Appendix

Overview of TIMSS Procedures for the Performance Assessment

History

TIMSS represents the continuation of a long series of studies conducted by the International Association for the Evaluation of Educational Achievement (IEA). Since its inception in 1959, the IEA has conducted more than 15 studies of cross-national achievement in curricular areas such as mathematics, science, language, civics, and reading. IEA conducted its First International Mathematics Study (FIMS) in 1964, and the Second International Mathematics Study (SIMS) in 1980-82. The First and Second International Science Studies (FISS and SISS) were conducted in 1970-71 and 1983-84, respectively. Since the subjects of mathematics and science are related in many respects and since there is broad interest in many countries in their students' abilities in both mathematics and science, the third studies were conducted together as an integrated effort.

The number of participating countries, the number of grades tested, and the assessment of mathematics and science simultaneously has resulted in TIMSS becoming the largest, most complex IEA study to date and the largest international study of educational achievement ever undertaken. Traditionally, IEA studies have systematically worked toward gaining more in-depth understanding of how various factors contribute to the overall outcomes of schooling. Particular emphasis has been given to refining our understanding of students' opportunity to learn as this opportunity becomes successively defined and implemented by curricular and instructional practices.

In an effort to extend what had been learned from previous studies and provide contextual and explanatory information, TIMSS expanded beyond the already substantial task of measuring achievement in two subject areas by also including a thorough investigation of curriculum and how it is delivered in classrooms around the world. In addition, extending the work of previous IEA studies, TIMSS included a performance assessment. In IEA's Second International Science Study a small subset of the participating countries administered practical tasks. TIMSS built on this experience and included more countries, more tasks, and a greater emphasis on an integration of conceptual knowledge and process skills. The inclusion of a performance assessment in TIMSS also reflected the current movement in education to measure students' understanding and competence with hands-on assessments.

THE COMPONENTS OF TIMSS

Continuing the approach of previous IEA studies, TIMSS addressed three conceptual levels of curriculum. The **intended curriculum** is composed of the mathematics and science instructional and learning goals as defined at the system level. The **implemented curriculum** is the mathematics and science curriculum as interpreted by teachers and made available to students. The **attained curriculum** is the mathematics and science content that students have learned and their attitudes towards these subjects. To aid in interpretation and comparison of results, TIMSS also collected extensive information about the social and cultural contexts for learning.

Nearly 50 countries participated in one or more of the various components of the TIMSS data collection effort, including the curriculum analysis. To gather information about the intended curriculum, mathematics and science specialists within each participating country worked section by section through curriculum guides, textbooks, and other curricular materials to categorize aspects of these materials in accordance with detailed specifications derived from the TIMSS mathematics and science curriculum frameworks.[1] Initial results from this component of TIMSS can be found in two companion volumes: *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Mathematics* and *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Science.*[2]

To measure the attained curriculum, TIMSS tested more than half a million students in mathematics and science at five grade levels. TIMSS included testing at three separate populations:

Population 1: Students enrolled in the two adjacent grades that contained the largest proportion of 9-year-old students at the time of testing – third- and fourth-grade students in most countries.

Population 2: Students enrolled in the two adjacent grades that contained the largest proportion of 13-year-old students at the time of testing – seventh- and eighth-grade students in most countries.

Population 3: Students in their final year of secondary education. As an additional option, countries could test two special subgroups of these students: students taking advanced courses in mathematics and students taking courses in physics.

Countries participating in the study were required to administer tests to the students in the two grades at Population 2 but could choose whether or not to participate at the other levels. Ten countries that participated in Population 1 and 21 countries that participated in Population 2 also administered the performance assessment to subsamples of the upper-grade students (eighth graders and fourth graders in most countries) who completed the written tests. Figure A.1 shows the countries that participated in the various components of TIMSS testing.

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

[2] Schmidt, W.H., McKnight, C.C., Valverde, G.A., Houang, R.T., and Wiley, D.E. (1997). *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Mathematics.* Dordrecht, the Netherlands: Kluwer Academic Publishers.

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.* Dordrecht, the Netherlands: Kluwer Academic Publishers.

TIMSS administered a broad array of questionnaires to collect data about how the curriculum is implemented in classrooms and about the social and cultural contexts for learning. Questionnaires were administered at the country level about decision-making and organizational features within the educational systems. The students who were tested answered questions pertaining to their attitudes towards mathematics and science, classroom activities, home background, and out-of-school activities. The mathematics and science teachers of sampled students responded to questions about teaching emphasis on the topics in the curriculum frameworks, instructional practices, textbook use, professional training and education, and their views on mathematics and science. The heads of schools responded to questions about school staffing and resources, mathematics and science course offerings, and support for teachers. In addition, a volume was compiled that presents descriptions of the educational systems of the participating countries.[3]

Achievement results and background data for Populations 1 and 2 (third, fourth, seventh, and eighth grades in many countries) have been published in four volumes.[4]

[3] Robitaille D.F. (Ed.). (1997). *National Contexts for Mathematics and Science Education: An Encyclopedia of Education Systems Participating in TIMSS*. Vancouver, B.C.: Pacific Educational Press.

[4] Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., and Smith, T.A. (1997). *Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College.

Martin, M.O., Mullis, I.V.S., Beaton, A.E., Gonzalez, E.J., Smith, T.A., and Kelly, D.L. (1997). *Science Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College

Beaton, A.E., Mullis, I.V.S., Martin, M.O. Gonzalez, E.J., Kelly, D.L., and Smith, T.A. (1996). *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College.

Beaton, A.E., Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., Smith, T.A., and Kelly, D.L. (1996). *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College.

Countries Participating in Components of TIMSS Testing

Figure A.1

Country	Population 1		Population 2		Population 3		
	Written Test	Performance Assessment	Written Test	Performance Assessment	Mathematics & Science Literacy	Advanced Mathematics	Physics
Argentina			●				
Australia	●	●	●	●	●	●	●
Austria	●		●		●	●	●
Belgium (Fl)			●				
Belgium (Fr)			●				
Bulgaria			●				
Canada	●	●	●	●	●	●	●
Colombia			●	●			
Cyprus	●	●	●	●	●	●	●
Czech Republic	●		●	●	●	●	●
Denmark			●		●	●	●
England	●		●	●			
France			●		●	●	●
Germany			●		●	●	●
Greece	●		●			●	
Hong Kong	●	●	●	●			
Hungary	●		●		●		
Iceland	●		●		●		
Indonesia	●		●				
Iran, Islamic Rep.	●	●	●	●			
Ireland	●		●				
Israel	●	●	●	●	●	●	●
Italy	●		●		●		
Japan	●		●				
Korea	●		●				
Kuwait	●		●				
Latvia	●		●				●
Lithuania			●		●	●	
Mexico	●		●		●	●	●
Netherlands	●		●	●	●		
New Zealand	●	●	●	●	●		
Norway	●		●	●	●		●
Philippines			●				
Portugal	●	●	●	●			
Romania			●	●			
Russian Federation			●		●	●	●
Scotland	●		●	●			
Singapore	●		●	●			
Slovak Republic			●				
Slovenia	●	●	●	●	●	●	●
South Africa			●		●		
Spain			●	●			
Sweden			●	●	●	●	●
Switzerland			●	●	●	●	●
Thailand	●		●				
United States	●	●	●	●	●	●	●

DEVELOPING THE TIMSS PERFORMANCE ASSESSMENT TASKS

The TIMSS performance assessment was developed by experts in mathematics, science, and performance assessment from participating countries. It was designed to reflect the TIMSS mathematics and science curriculum frameworks and be feasible for administration in a large-scale international assessment. In particular, attention was focused on developing tasks that represented the range of performance expectations in the TIMSS curriculum frameworks. The TIMSS Performance Assessment Committee developed a set of tasks, some of which were adapted from versions used in assessments in the United Kingdom, Australia, New Zealand, and the United States. In 1994, 22 tasks (at each population level) were field-tested in 19 countries. Based on the student results from this field test and evaluations of each task by field-test administrators, National Research Coordinators, and mathematics and science subject matter experts in the participating countries, 12 tasks for each population were selected for the performance assessment. Task selection was based on breadth of coverage, feasibility of obtaining materials and administering the tasks, time constraints, quality considerations noted by task reviewers, and field-test item statistics.[5] Difficulties in standardizing the use of live materials and soils, and differing climate effects – for example, great difficulty in keeping a moisture indicator dry in maritime climates – resulted in elimination of certain tasks, chiefly in the life and earth science areas, and reduced the overall content coverage to the physical sciences, mathematics, and human biology. There were 13 tasks altogether; 11 were administered at both grades, although adaptations were made in the form of scaffolding for the younger age group, one unique task was administered to the fourth grade, and one unique task administered to the eighth grade. Table A.1 summarizes the mathematics and science content areas and the performance expectations associated with each of the performance assessment tasks.

The TIMSS performance assessment instruments were prepared in English and translated into the languages of administration. In addition, it sometimes was necessary to adapt the international versions for cultural purposes, even in the countries that tested in English. This process represented an enormous effort for the national centers, with many checks along the way. The translation effort included: 1) developing explicit guidelines for translation and cultural adaptation, 2) translation of the instruments by the national centers in accordance with the guidelines and by using two or more independent translators, 3) consultation with subject matter experts regarding cultural adaptations to ensure that the meaning and difficulty of items did not change, 4) verification of the quality of the translations by professional translators from an independent translation company, 5) correction by the national centers in accordance with the suggestions made, 6) verification that corrections were made, and 7) a series of statistical checks after the testing to detect items that did not perform comparably across countries.[6]

[5] See Chapter 1 of this report for a display of each task and student responses. Details of the criteria used for task selection are provided in Harmon, M. and Kelly, D.L. (1996). "Performance Assessment" in M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report, Volume I.* Chestnut Hill, MA: Boston College.

[6] More details about the translation verification procedures can be found in: Mullis, I.V.S., Kelly, D.L., and Haley, K. (1996). "Translation Verification Procedures" in M.O. Martin and I.V.S. Mullis (Eds.), *Third International Mathematics and Science Study: Quality Assurance in Data Collection.* Chestnut Hill, MA: Boston College. Maxwell, B. (1996). "Translation and Cultural Adaptation of the TIMSS Instruments" in M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report, Volume I.* Chestnut Hill, MA: Boston College.

Description of Performance Assessment Tasks with Associated Content Knowledge Areas and Performance Expectations (1 of 3)

Table A.1

Science Tasks

Task	Description	Content Areas		Performance Expectations
		Overall Task Content Areas	Specific Knowledge Areas	
Pulse	Student investigates changes in pulse rate during exercise; records and analyzes data; and explains results.	Life Science	• Life Process and Systems – Energy handling • Human Biology	• Conducting investigations • Gathering, organizing, and representing data • Interpreting investigational data • Applying scientific principles to develop explanations
Magnets	Student determines the stronger of two magnets and describes strategies to support conclusion.	Physical Science	• Energy and Physical Processes – Magnetism	• Conducting investigations • Interpreting investigational data • Formulating conclusions from investigational data
Batteries	Student determines which of four batteries are worn out; describes strategy; and uses concept knowledge to explain proper arrangement of batteries in a flashlight.	Physical Science	• Energy and Physical Processes – Electricity	• Conducting investigations • Interpreting investigational data • Formulating conclusions from investigational data • Applying scientific principles to solve problems and develop explanations
Rubber Band	Student investigates the effect on the length of a rubber band from attaching increasing numbers of masses; then explains results.	Physical Science	• Physical Properties of Matter – Elasticity	• Conducting investigations • Gathering, organizing, and representing data • Interpreting and extrapolating data • Applying scientific principles to develop explanations
Solutions	Student investigates the effect of different solvent temperatures on rate of solution; collects, records, and analyzes data; and explains results.	Physical Science	• Physical Properties of Matter – Solubility • Structure of Matter – Atoms, ions, molecules • Energy and Physical Processes – Heat and temperature • Physical Transformation – Dissolving – Explanations of physical changes	• Designing and conducting investigations • Using equipment • Gathering, organizing, and representing data • Formulating conclusions from investigational data • Applying scientific principles to develop explanations
Containers	Student investigates the effect of different container materials on heat transfer; draws a conclusion about the best insulator; and applies concept to a new, seemingly quite different problem.	Physical Science	• Physical Properties of Matter – Specific heat and temperature	• Conducting investigations • Using equipment • Gathering, organizing, and representing data • Formulating conclusions from investigational data • Applying scientific principles to develop explanations and solve new problems

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table A.1 Description of Performance Assessment Tasks with Associated Content Knowledge Areas and Performance Expectations (Continued – 2 of 3)

Combination Tasks

Task	Description	Content Areas		Performance Expectations
		Overall Task Content Areas	Specific Knowledge Areas	
Shadows	Student manipulates the positions of light source and object to find three positions where the shadow is twice the width of the object, and expresses the relationships of distances of the light and object to the screen as a general rule.	Physical Science	• Energy and Physical Processes – Light	**Science** • Conducting investigations • Gathering, organizing, and representing data • Interpreting investigational data • Formulating conclusions from investigational data • Applying scientific principles to develop explanations **Mathematics** • Performing routine and complex mathematical procedures • Problem solving • Conjecturing • Generalizing
		Mathematics	• Measurement and Units • Geometry: Position, Visualization, and Shape – Two-dimensional polygons • Geometry: Symmetry, Congruency and Similarity • Proportionality Problems	
Plasticine	Given only two standard masses, student develops and describes strategies to weigh lumps of various specified masses.	Physical Science	• Physical Properties of Matter – Weight and balance	**Science** • Conducting routine experimental procedures • Applying scientific principles to solve quantitative problems **Mathematics** • Performing routine mathematical procedures • Problem solving • Developing and describing strategy
		Mathematics	• Measurement and Units • Proportionality Concepts and Problems	

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Description of Performance Assessment Tasks with Associated Content Knowledge Areas and Performance Expectations (Continued – 3 of 3)

Table A.1

Mathematics Tasks

Task	Description	Content Areas		Performance Expectations
		Overall Task Content Areas	Specific Knowledge Areas	
Dice	Student applies a given algorithm to numbers that are obtained from successive throws of a die, then explains why one resulting number occurs more frequently than others.	Mathematics	• Whole Number Operations • Data Representation and Analysis • Probability	• Performing routine and complex mathematical procedures • Conjecturing
Calculator	Student uses calculator for a series of multiplications, identifies pattern in the products, describes and extrapolates the pattern to solve a new problem. At eighth grade, student also draws on knowledge of number properties and factoring to find a set of factors.	Mathematics	• Whole Numbers: Meaning and Operations • Data Representation and Analysis	• Using equipment • Recalling mathematical objects and properties • Performing routine and complex mathematical procedures • Developing and describing strategy • Predicting
Folding and Cutting	Student reproduces patterns of increasing complexity by folding along axes of symmetry and cutting paper. At eighth grade, this is extended to drawing lines of symmetry without manipulating materials.	Mathematics	• Geometry: Symmetry Transformations	• Problem solving • Predicting
Around the Bend	Student uses models to determine which "furniture" will go around the bend in a corridor, uses scale to convert model dimensions (in centimeters) to real furniture dimensions (in meters), makes judgements about real-world applications, and develops a general rule.	Mathematics	• Measurement and Units • Geometry: Position, Visualization and Shape – Two-dimensional polygons – Three-dimensional • Proportionality Problems	• Performing routine and complex mathematical procedures • Problem solving
Packaging	Student constructs boxes for three unique arrangements of four balls tightly packed; sketches nets for each box, and draws one net to actual size.	Mathematics	• Measurement and Units • Geometry: Position, Visualization and Shape – Three-dimensional	• Performing routine and complex mathematical procedures • Problem solving

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Performance Assessment Design and Administration Procedures

The performance assessment was administered in a "circus" format in which the materials for 12 tasks (at each grade) were assembled at stations and students visited the stations according to one of two rotation plans to which they were assigned during the sampling process. In each administration, there were nine stations with materials for either one or two tasks. Students visited three stations, completing three to five tasks altogether. Each student spent 30 minutes at each station. The administration was designed to accommodate 9 students; at each school, either 9 or 18 students participated in the performance assessment.[7]

After completing the tasks at each station, students submitted their work booklets to the performance assessment administrator, together with any products. The work recorded in the booklets and any products created during the assessment were evaluated by coders specially trained to use the TIMSS scoring rubrics (see section on scoring the TIMSS performance assessment).

Each participating country was responsible for providing the materials for the performance assessment tasks and for administering the performance assessment, in accordance with the international procedures. The *Performance Assessment Administration Manual* specified the materials required for the tasks, the organization of tasks at stations in a "circus," assignment of students to stations, and all other aspects of the administration session. During the administration, the performance assessment administrator ensured that the students visited the correct stations and that supplies were replenished as necessary, and collected students' work. Several regional training sessions were conducted around the world during which representatives from the participating countries were trained in equipment set-up and administration procedures.

Sample Implementation And Participation Rates

The selection of valid and efficient samples is crucial to the quality and success of an international comparative study such as TIMSS. The accuracy of the survey results depends on the quality of the available sampling information and on the quality of the sampling activities themselves. For TIMSS, National Research Coordinators (NRCs) worked on all phases of sampling with staff from Statistics Canada. NRCs received training in how to select the school and student samples and in the use of the sampling software. In consultation with the TIMSS sampling referee (Keith Rust, Westat, Inc.), staff from Statistics Canada reviewed the national sampling plans, sampling data, sampling frames, and sample execution. This documentation was used by the TIMSS International Study Center in consultation with Statistics Canada, the sampling referee, and the Technical Advisory Committee to evaluate the quality of the samples.

The sample of schools and students for the performance assessment was a subsample of the schools and students that participated in the main written assessment. Consequently, the characteristics of each country's performance assessment sample reflect the quality of the sampling for the written assessment and compliance with the guidelines for the performance assessment sampling.

In a few situations where it was not possible to implement TIMSS for all of Populations 1 and 2, as specified by the international desired population definition – all students in the upper grade of the two adjacent grades with the largest proportion of 9-year-olds (Population 1) and 13-year-olds (Population 2) – countries were permitted to define a national desired population that did not include part of the international desired population. Tables A.2 and A.3 show any differences in coverage between the international and national desired populations, at the upper grades of the target

[7] For more information on the performance assessment design see Harmon, M. and Kelly, D.L. (1996). "Performance Assessment" in M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report Volume I.* Chestnut Hill, MA: Boston College.

populations (eighth grade and fourth grade in most countries). Most participants achieved 100% coverage. The countries with less than 100% coverage are identified in the tables in this report. Israel defined its tested population according to the structure of its school system and tested only schools in the Hebrew education system. Switzerland administered TIMSS in the German-speaking cantons only.

For the main written assessment, within the desired population countries could define a population that excluded a small percentage (less than 10%) of certain kinds of schools or students that would be very difficult or resource-intensive to test (e.g., schools for students with special needs or schools that were very small or located in extremely remote areas). For the performance assessment, in the interest of ensuring the quality of the administration, countries could exclude additional schools if the schools had fewer than nine students in the upper grade and thus could not provide a full complement of students for the performance assessment rotation or if the schools were in a remote region. The exclusion rate for the performance assessment sample was not to exceed 25% of the national desired population. Tables A.2 and A.3 show the main assessment school exclusion rates, the performance assessment school exclusion rates, the within-sample exclusion rates, and the overall exclusion rates for the eighth and fourth grades, respectively. For various reasons, at the eighth grade England and Romania exceeded the 25% limit for performance assessment exclusions. At the fourth grade only New Zealand exceeded this limit. The exclusion rates for these countries are noted in the tables in this report.

For the main assessment, TIMSS used a two-stage sample design at Populations 1 and 2, where the first stage involved selecting 150 public and private schools within each country. Within each school, countries were to use random procedures to select one fourth grade and one third grade mathematics class for Population 1 and one eighth grade and one seventh grade mathematics class at Population 2 (or the corresponding upper and lower grades in that country). All of the students in those two classes were to participate in the TIMSS testing. This approach was designed to yield a representative sample of approximately 7,500 students per country, with approximately 3,750 students at each grade.[8]

For the performance assessment, TIMSS participants were to sample at least 50 schools from those already selected for the written assessment, and from each school a sample of either 9 or 18 upper-grade students already selected for the written assessment. This yielded a sample of about 450 students in each of the eighth and fourth grades in each country. Typically, between 150 and 300 students in a country responded to each performance assessment task. Tables A.4 and A.5 show the school and student sample sizes for the main assessment sample and the performance assessment subsample for the eighth grade. Tables A.6 and A.7 show the corresponding information for the fourth grade.

Countries were required to obtain a participation rate of at least 85% of both schools and students, or a combined rate (the product of school and student participation) of 75%. Tables A.8 and A.9 present, for the eighth and fourth grades, respectively, the school, student, and overall participation rates for the main assessment and the performance assessment. Because the performance assessment sample is drawn from the main assessment sample, the participation rates achieved for the performance assessment reflect the participation of schools and students in the main assessment, as well as those for the performance assessment administration.

[8] The sample design for TIMSS is described in detail in Foy, P., Rust, K., and Schleicher, A. (1996). "TIMSS Sample Design" in M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study, Technical Report, Volume I.* Chestnut Hill, MA: Boston College.

Table A.2 Coverage of TIMSS Target Population - Performance Assessment – Eighth Grade*

The international desired target population is defined as follows:
Eighth Grade - All students enrolled in the higher of the two adjacent grades with the largest proportion of 13-year-old students at the time of testing.

Country	International Desired Target Population		National Desired Target Population			
	Coverage	Notes	Main Assessment School-Level Exclusions	Performance Assessment School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
Australia	100%		0.2%	16.3%	0.6%	17.0%
Canada	100%		2.4%	15.0%	1.8%	19.1%
Colombia	100%		3.8%	0.0%	0.0%	3.8%
Cyprus	100%		0.0%	0.0%	0.0%	0.0%
Czech Republic	100%		4.9%	0.0%	0.0%	4.9%
[2] England	100%		8.4%	16.6%	2.4%	27.3%
Hong Kong	100%		2.0%	1.0%	0.0%	3.0%
Iran, Islamic Rep.	100%		0.3%	17.0%	0.0%	17.3%
[1] Israel	74%	Hebrew Public Education System	3.1%	0.0%	0.0%	3.1%
Netherlands	100%		1.2%	0.0%	0.0%	1.2%
New Zealand	100%		1.3%	10.5%	0.4%	12.1%
Norway	100%		0.3%	22.6%	1.5%	24.4%
Portugal	100%		0.0%	0.0%	0.3%	0.3%
[3] Romania	100%		2.8%	28.5%	0.0%	31.3%
Scotland	100%		0.3%	9.3%	1.7%	11.3%
Singapore	100%		4.6%	0.0%	0.0%	4.6%
Slovenia	100%		2.4%	0.7%	0.2%	3.2%
Spain	100%		6.0%	1.7%	2.6%	10.3%
Sweden	100%		0.0%	23.5%	0.7%	24.2%
[1] Switzerland	75%	German Cantons	4.4%	8.4%	0.8%	13.6%
United States	100%		0.4%	1.3%	1.7%	3.4%

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

* Eighth grade in most countries; see Table 2 for information about the grades tested in each country.
[1] National Desired Population does not cover all of International Desired Population.
[2] National Defined Population covers less than 90 percent of National Desired Population for the main assessment (school-level plust within-sample exclusions).
[3] School-level exclusions for performance assessment exceed 25% of the National Desired Population.
Because results are rounded to the nearest whole number, some totals may appear inconsistent.

Coverage of TIMSS Target Population - Performance Assessment – Fourth Grade* Table A.3

The international desired target population is defined as follows:
Fourth Grade - All students enrolled in the higher of the two adjacent grades with the largest proportion of 9-year-old students at the time of testing.

Country	International Desired Target Population		National Desired Target Population			
	Coverage	Notes	Main Assessment School-Level Exclusions	Performance Assessment School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
Australia	100%		0.1%	15.1%	1.4%	16.7%
Canada	100%		2.5%	15.4%	3.1%	21.0%
Cyprus	100%		3.1%	0.0%	0.1%	3.2%
Hong Kong	100%		2.6%	1.9%	0.0%	4.6%
Iran, Islamic Rep.	100%		0.3%	17.5%	0.9%	18.7%
[2] Israel	72%	Hebrew Public Education System	1.1%	0.0%	0.1%	1.2%
[1] New Zealand	100%		0.7%	25.8%	0.4%	27.0%
Portugal	100%		6.6%	0.0%	0.7%	7.3%
Slovenia	100%		1.9%	0.7%	0.0%	2.6%
United States	100%		0.4%	0.0%	4.3%	4.7%

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

* Fourth grade in most countries; see Table 2 for information about the grades tested in each country.
[1] School-level exclusions for performance assessment exceed 25% of the National Desired Population.
[2] National Desired Population does not cover all of International Desired Population.
Because results are rounded to the nearest whole number, some totals may appear inconsistent.

Table A.4 TIMSS School Sample Sizes - Performance Assessment - Eighth Grade*

Country	Main Assessment					Performance Assessment				
	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample That Participated in Main Assessment	Number of Replacement Schools That Participated in Main Assessment	Total Number of Schools That Participated in Main Assessment	Number of Schools Eligible for Performance Assessment	Number of Schools Sampled for Performance Assessment	Number of Original Schools That Participated	Number of Replacement Schools That Participated	Total Number of Schools That Participated in Performance Assessment
Australia	214	214	158	3	161	127	51	37	5	42
Canada	413	388	363	1	364	312	80	77	0	77
Colombia	150	150	136	4	140	150	54	49	0	49
Cyprus	55	55	55	0	55	55	50	48	0	48
Czech Republic	150	149	143	6	149	149	50	47	3	50
England	150	144	80	41	121	100	50	26	24	50
Hong Kong	105	104	85	0	85	84	50	27	0	27
Iran, Islamic Rep.	192	191	191	0	191	150	50	49	0	49
Israel	100	100	45	1	46	20	20	19	1	20
Netherlands	150	150	36	59	95	74	50	18	31	49
New Zealand	150	150	137	12	149	134	50	45	5	50
Norway	150	150	136	10	146	113	50	44	5	49
Portugal	150	150	142	0	142	150	50	48	0	48
Romania	176	176	163	0	163	95	50	50	0	50
Scotland	153	153	119	8	127	136	50	39	9	48
Singapore	137	137	137	0	137	137	50	46	4	50
Slovenia	150	150	121	0	121	149	50	49	1	50
Spain	155	154	147	6	153	146	50	47	3	50
Sweden	120	120	116	0	116	91	50	50	0	50
Switzerland	259	258	247	3	250	158	50	36	8	44
United States	220	217	169	14	183	216	107	76	6	82

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

* Eighth grade in most countries; see Table 2 for information about the grades tested in each country.

TIMSS Student Sample Sizes - Performance Assessment – Eighth Grade* Table A.5

Country	Main Assessment						Performance Assessment
	Number of Sampled Students in Participating Schools	Number of Students Withdrawn from Class/School	Number of Students Excluded	Number of Students Eligible	Number of Students Absent	Total Number of Students Assessed	Total Number of Students Assessed
Australia	8027	63	61	7903	650	7253	564
Canada	9240	134	206	8900	538	8362	1240
Colombia	2843	6	0	2837	188	2649	455
Cyprus	3045	15	0	3030	107	2923	414
Czech Republic	3608	6	0	3602	275	3327	450
England	2015	37	60	1918	142	1776	440
Hong Kong	3415	12	0	3403	64	3339	217
Iran, Islamic Rep.	3770	20	0	3750	56	3694	436
Israel	1453	6	0	1447	32	1415	171
Netherlands	2112	14	1	2097	110	1987	435
New Zealand	4038	121	12	3905	222	3683	824
Norway	3482	26	49	3407	140	3267	438
Portugal	3589	70	13	3506	115	3391	430
Romania	3899	0	0	3899	174	3725	450
Scotland	3289	0	46	3243	380	2863	424
Singapore	4910	18	0	4892	248	4644	450
Slovenia	2869	15	8	2846	138	2708	451
Spain	4198	27	102	4069	214	3855	449
Sweden	4483	71	28	4384	309	4075	433
Switzerland	4989	16	24	4949	94	4855	396
United States	8026	104	108	7814	727	7087	712

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

* Eighth grade in most countries; see Table 2 for information about the grades tested in each country.

Table A.6 TIMSS School Sample Sizes - Performance Assessment - Fourth Grade*

Country	Main Assessment					Performance Assessment				
	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample That Participated in Main Assessment	Number of Replacement Schools That Participated in Main Assessment	Total Number of Schools That Participated in Main Assessment	Number of Schools Eligible for Performance Assessment	Number of Schools Sampled for Performance Assessment	Number of Original Schools That Participated	Number of Replacement Schools That Participated	Total Number of Schools That Participated in Performance Assessment
Australia	268	268	169	9	178	122	50	41	5	46
Canada	423	420	390	0	390	319	84	75	1	76
Cyprus	150	150	146	0	146	150	50	49	1	50
Hong Kong	156	148	124	0	124	120	50	37	10	47
Iran, Islamic Rep.	180	180	180	0	180	140	50	49	1	50
Israel	100	100	40	0**	87	100	54	27	18	45
New Zealand	150	150	120	29	149	103	50	39	11	50
Portugal	150	150	143	0	143	150	50	48	0	48
Slovenia	150	150	121	0	121	149	50	49	1	50
United States	220	213	182	0	182	212	106	88	1	89

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

* Fourth grade in most countries; see Table 2 for information about the grades tested in each country.

**Does not include 47 replacement schools that were selected using unapproved methods.

TIMSS Student Sample Sizes - Performance Assessment – Fourth Grade* Table A.7

Country	Main Assessment						Performance Assessment
	Number of Sampled Students in Participating Schools	Number of Students Withdrawn from Class/School	Number of Students Excluded	Number of Students Eligible	Number of Students Absent	Total Number of Students Assessed	Total Number of Students Assessed
Australia	6930	37	104	6789	282	6507	513
Canada	9193	81	268	8844	436	8408	1150
Cyprus	3972	4	3	3965	589	3376	444
Hong Kong	4475	0	1	4474	63	4411	421
Iran, Islamic Rep.	3521	5	36	3480	95	3385	440
Israel	2486	0	3	2483	132	2351	402
New Zealand	2627	82	20	2525	104	2421	613
Portugal	2994	15	16	2963	110	2853	430
Slovenia	2720	3	0	2717	151	2566	447
United States	8224	61	412	7751	455	7296	777

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

* Fourth grade in most countries; see Table 2 for information about the grades tested in each country.

Table A.8 TIMSS Participation Rates - Performance Assessment - Eighth Grade*

Country	Main Assessment					Performance Assessment				
	School Participation Rate Before Replacement (Weighted Percentage)	School Participation Rate After Replacement (Weighted Percentage)	Within-School Student Participation Rate (Weighted Percentage)	Overall Participation Rate Before Replacement (Weighted Percentage)	Overall Participation Rate After Replacement (Weighted Percentage)	School Participation Rate Before Replacement (Weighted Percentage)	School Participation Rate After Replacement (Weighted Percentage)	Within-School Student Participation Rate (Weighted Percentage)	Overall Participation Rate Before Replacement (Weighted Percentage)	Overall Participation Rate After Replacement (Weighted Percentage)
Australia	75%	77%	92%	69%	70%	51%	58%	73%	37%	43%
Canada	90%	91%	93%	84%	84%	97%	97%	92%	89%	89%
Colombia	91%	93%	94%	85%	87%	91%	91%	96%	88%	88%
Cyprus	100%	100%	97%	97%	97%	96%	96%	93%	88%	88%
Czech Republic	96%	100%	92%	89%	92%	94%	100%	82%	77%	82%
England	56%	85%	91%	51%	77%	46%	85%	84%	38%	71%
Hong Kong	82%	82%	98%	81%	81%	44%	44%	77%	34%	34%
Iran, Islamic Rep.	100%	100%	98%	98%	98%	98%	98%	93%	91%	91%
Israel	45%	46%	98%	44%	45%	44%**	46%**	30%**	13%**	14%**
Netherlands	24%	63%	95%	23%	60%	18%	48%	89%	16%	43%
New Zealand	91%	99%	94%	86%	94%	90%	100%	88%	79%	88%
Norway	91%	97%	96%	87%	93%	87%	96%	91%	79%	88%
Portugal	95%	95%	97%	92%	92%	96%	96%	91%	87%	87%
Romania	94%	94%	96%	89%	89%	90%	90%	94%	84%	84%
Scotland	79%	83%	88%	69%	73%	78%	96%	85%	66%	81%
Singapore	100%	100%	95%	95%	95%	90%	100%	87%	79%	87%
Slovenia	81%	81%	95%	77%	77%	98%	100%	93%	91%	93%
Spain	96%	100%	95%	91%	94%	94%	100%	93%	87%	93%
Sweden	97%	97%	93%	90%	90%	99%	99%	88%	87%	87%
Switzerland	93%	95%	98%	92%	94%	65%	81%	97%	63%	78%
United States	77%	85%	92%	71%	78%	71%	77%	86%	61%	66%

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

* Eighth grade in most countries; see Table 2 for information about the grades tested in each country.
**Unweighted participation rates.

TIMSS Participation Rates - Performance Assessment - Fourth Grade* Table A.9

Country	Main Assessment					Performance Assessment				
	School Participation Rate Before Replacement (Weighted Percentage)	School Participation Rate After Replacement (Weighted Percentage)	Within-School Student Participation Rate (Weighted Percentage)	Overall Participation Rate Before Replacement (Weighted Percentage)	Overall Participation Rate After Replacement (Weighted Percentage)	School Participation Rate Before Replacement (Weighted Percentage)	School Participation Rate After Replacement (Weighted Percentage)	Within-School Student Participation Rate (Weighted Percentage)	Overall Participation Rate Before Replacement (Weighted Percentage)	Overall Participation Rate After Replacement (Weighted Percentage)
Australia	66%	69%	96%	63%	66%	47%	56%	76%	36%	43%
Canada	90%	90%	96%	86%	86%	91%	92%	95%	87%	88%
Cyprus	97%	97%	86%	83%	83%	98%	100%	86%	85%	86%
Hong Kong	84%	84%	98%	83%	83%	61%	77%	95%	58%	73%
Iran, Islamic Rep.	100%	100%	97%	97%	97%	97%	100%	93%	90%	93%
Israel	40%	40%	94%	38%	38%	50% **	83% **	30% **	15% **	25% **
New Zealand	80%	99%	96%	77%	95%	72%	93%	90%	65%	83%
Portugal	95%	95%	96%	92%	92%	96%	96%	94%	91%	91%
Slovenia	81%	81%	94%	76%	76%	98%	100%	91%	89%	91%
United States	85%	85%	94%	80%	80%	83%	84%	88%	73%	74%

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

* Fourth grade in most countries; see Table 2 for information about the grades tested in each country.
**Unweighted participation rates.

Compliance With Sampling Guidelines

Figure A.2 shows how countries have been grouped in tables in this report. Countries that complied with the TIMSS guidelines for grade selection and classroom sampling, and that achieved acceptable participation rates – 85% of both the schools and students or a combined rate (the product of school and student participation) of 75% with or without replacement schools, are shown in the first panel of Figure A.2. These countries (12 at the eighth grade and 5 at the fourth grade) appear in the tables in this report ordered by achievement. Countries that met the guidelines only after including replacement schools are so labeled.

Countries that did not reach at least 50% school participation without the use of replacement schools, or that failed to reach the sampling participation standard even with the inclusion of replacement schools, are shown in the second panel of Figure A.2. These countries are presented in a separate section of the tables presenting task performance.

To provide a better curricular match, for the written assessment at Population 2, Colombia, Romania, and Slovenia elected to test their seventh- and eighth-grade students, even though that meant not testing the two grades with the most 13-year-olds. Their students were thus somewhat older than those in the other countries. As a result, the students sampled for the performance assessment (eighth graders) also are somewhat older than those in other countries. At Population 1, Slovenia tested their third- and fourth-grade students for the written assessment even though these were not the two grades with the most 9-year-olds. Consequently, their fourth graders who were sampled for the performance assessment are somewhat older than students in other countries. Colombia and Romania did not participate in TIMSS at the primary grades. See Table A.10 for the percentages of 9- and 13-year-olds in the target grades. In this report, Colombia, Romania, and Slovenia are presented in alphabetical order in a separate section of the tables presenting task results.

At the eighth grade, Hong Kong's sample size for the performance assessment was very small due to low school participation, and thus its eighth-grade results are presented in Appendix B. Israel did not completely comply with the TIMSS within-school sampling procedures at the eighth and fourth grades and it had a small sample size at the eighth grade; its results are also presented in Appendix B.

Countries Grouped for Reporting of Performance Assessment Results According to Their Compliance with Guidelines for Sample Implementation and Participation Rates

Figure A.2

	Eighth Grade	Fourth Grade
Countries satisfying guidelines for sample participation rates, grade selection and sampling procedures	Canada Cyprus Czech Republic Iran, Islamic Republic New Zealand Norway Portugal † Scotland Singapore Spain Sweden †1 Switzerland	Canada Cyprus Iran, Islamic Republic †3 New Zealand Portugal
Countries not satisfying guidelines for sample participation rates	Australia 2 England Netherlands United States	Australia Hong Kong United States
Countries not meeting age/grade specifications (high percentage of older students)	Colombia 3 Romania Slovenia	Slovenia
Countries with small sample sizes	Hong Kong	
Countries with unapproved sampling procedures	4 Israel	Israel

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

† Met guidelines for sample participation rates only after replacement schools were included.

1 National Desired Population does not cover all of International Desired Population (see Table A.2) - German-speaking cantons only.

2 National Defined Population covers less than 90 percent of National Desired Population for the main assessment (see Table A.2).

3 School-level exclusions for performance assessment exceed 25% of the National Desired Population (see Tables A.2 and A.3).

4 Israel also had a small size at the eighth grade.

Table A.10 Coverage of 13-Year Old and 9-Year-Old Students

Country	13-Year-Old Students			9-Year-Old-Students		
	Percent in Lower Grade (Seventh Grade*)	Percent in Upper Grade (Eighth Grade*)	Percent in Both Grades	Percent in Lower Grade (Third Grade*)	Percent in Upper Grade (Fourth Grade*)	Percent in Both Grades
Australia	64	28	92	65	29	94
Canada	48	43	91	46	48	94
Colombia	30	15	45	.	.	.
Cyprus	28	70	98	35	63	98
Czech Republic	73	17	90	.	.	.
England	57	42	99	.	.	.
Hong Kong	44	46	90	43	50	93
Iran, Islamic Rep.	47	25	72	51	32	83
Israel	-	-	-	-	-	-
Netherlands	59	31	90	.	.	.
New Zealand	52	47	99	50	49	99
Norway	43	57	100	.	.	.
Portugal	44	32	76	45	48	93
Romania	67	9	76	.	.	.
Scotland	24	75	99	.	.	.
Singapore	82	15	97	.	.	.
Slovenia	65	2	67	60	0	60
Spain	46	39	85	.	.	.
Sweden	45	54	99	.	.	.
Switzerland	48	44	92	.	.	.
United States	58	33	91	61	34	95

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

* Seventh, eighth, third, and fourth grades in most countries; see Table 2 for information about the upper grades tested in each country. The international definition is the two adjacent grades with the largest proportion of 13-year-old students, and the two with the largest proportion of 9-year-old students.

A dash (-) indicates data are not available. Israel did not test the lower grades.

A dot (.) indicates country did not participate in performance assessment at the fourth grade.

Scoring the Performance Assessment

TIMSS developed detailed scoring rubrics in order to obtain the maximum amount of information from the constructed responses and to evaluate students' work reliably. The scoring system for the performance assessment used the same type of two-digit codes as the free-response items of the written test.[9] The first digit designates the correctness level of the response (3, 2, 1, or 0 points). The second digit, combined with the first, represents a diagnostic code used to identify specific types of approaches, strategies, or common errors and misconceptions. Although not used in this report, analyses of responses based on the complete two-digit code should provide insight into ways to help students better understand science and mathematics concepts and problem-solving approaches.

To meet the goal of implementing reliable scoring procedures based on the TIMSS rubrics, the TIMSS International Study Center prepared guides containing the rubrics and explanations of how to apply them, together with example student responses for the various rubric categories. These guides, together with additional practice responses, were used as a basis for a series of regional training sessions. These were designed to assist representatives of national centers who would then be responsible for training personnel in their respective countries to apply the two-digit codes reliably.[10]

To gather and document empirical information about the within-country agreement among scorers, TIMSS developed a procedure whereby systematic subsamples of approximately 10% of the students' responses in each country were to be coded independently by two different scorers. Tables A.11 and A.12 display the inter-coder agreement for the eighth and fourth grades, respectively. Data are presented for 12 countries at the eighth grade and for 4 countries at the fourth grade. Unfortunately, lack of resources prevented several countries from providing this information. The range and average across all performance assessment items of percent exact agreement are reported for both the correctness score and the full two-digit diagnostic code. A high percentage of exact agreement was observed for most items, especially at the correctness score level. At the eighth grade, the average percent exact agreement across items for the correctness score ranged from 79% to 100% across countries, with an overall average for all 12 countries of 91%. At the fourth grade, the country-level averages ranged from 91% to 99%, with an overall average of 93%. It should be noted that due to the smaller sample sizes in the performance assessment, in some countries only a small number of student responses for each item were available in the reliability sample.

[9] For more information on the TIMSS scoring procedures, see Lie, S., Taylor, A., and Harmon, M. (1996). "Scoring Techniques and Criteria" in M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report, Volume I.* Chestnut Hill, MA: Boston College.

[10] The procedures used in the training sessions are documented in Mullis, I.V.S., Garden, R.A., and Jones, C.A. (1996). "Training for Scoring the TIMSS Free-Response Items" in M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report, Volume I.* Chestnut Hill, MA: Boston College.

Table A.11 TIMSS Inter-Coder Agreement for Performance Assessment – Eighth Grade*

Country	Correctness Score Agreement			Diagnostic Code Agreement			Average Number of Student Responses per Item in the Reliability Sample▼
	Average Percent of Exact Agreement Across Tasks	Range of Percent of Exact Agreement		Average Percent of Exact Agreement Across Tasks	Range of Percent of Exact Agreement		
		Min	Max		Min	Max	
Australia	92	63	100	83	43	100	30
Colombia	94	68	100	82	40	100	18
Czech Republic	96	78	100	91	70	100	27
Hong Kong	89	56	100	80	44	100	9
Netherlands	82	52	100	71	22	100	23
Norway	88	67	100	81	40	100	15
Portugal	100	91	100	96	73	100	12
Scotland	79	46	100	70	27	100	12
Singapore	97	76	100	94	68	100	25
Spain	93	68	100	88	52	100	24
Switzerland	96	77	100	92	77	100	24
United States	85	62	100	74	46	100	59
AVERAGE	91	67	100	84	50	100	23

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

* Eighth grade in most countries; see Table 2 for information about the grades tested in each country.

▼ Number of student responses per item in reliability sample averaged over all items.

Note: Reliablity data based on 64 scored item parts. Percent agreement was computed separately for each part, and each part was treated as a separate item in computing averages and ranges. Reliability data are not available for one item (Magnets, Item 2).

Reliability data are not available for the following countries: Canada, Cyprus, England, Iran, Israel, New Zealand, Romania, Slovenia, and Sweden.

Because results are rounded to the nearest whole number, some totals may appear inconsistent.

TIMSS Inter-Coder Agreement for Performance Assessment – Fourth Grade* Table A.12

Country	Correctness Score Agreement			Diagnostic Code Agreement			Average Number of Student Responses per Item in the Reliability Sample▼
	Average Percent of Exact Agreement Across Tasks	Range of Percent of Exact Agreement		Average Percent of Exact Agreement Across Tasks	Range of Percent of Exact Agreement		
		Min	Max		Min	Max	
Australia	91	69	100	80	41	100	30
Hong Kong	93	75	100	86	56	100	16
Portugal	99	89	100	97	83	100	18
United States	89	60	100	77	41	100	67
AVERAGE	93	73	100	85	56	100	33

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

* Fourth grade in most countries; see Table 2 for information about the grades tested in each country.

▼ Number of student responses per item in reliability sample averaged over all items.

Note: Reliablity data based on 56 scored item parts. Percent agreement was computed separately for each part, and each part was treated as a separate item in computing averages and ranges. Reliability data are not available for one item (Magnets, Item 2).

Reliability data are not available for the following countries: Canada, Cyprus, Iran, Israel, New Zealand, Slovenia.

Because results are rounded to the nearest whole number, some totals may appear inconsistent.

Performance Assessment Test Reliability

Table A.13 displays a measure of the reliability of the performance assessment test as a whole for each country for the eighth and fourth grades. This coefficient is the KR-21 reliability coefficient across the items in all tasks computed from the correlation matrix based on all available data for each country. Reliabilities for the fourth grade ranged from .85 to .89 and in the eighth grade from .86 to .94. The international median, shown in the last row of the table, is the median of the reliability coefficients for all countries. These international medians are .88 for the fourth grade and .90 for the eighth grade.

Data Processing

To ensure the availability of comparable, high-quality data for analysis, TIMSS carried out a set of rigorous quality control steps to create the international database.[11] TIMSS prepared manuals and software for countries to use in entering their data so that the information would be in a standardized international format before being forwarded to the IEA Data Processing Center in Hamburg for creation of the international database. Upon arrival at the IEA Data Processing Center, the data from each country underwent an exhaustive cleaning process. The data cleaning process involved several iterative steps and procedures designed to identify, document, and correct deviations from the international instruments, file structures, and coding schemes. This process also emphasized consistency of information within national data sets and appropriate linking among the many student, teacher, and school data files.

Throughout the process, the data were checked and double-checked by the IEA Data Processing Center, the TIMSS International Study Center, and the national centers. The national centers were contacted regularly and given multiple opportunities to review the data for their countries. In conjunction with the Australian Council for Educational Research (ACER), the TIMSS International Study Center reviewed item statistics for each performance assessment item in each country to identify poorly performing items. Usually the poor statistics were a result of translation, adaptation, or printing deviations.

[11] These steps are detailed in Jungclaus, H. and Bruneforth, M. (1996). "Data Consistency Checking Across Countries" in M.O. Martin and D.L. Kelly, (Eds.), *Third International Mathematics and Science Study Technical Report, Volume I.* Chestnut Hill, MA: Boston College.

Reliability Coefficients[1] for the TIMSS Performance Assessment Eighth and Fourth Grades*

Table A.13

Country	Eighth Grade	Fourth Grade
Australia	0.90	0.87
Canada	0.89	0.85
Colombia	0.89	.
Cyprus	0.92	0.88
Czech Republic	0.93	.
England	0.93	.
Hong Kong	-	0.87
Iran, Islamic Rep.	0.89	0.89
Israel	-	-
Netherlands	0.88	.
New Zealand	0.90	0.85
Norway	0.89	.
Portugal	0.89	0.88
Romania	0.89	.
Scotland	0.94	.
Singapore	0.91	.
Slovenia	0.86	0.88
Spain	0.89	.
Sweden	0.92	.
Switzerland	0.90	.
United States	0.92	0.88
International Median	0.90	0.88

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

* Eighth and fourth grades in most countries; see Table 2 for information about the grades tested in each country.

[1] The reliability coefficient for each country is the KR-21 reliability coefficient across the tasks computed from the correlation matrix based on all available data for the country.

A dash (-) indicates data are not available.

A dot (.) indicates country did not participate at the fourth grade.

Data Analysis

The analytic approach underlying the majority of the results presented in this report involved calculating the average percentage score on each item within each task. The percentage score on an item is the score achieved by a student expressed as a percentage of the maximum points available on that item. The average percentage score shown for each item in Chapter 1 is this score averaged over the students in each country.

The overall task averages for each country shown in Chapter 1 (also shown in Tables 2.1 and 2.2) were obtained by averaging that country's average percentage scores across all items in a task, with each item being weighted equally. The unweighted average of items within a task was chosen to equalize the contribution of each item, since the scoring scheme for each item was developed independently, and the maximum point values were not required to be comparable across items. The overall averages for each country shown in Chapter 2 (also shown in Tables 3.1 to 3.4) reflect that country's task-level average percentage scores averaged across all tasks, with each task weighted equally. The international averages shown in all tables in Chapters 1 and 2 are the unweighted averages of the country-level average percentage scores.

Two different methods of analysis were used for the results shown in Chapter 3 on performance expectations. The average percentage scores by performance expectation categories in Tables 3.1 to 3.4 were computed by the method described in the previous paragraph. In these tables, however, average percentage scores for subsets of items were computed based on their assignment to performance expectation categories. The average of percentage scores across all tasks (shown in the first column) are the same as the overall averages computed in Chapter 2. Again, the international averages reflect the unweighted average of the country-level average percentage scores for each category.

Results shown in Figures 3.2 to 3.6 are based on calculating the percentage of students internationally obtaining full credit (maximum points) and also the percentage obtaining partial credit (one point on a two-point item; one or two points on a three-point item) on each example item.

Estimating Sampling Error

Because the statistics presented in this report are estimates of national performance based on samples of students, rather than the values that could be calculated if every student in every country had answered every question, it is important to have measures of the degree of uncertainty of the estimates. The jackknife procedure was used to estimate the standard error associated with each statistic presented in this report. The use of confidence intervals, based on the standard errors, provides a way to make inferences about the population means and proportions in a manner that reflects the uncertainty associated with the sample estimates. An estimated sample statistic plus or minus two standard errors represents a 95% confidence interval for the corresponding population result.