE TIMSS ACHIEVEMENT TESTS

The measurement of student achievement in a school subject is a challenge under any circumstances. The measurement of student achievement in two subjects at three student levels in 45 countries (through the local language of instruction), in a manner that does justice to the curriculum to which the students have been exposed and that allows the students to display the full range of their knowledge and abilities, is indeed a formidable task. This, nonetheless, is the task that TIMSS set for itself.

Although the IEA had conducted separate studies of student achievement in mathematics and science on two earlier occasions (mathematics in 1964 and 1980-82, and science in 1970-71 and 1983-84), TIMSS was the first IEA study to test mathematics and science together. Since there is a limit to the amount of student testing time that may reasonably be requested, assessing student achievement in two subjects simultaneously constrains the number of questions that may be asked, and therefore limits the amount of information that may be collected from any one student.

Recent IEA studies, particularly the Second International Mathematics Study (Robitaille and Garden, 1989), placed great emphasis on the role of curriculum in all its manifestations in the achievement of students. This concern with curriculum coverage, together with the desire of curriculum specialists and educators generally to ensure that both subjects be assessed as widely as possible, led to pressure for very ambitious coverage in the TIMSS achievement tests. Further, there was concern that the assessment of student knowledge and abilities be as “authentic” as possible, with the questions asked and the problems posed in a form that students are used to in their everyday school experience. In particular, there was a requirement that test items make use of a variety of task types and response formats, and not be exclusively multiple choice.

Reconciling the demands for the form and extent of the TIMSS achievement tests was a lengthy and difficult process. It involved extensive consensus building through which the concerns of all interested parties had to be balanced in the interests of producing a reliable measuring instrument that could serve as a valid index of student achievement in mathematics and science in all of the participating countries. The tests that finally emerged were necessarily a compromise between what might have been attempted in an ideal world of infinite time and resources, and the real world of short timelines and limited resources.

Despite the need for compromise in some areas, the TIMSS achievement tests have gone a long way toward meeting the ideals of their designers. They cover a wide range of subject matter, yielding, in Population 2 for example, estimates of student proficiency in 11 areas or “reporting categories” of mathematics and science, as well as overall mathematics and science scores. The test items include both multiple-choice and free-response items. The latter come in two varieties: “short-answer,” where the student supplies a brief written response; and “extended-response,” where students must provide a more extensive written answer, and sometimes explain their reasoning. The free-response items are scored using a unique two-digit coding rubric that yields both a score for the response and an indication of the nature of the response. The free-response data will be a rich source of information about student understanding and misunderstanding of mathematics and science topics.

The wide coverage and detailed reporting requirements of the achievement tests resulted in a pool of mathematics and science items in Population 2 that, if all of it were to be administered to any one student, would take almost seven hours of testing. Since the consensus among the National Research Coordinators was that 90 minutes was the most that could be expected for this population, a way of dividing the item pool among the students had to be found. Matrix sampling provided a solution to this problem by assigning subsets of items to individual students in such a way as to produce reliable estimates of the performance of the population on all the items, even though no student has responded to the entire item pool. The TIMSS test design uses a variant of matrix sampling to map the mathematics and science item pool into eight student booklets for each Population 1 and Population 2, and nine booklets for Population 3.

The TIMSS test design sought breadth of subject-matter coverage and reliable reporting of summary statistics for each of the reporting categories. However, because of the interest in the details of student performance at the item level, at least some of the items also had to be administered to enough students to permit accurate reporting of their item statistics. The TIMSS item pool for both Populations 1 and 2 was therefore divided into 26 sets, or clusters, of items. These were then arranged in various ways to make up eight test booklets, each containing seven item clusters. One cluster, the core cluster, appears in each booklet. Seven "focus" clusters appear in three of the eight booklets. The items in these eight clusters should be sufficient to permit accurate reporting of their statistics. There are also 12 "breadth" clusters, each of which appears in just one test booklet. These help ensure wide coverage, but the accuracy of their statistics may be relatively low. Finally, there are eight "free-response clusters," each of which appears in two booklets. These items are a rich source of information about the nature of student responses, and should have relatively accurate statistics.

The eight student booklets were distributed systematically in each classroom, one per student. This is efficient from a sampling viewpoint, and, since there are eight substantially different booklets in use in each classroom, it reduces the likelihood of students copying answers from their neighbors.

The approach to assessing achievement in mathematics and science at Population 3 was quite different from that for the younger populations. At Population 3 there are really three populations to be tested. For all students in the population (the final year of secondary schooling), TIMSS plans to report measures of mathematics and science literacy. The item pool for this domain consists of four clusters of items, assembled in two booklets distributed across the entire Population 3 sample. The other two populations to be reported on are the students at Population 3 who are taking advanced mathematics courses, and those taking physics courses. Since each group will have received advanced preparation, item pools had to be developed that treated these subjects with the appropriate depth of coverage and level of difficulty.

There are four clusters of advanced mathematics items, assembled into three booklets for students taking courses in advanced mathematics. The pool of physics items is also grouped into four clusters and assembled into three booklets distributed to the students in the sample who are taking physics courses. A ninth booklet, consisting of one cluster of literacy items, one of mathematics items, and one of physics items, is designed just for students taking courses in both advanced mathematics and physics.

1.8 PERFORMANCE ASSESSMENT

Educators have long advocated the use of using practical tasks to assess student performance in mathematics and particularly in science. The inclusion of such a "performance assessment" was a design goal from the beginning of TIMSS. The performance expectations aspect of the TIMSS curriculum frameworks explicitly mentions skills such as measurement, data collection, and use of equipment that cannot be assessed with traditional paper and pencil tests. However, the obstacles to including a performance assessment component in a study like TIMSS are formidable. The difficulties inherent in developing a valid international measure of student achievement using just paper and pencil are greatly compounded when developing a practical test of student performance. In addition to the usual difficulties of translation and adaptation, there is the question of standardization of materials and of administration procedures, and the greatly increased cost of data collection. The TIMSS performance assessment was designed to obtain measures of students' responses to hands-on tasks in mathematics and science and to demonstrate the feasibility of including a performance assessment in a large-scale international student assessment. It was conducted at Populations 1 and 2 only.

The performance assessment in TIMSS consists of a set of 13 tasks, 12 of which are used at Population 1 and 12 at Population 2. While 11 of the tasks are common to both populations, there were important differences in presentation. For the younger students (Population 1), the tasks were presented with more explicit instructions, or "scaffolding," while for the older students (Population 2) there were usually more activities to be done, or additional questions to be answered.

The tasks were organized into a circuit of nine stations, with each station consisting of one long task (taking about 30 minutes to complete) or two shorter tasks (which together took about 30 minutes). An administration of the performance assessment required nine students, who were a subsample of the students selected for the main survey, and 90 minutes of testing time. Each student visited three of the stations during this time; the choice of stations and the order in which they were visited was determined by a task assignment plan.

Because of the cost and complexity of this kind of data collection endeavor, the performance assessment was an optional component of the study. The performance assessment component of TIMSS was conducted by 21 of the 45 countries participating in Population 2, and by 11 of the 28 countries participating in Population 1.

1.9 THE CONTEXT QUESTIONNAIRES

To obtain information about the contexts for learning mathematics and science, TIMSS included questionnaires for the participating students, their mathematics and science teachers, and the principals of their schools. The student and school questionnaires were administered at all three populations, and the questionnaires for mathematics and science teachers at Populations 1 and 2. National Research Coordinators (NRCs) provided information about the structure of their educational systems, educational decision-making processes, qualifications required for teaching, and course structures in mathematics and science. In an exercise to investigate the curricular relevance of the TIMSS achievement tests, NRCs were asked to indicate which items, if any, are not included in their country's intended curriculum. The results of this Test-Curriculum Matching Analysis will be reported in the first international reports.

The student questionnaire addresses students' attitudes towards mathematics and science, parental expectations, and out-of-school activities. Students also were asked about their classroom activities in mathematics and the sciences, and about the courses they had taken. A special version of the student questionnaire was prepared for countries where physics, chemistry, and biology are taught as separate subjects. Although not strictly related to the question of what students have learned in mathematics or science, characteristics of pupils can be important correlates for understanding educational processes and attainments. Therefore, students also provided general home and demographic information.

The teacher questionnaires had two sections. The first section covered general background information about preparation, training, and experience, and about how teachers spend their time in school. Teachers also were asked about the amount of support and resources they received in fulfilling their teaching duties. The second part of the questionnaire related to instructional practices in the classrooms selected for TIMSS testing. To obtain information about the implemented curriculum, teachers were asked how many periods the class spent on topics from the TIMSS curriculum frameworks. They also were asked about their use of textbooks in teaching mathematics and science and about the instructional strategies used in the class, including the use of calculators and computers. In optional sections of the questionnaire, teachers were asked to review selected items from the achievement tests and indicate whether their students had been exposed to the content covered by the items, and to respond to a set of questions that probed their pedagogic beliefs. At Population 2, there were separate versions of the questionnaire for mathematics teachers and science teachers. The TIMSS design did not include a teacher questionnaire for teachers of Population 3 students.

The school questionnaire was designed to provide information about overall organization and resources. It asked about staffing, facilities, staff development, enrollment, course offerings, and the amount of school time for students, primarily in relation to mathematics and science instruction. School principals also were asked about the functions that schools perform in maintaining relationships with the community and students' families.

1.10 MANAGEMENT AND OPERATIONS

Like all previous IEA studies, TIMSS was essentially a cooperative venture among independent research centers around the world. While country representatives came together to plan the study and to agree on instruments and procedures, participants were each responsible for conducting TIMSS in their own country, in accordance with the international standards. Each national center provided its own funding, and contributed to the support of the international coordination of the study. A study of the scope and magnitude of TIMSS offers a tremendous operational and logistic challenge. In order to yield comparable data, the achievement survey must be replicated in each participating country in a timely and consistent manner. This was the responsibility of the NRC in each country. Among the major responsibilities of NRCs in this regard were the following.

- Meeting with other NRCs and international project staff to plan the study and to develop instruments and procedures
- Defining the school populations from which the TIMSS samples were to be drawn, selecting the sample of schools using an approved random sampling procedure, contacting the school principals and securing their agreement to participate in the study, and selecting the classes to be tested, again using an approved random sampling procedure
- Translating and adapting all of the tests, questionnaires, and administration manuals into the language of instruction of the country (and sometimes into more than one language) prior to data collection
- Assembling, printing, and packaging the test booklets and questionnaires, and shipping the survey materials to the participating schools
- Ensuring that the tests and questionnaires were administered in participating schools, either by teachers in the school or by an external team of test administrators, and that the completed test protocols were returned to the TIMSS national center
- Conducting a quality assurance exercise in conjunction with the test administration, whereby some testing sessions were observed by an independent observer to confirm that all specified procedures were followed
- Recruiting and training individuals to score the free-response questions in the achievement tests, and implementing the plan for coding the student responses, including the plan for assessing the reliability of the coding procedure
- Recruiting and training data entry personnel for keying the responses of students, teachers, and principals into computerized data files, and conducting the data-entry operation, using the software provided
- Checking the accuracy and integrity of the data files prior to shipping them to the IEA Data Processing Center in Hamburg.

In addition to their role in implementing the TIMSS data collection procedures, NRCs were responsible for conducting analyses of their national data, and for reporting on the results of TIMSS in their own countries.[2]

The TIMSS International Study Director was responsible for the overall direction and coordination of the project. The TIMSS International Study Center, located at Boston College in the United States, was responsible for supervising all aspects of the design and implementation of the study at the international level. This included the following.

- Planning, conducting and coordinating all international TIMSS activities, including meetings of the International Steering Committee, NRCs, and advisory committees
- Development, including field testing, of all data collection instruments

[2] A list of the TIMSS National Research Coordinators is provided in Appendix A.

- Development of sampling procedures for efficiently selecting representative samples of students in each country, and monitoring sampling operations to ensure that they conformed to TIMSS requirements
- Development and documentation of operational procedures to ensure efficient collection of all TIMSS data
- Design and implementation of a quality assurance program encompassing all aspects of the TIMSS data collection, including monitoring of test administration sessions in participating countries
- Supervision of the checking and cleaning of the data from the participating countries, and construction of the TIMSS international database, including the computation of sampling weights and the scaling of the achievement items
- Analysis of international data, and writing and dissemination of international reports.

The International Study Center was supported in its work by the following advisory committees.[3]

- The International Steering Committee advises on policy issues and on the general direction of the study.
- The Subject Matter Advisory Committee advises on all matters relating to mathematics and science subject matter, particularly the content of the achievement tests.
- The Technical Advisory Committee advises on all technical issues related to the study, including study design, sampling design, achievement test construction and scaling, questionnaire design, database construction, data analysis, and reporting.
- The Performance Assessment Committee developed the TIMSS performance assessment and advises on the analysis and reporting of the performance assessment data.
- The Free-Response Item Coding Committee developed the coding rubrics for the free-response items.
- The Quality Assurance Committee helped to develop the TIMSS quality assurance program.
- The Advisory Committee on Curriculum Analysis advised the International Study Director on matters related to the curriculum analysis.

Several important TIMSS functions, including test and questionnaire development, translation checking, sampling consultations, data processing, and data analysis, were conducted by centers around the world, under the direction of the TIMSS International Study Center. In particular, the following centers have played important roles in the TIMSS project.

- The International Coordinating Center (ICC), in Vancouver, Canada, was responsible for international project coordination prior to the establishment of the International Study Center in August 1993. Since then, the ICC has provided support to the International Study Center, and in particular has managed translation verification in the achievement test development process; and has published several monographs in the TIMSS monograph series.

[3] See Appendix A for membership of TIMSS committees.

- The IEA Data Processing Center (DPC), located in Hamburg, Germany is responsible for checking and processing all TIMSS data and for constructing the international database. The DPC played a major role in developing and documenting the TIMSS field operations procedures.
- Statistics Canada, located in Ottawa, Canada, is responsible for advising NRCs on their sampling plans, for monitoring progress in all aspects of sampling, and for the computation of sampling weights.
- The Australian Council for Educational Research (ACER), located in Melbourne, Australia, has participated in the development of the achievement tests, has conducted psychometric analyses of field trial data, and was responsible for the development of scaling software and for scaling the achievement test data.

As Sampling Referee, Keith Rust of WESTAT, Inc., (United States) worked with Statistics Canada and the NRCs to ensure that sampling plans met the TIMSS standards, and advised the International Study Director on all matters relating to sampling.

1.11 SUMMARY OF THE REPORT

In Chapter 2, Robert Garden and Graham Orpwood (subject-matter coordinators in mathematics and science, respectively) describe the long and sometimes arduous process of developing the TIMSS achievement tests. They outline the tensions between the wish for comprehensive coverage of mathematics and science curricula, the limited time available for student testing, the need to be sensitive to curricular variations from country to country, the desire to include innovative assessment methods, and the requirements of a rigorous approach to measuring student achievement. The authors describe how these tensions were resolved, the compromises that were made, and the characteristics of the final pool of achievement items. They show how the items in this pool address a wide range of subject matter in mathematics and science, and seek to evoke from the students a wide range of performance types, from exhibiting acquired knowledge to engaging in complex problem solving.

In Chapter 3, Raymond Adams and Eugenio Gonzalez present the TIMSS achievement test design. The design describes, for each student population, how the pool of achievement questions was organized into achievement booklets that were given to the students selected to take part in TIMSS. Since the entire item pool was much too large to be administered to every student in the time available, a matrix sampling approach was used in which subsets of items drawn from the total item pool were administered to random subsamples of students. This procedure provides accurate estimates of population parameters based on all items, even though not every student responds to every item.

Pierre Foy, Keith Rust, and Andreas Schleicher describe in Chapter 4 the student populations that were the target of TIMSS, and the designs that were developed to draw samples from these populations. They pay particular attention to the principles by which the target populations were defined in participating countries. This process involved specifying exactly which students were eligible for selection as a sample, and which subgroups (e.g., mentally handicapped students), if any, were to be excluded. The authors present the sampling-precision requirements of TIMSS, and show how these were used to determine sample size in the participating countries. They go on to describe the TIMSS sampling designs, including the use of stratification and multistage sampling, and illustrate the general method used in sampling schools in TIMSS (the sampling of classrooms is described in Chapter 9 on field operations).

The background of the development of the student, teacher, and school questionnaires is the subject of Chapter 5, by William Schmidt and Leland Cogan. The difficulties devising achievement tests that can provide valid and reliable data from countries with diverse social, cultural, and educational traditions are discussed in Chapter 2; the difficulties of developing questionnaires that can elicit useful information about the educational context in an array of countries are no less formidable. Factors that are fundamental to the understanding of student achievement in one country may be much less relevant in another. Schmidt and Cogan recount the enormous efforts made to build consensus on the conceptual framework for TIMSS and to derive from it a set of data collection instruments that would be useable in all participating countries and would do justice to the aims of the project.

In measuring student achievement, TIMSS sought to ensure the validity of its tests as well as their reliability by combining traditional objective multiple-choice items with innovative and challenging free-response questions. The economic realities of a large-scale international assessment dictated that the data be collected mainly through written responses to written questions. However, TIMSS acknowledged that a comprehensive assessment of student competency in mathematics, and particularly in science, demands that students also be given realistic problems that must be answered by manipulating tools and materials. The TIMSS performance assessment, which is described by Maryellen Harmon and Dana Kelly in Chapter 6, was developed to meet that need. The performance assessment is a set of tasks that were administered to a subsample of the students selected for the main survey. Although an integral feature of the overall TIMSS design, the extra expense of conducting the performance assessment was beyond the resources of some participants. Accordingly, it was presented as an optional component.

TIMSS was committed from an early stage to measuring student achievement through a variety of item types. The main survey contains three types of items: multiple-choice, short-answer, and extended-response. All of the items in the performance assessment are either short-answer or extended-response. Unlike multiple-choice items, short-answer and extended-response items are free-response items that require a coding rubric, or set of rules, so that a code may be assigned to each response. Svein Lie, Alan Taylor, and Maryellen Harmon in Chapter 7, describe the evolution of the innovative two-digit coding rubric used with all TIMSS free-response items. This coding rubric provides for coding the “correctness” of a response to an item as a score level in the left digit, and information about the detailed nature of the response in both digits together. This coding rubric promises to provide a rich store of information about the most common student responses to the free-response items, and in particular about the most common misconceptions about mathematics and science concepts..

In order to implement the TIMSS survey in the 45 participating countries, it was necessary to translate the achievement tests, the student, teacher, and school questionnaires, and in many cases the manuals and tracking forms from English, the language in which they were developed, to the language of the country. In all, the TIMSS instruments were translated into 30 languages. Even where the language of testing was English, cultural adaptations had to be made to suit national language usage. In Chapter 8, Beverley Maxwell describes the procedures that were used to ensure that the translations and cultural adaptations made in each country produced local versions that corresponded closely in meaning to the international versions, and in particular that the items in the achievement tests were not made easier or more difficult through translation.

As a comparative sample survey of student achievement conducted simultaneously in 45 countries, TIMSS depends crucially on its data collection procedures to obtain high-quality data. In Chapter 9, Andreas Schleicher and Maria Teresa Siniscalco describe the procedures developed to ensure that the TIMSS data were collected in a timely and cost-effective manner while adhering to the highest standards of survey research. The authors outline the extensive list of procedural manuals that describe in detail all aspects of the TIMSS field operations. They describe also the software systems that were provided to participants to help them conduct their data collection activities. The development of practical and effective methods of sampling classes within schools that would be generally applicable was a particular challenge to TIMSS. The authors present these methods in some detail, and particularly emphasize the system of documentation that is an integral component of each procedure. This documentation consists of a series of tracking forms, which record how schools, classes, teachers, and students were selected to take part in the study. This documentation facilitates the operational aspects of data collection, and provides vital information for other aspects of the study (e.g., the computation of sampling weights, the provision of checks for quality assurance).

As a consequence of its emphasis on authentic and innovative assessment of achievement, much of the testing time was used to provide answers to free-response items. Approximately one-third of student time for the main survey, and all of the time for the performance assessment, was devoted to free-response items. This resulted in a huge number of student responses that had to be coded using the two-digit coding scheme described in Chapter 7. In order to code reliably and in the same way in each country, and to ensure that the performance assessment was administered consistently across countries, an extensive training program for National Research Coordinators and their staff members was conducted. In Chapter 10, Ina Mullis, Chancey Jones, and Robert Garden outline the content and format of this training program, and describe the logistics of conducting ten training meetings in locations all around the globe.

A major part of the role of the TIMSS International Study Center was to ensure that all aspects of the study were carried out to the highest standards of survey research. In Chapter 11, Michael Martin, Ina Mullis, and Dana Kelly describe the procedures and activities that comprised the TIMSS quality assurance program. The International Study

Center sought to ensure a high-quality study in the first instance by providing participants with complete and explicit documentation of all procedures and materials, supplemented by meetings with consultants and training meetings at every opportunity. An integral part of the documentation for each procedure was a set of forms that had to be completed in order to carry out the procedure. The completed forms constitute a record that can be reviewed for quality control purposes. A major component of the quality assurance activities was a program of visits to each participating country by monitors appointed by the International Study Center. These quality control monitors visited the national research centers and interviewed the NRCs about all aspects of the implementation of TIMSS. They also visited a sample of ten of the schools taking part in the study to interview the School Coordinator and Test Administrator, and to observe a test administration in one classroom.

1.12 SUMMARY

This report provides an overview of the main features of the TIMSS project, and summarizes the technical background to the development of the study. The development of the achievement tests and questionnaires, the sampling and operations procedures, the procedures for coding the free-response items, and quality assurance activities are all described in detail. The activities involved in the collection of the TIMSS data, and in analysis and reporting, will be presented in a subsequent volume.

REFERENCES

Beaton, A.E., Martin, M.O., and Mullis, I.V.S. (in press). Providing Data for Educational Policy in an International Context: The Methodologically Ambitious Nature of TIMSS. *European Journal of Psychological Assessment.*

Robitaille, D.F. and Garden, R.A. (1996). "Design of the Study" in D.F. Robitaille and R.A. Garden (eds.), *TIMSS Monograph No. 2: Research Questions & Study Design*. Vancouver, Canada: Pacific Educational Press.

Robitaille, D.F. and Garden, R.A. (1989). *The IEA Study of Mathematics II: Contexts and Outcomes of School Mathematics.* Oxford: Pergamon Press.

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

Schmidt, W.H., McKnight, C.C., Valverde, G.A., Houang, R.T., and Wiley, D.E. (in press). *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Mathematics*. Norwell, MA: Kluwer Academic Press.

Schmidt, W.H., Raizen, S.A., Britton, E.D., Bianchi, L.J., and Wolfe, R.G. (in press). *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Science.* Norwell, MA: Kluwer Academic Press.

Travers, K.J. and Westbury, I. (1989). *The IEA Study of Mathematics I: Analysis of Mathematics Curricula*. Oxford: Pergamon Press.