Appendix A

Overview of TIMSS Prodecures: Mathematics Achievement Results for Third- and Fourth-Grade Students

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, the third studies were conducted together as an integrated effort.[1]

The number of participating countries, the number of grades tested, and testing both mathematics and science 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, the magnitude of TIMSS expanded beyond the already substantial task of measuring achievement in two subject areas to also include a thorough investigation of curriculum and how it is delivered in classrooms around the world.

[1] Because a substantial amount of time has elapsed since earlier IEA studies in mathematics and science, curriculum and testing methods in these two subjects have undergone many changes. Because TIMSS has devoted considerable energy towards reflecting the most current educational and measurement practices, changes in items and methods as well as differences in the populations tested make comparisons of TIMSS results with those of previous studies very difficult. For example, SIMS did not include students at the lower grade levels. The focus of TIMSS is not on measuring achievement trends, but rather on providing up-to-date information about the current quality of education in mathematics and science.

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 meaningful interpretation and comparison of results, TIMSS also collected extensive information about the social and cultural contexts for learning, many of which are related to variations among different educational systems.

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.[2] Initial results from this component of TIMSS can be found in two companion volumes: *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intention in School Mathematics* and *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Science*.[3]

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:

1) Students taking advanced courses in mathematics, and
2) Students taking courses in physics.

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

[3] 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.

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. In about half of the countries at Populations 1 and 2, subsets of the upper-grade students who completed the written tests also participated in a performance assessment. In the performance assessment, students engaged in a number of hands-on mathematics and science activities. The students designed experiments, tested hypotheses, and recorded their findings. For example, in one task, students were asked to investigate probability by repeatedly rolling a die, applying a computational algorithm, and proposing explanations in terms of probability for patterns that emerged. Figure A.1 shows the countries that participated in the various components of TIMSS achievement testing.

TIMSS also administered a broad array of questionnaires to collect data about how the curriculum is implemented in classrooms, including the instructional practices used to deliver it. The questionnaires also were used to collect information about the social and cultural contexts for learning. Questionnaires were administered at the country level about decision-making and organizational features within their 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 teacher support. In addition, a volume was compiled that presents descriptions of the educational systems of the participating countries.[4]

With its enormous array of data, TIMSS has numerous possibilities for policy-related research, focused studies related to students' understandings of mathematics and science subtopics and processes, and integrated analyses linking the various components of TIMSS. The initial round of reports is only the beginning of a number of research efforts and publications aimed at increasing our understanding of how mathematics and science education functions across countries, investigating what impacts student performance, and helping to improve mathematics and science education.

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

Figure A.1

Countries Participating in Components of TIMSS Testing

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 Mathematics Test

The TIMSS curriculum framework underlying the mathematics tests at all three populations was developed by groups of mathematics educators with input from the TIMSS National Research Coordinators (NRCs). As shown in Figure A.2, the mathematics curriculum framework contains three dimensions or aspects. The content aspect represents the subject matter content of school mathematics. The performance expectations aspect describes, in a non-hierarchical way, the many kinds of performances or behaviors that might be expected of students in school mathematics. The perspectives aspect focuses on the development of students' attitudes, interest, and motivations in mathematics.[5]

Working within the mathematics curriculum framework, mathematics test specifications were developed for each population that included items representing a wide range of mathematics topics and eliciting a range of skills from the students. The tests were developed through an international consensus involving input from experts in mathematics and measurement specialists. The TIMSS Subject Matter Advisory Committee, including distinguished scholars from 10 countries, ensured that the test reflected current thinking and priorities within the field of mathematics. The items underwent an iterative development and review process, with multiple pilot testing efforts. Every effort was made to help ensure that the tests represented the curricula of the participating countries and that the items did not exhibit any bias towards or against particular countries, including modifying specifications in accordance with data from the curriculum analysis component, obtaining ratings of the items by subject matter specialists within the participating countries, and conducting thorough statistical item analysis of data collected in the pilot testing. The final forms of the test were endorsed by the NRCs of the participating countries.[6] In addition, countries had an opportunity to match the content of the test to their curricula at the third and fourth grades. They identified items measuring topics not covered in their intended curriculum. The information from this Test-Curriculum Matching Analysis indicates that omitting such items has little effect on the overall pattern of results (see Appendix B).

Table A.1 presents the six content areas included in the Population 1 mathematics test and the numbers of items and score points in each category. Distributions also are included for the four performance categories derived from the performance expectations aspect of the curriculum framework. Approximately one-fourth of the items were in the free-response format, requiring students to generate and write their own answers. Designed to represent approximately one-third of students' response time, some free-response questions asked for short answers while others required

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

[6] For a full discussion of the TIMSS test development effort, please see: Garden, R.A. and Orpwood, G. (1996). "TIMSS Test Development" in M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report, Volume I*. Chestnut Hill, MA: Boston College; and Garden, R.A.(1996). *No.2: Research Questions and Study Design*. Vancouver, B.C.: Pacific Educational Press.

extended responses where students needed to show their work. The remaining questions used a multiple-choice format. In scoring the tests, correct answers to most questions were worth one point. Consistent with the approach of allotting students longer response time for the constructed-response questions than for multiple-choice questions, however, responses to some of these questions (particularly those requiring extended responses) were evaluated for partial credit, with a fully correct answer being awarded two points (see later section on scoring). This, in addition to the fact that several items had two parts, means that the total number of score points available for analysis somewhat exceeds the number of items included in the test.

The TIMSS instruments were prepared in English and translated into the additional languages used for testing. In addition, it sometimes was necessary to adapt the international versions for cultural purposes, including 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 using two or more independent translations, 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) corrections by the national centers in accordance with the suggestions made, 6) verification that corrections were implemented, and 7) a series of statistical checks after the testing to detect items that did not perform comparably across countries.[7]

[7] 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; and 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.

Figure A.2

The Three Aspects and Major Categories of the Mathematics Framework

Content

- Numbers
- Measurement
- Geometry
- Proportionality
- Functions, relations, and equations
- Data representation, probability, and statistics
- Elementary analysis
- Validation and structure

Performance Expectations

- Knowing
- Using routine procedures
- Investigating and problem solving
- Mathematical reasoning
- Communicating

Perspectives

- Attitudes
- Careers
- Participation
- Increasing interest
- Habits of mind

Table A.1

Distribution of Mathematics Items by Content Reporting Category and Performance Expectation - Population 1

Content Category	Percentage of Items	Number of Items	Number of Multiple-Choice Items	Number of Short-Answer Items	Number of Extended-Response Items	Number of Score Points[1]
Whole Numbers	25%	25	19	5	1	27
Fractions and Proportionality	21%	21	15	2	4	26
Measurement, Estimation, and Number Sense	20%	20	16	3	1	21
Data Representation, Analysis, and Probability	12%	12	8	2	2	15
Geometry	14%	14	12	2	0	14
Patterns, Relations, and Functions	10%	10	9	1	0	10
Total	102%	102	79	15	8	113

Performance Expectation	Percentage of Items	Number of Items	Number of Multiple-Choice Items	Number of Short-Answer Items	Number of Extended-Response Items	Number of Score Points[1]
Knowing	41%	42	35	7	0	38
Performing Routine Procedures	16%	16	13	3	0	16
Using Complex Procedures	24%	24	21	2	1	25
Solving Problems[2]	20%	20	10	3	7	34

[1] In scoring the tests correct answers to most items were worth one point. However, responses to some constructed-response items were evaluated for partial credit with a fully correct answer awarded up to two points. In addition, some items had two parts. Thus, the number of score points exceeds the number of items in the test.

[2] Includes one extended-response item classified as "Justifying and Proving" and three extended-response items and one short-answer item classified as "Communicating."

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

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

TIMSS Test Design

Not all of the students in Population 1 responded to all of the mathematics items. To ensure broad subject matter coverage without overburdening individual students, TIMSS used a rotated design that included both the mathematics and science items. Thus, the same students participated in both the mathematics and science testing. The TIMSS Population 1 test consisted of eight booklets, with each booklet requiring 64 minutes of student response time. The booklets were designed to be administered in two consecutive testing sessions with a 15- to 20-minute break in between. Students took four clusters of items (37 minutes) prior to the break and three clusters of items (27 minutes) after the break. In accordance with the design, the mathematics and science items were assembled into 26 different clusters (labeled A through Z). Cluster A was designed to take students 10 minutes to complete and the remaining clusters were designed to take 9 minutes each. In all, the design provided a total of 235 unique testing minutes, 118 for mathematics and 117 for science. Cluster A was a core cluster assigned to all booklets. The remaining clusters were assigned to the booklets in accordance with the rotated design so that representative samples of students responded to each cluster.[8]

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, 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 International Study Center in consultation with Statistics Canada, the sampling referee, and the Technical Advisory Committee to evaluate the quality of the samples.

In a few situations where it was not possible to implement TIMSS testing for all of Population 1, as specified by the international desired definition (all students in the two adjacent grades with the greatest proportion of 9-year-olds), countries were permitted to define a national desired population that did not include part of the international desired population. Table A.2 shows any differences in coverage between the international and national desired populations. Most participants achieved 100% coverage (24 out of 26). The countries with less than 100% coverage are annotated

[8] The design is fully documented in Adams, R. and Gonzalez, E. (1996). "Design of the TIMSS Achievement Instruments" in D.F. Robitaille and R.A. Garden (Eds.), *TIMSS Monograph No. 2: Research Questions and Study Design*. Vancouver, B.C.: Pacific Education Press; and Adams, R. and Gonzalez, E. (1996). "TIMSS Test 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.

in tables in this report. Israel and Latvia, as a matter of practicality, needed to define their tested populations according to the structure of their school systems. Because coverage fell below 65% for Latvia, the Latvian results have been labeled "Latvia (LSS)," for Latvian Speaking Schools, throughout the report.

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). Table A.2 also shows that the degree of such exclusions was small. Only England exceeded the 10% limit, and this is annotated in the tables in this report. This primarily was because schools which were taking part in trials for National Curriculum Assessment (5.8% of students) were excluded.

Countries were required to test the two adjacent grades with the greatest proportion of 9-year-olds. Table A.3 presents, for each country, the percentage of 9-year-olds in the lower grade tested, the percentage in the upper grade, and the percentage in the upper and lower grades combined.

Within countries, TIMSS used a two-stage sample design at Population 1, where the first stage involved selecting 150 public and private schools within each country. Within each school, the basic approach required countries to use random procedures to select one mathematics class at the fourth grade and one at the third grade (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 7,500 students per country, with approximately 3,750 students at each grade.[9] Typically, between 450 and 3,750 students responded to each item at each grade level, depending on the booklets in which the items were located.

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.4 through A.8 present the participation rates and achieved sample sizes for the fourth and third grades.

[9] 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

The International Desired Population is defined as follows:
Population 1 - All students enrolled in the two adjacent grades with the largest proportion of 9-year-old students at the time of testing.

Country	International Desired Population		National Desired Population		
	Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
Australia	100%		0.1%	1.6%	1.8%
Austria	100%		2.6%	0.2%	2.8%
Canada	100%		2.5%	3.6%	6.2%
Cyprus	100%		3.1%	0.1%	3.2%
Czech Republic	100%		4.1%	0.0%	4.1%
[2] England	100%		8.6%	3.5%	12.1%
Greece	100%		1.5%	4.0%	5.4%
Hong Kong	100%		2.6%	0.0%	2.7%
Hungary	100%		3.8%	0.0%	3.8%
Iceland	100%		1.9%	4.3%	6.2%
Iran, Islamic Rep.	100%		0.3%	1.0%	1.3%
Ireland	100%		5.3%	1.6%	6.9%
[1] Israel	72%	Hebrew Public Education System	1.1%	0.1%	1.2%
Japan	100%		3.0%	0.0%	3.0%
Korea	100%		3.9%	2.6%	6.6%
Kuwait	100%		0.0%	0.0%	0.0%
[1] Latvia (LSS)	60%	Latvian-speaking schools	2.1%	0.0%	2.1%
Netherlands	100%		4.0%	0.4%	4.4%
New Zealand	100%		0.7%	0.6%	1.3%
Norway	100%		1.1%	2.0%	3.1%
Portugal	100%		6.6%	0.7%	7.3%
Scotland	100%		2.4%	4.3%	6.7%
Singapore	100%		0.0%	0.0%	0.0%
Slovenia	100%		1.9%	0.0%	1.9%
Thailand	100%		6.8%	1.5%	8.3%
United States	100%		0.4%	4.3%	4.7%

[1] National Desired Population does not cover all of International Desired Population. Because coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.
[2] National Defined Population covers less than 90 percent of National Desired Population.

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

Table A.3

Coverage of 9-Year-Old Students

Country	Percent of 9-Year-Olds in Lower Grade (Third Grade*)	Percent of 9-Year-Olds in Upper Grade (Fourth Grade*)	Percent of 9-Year-Olds in Both Grades
Australia	65	29	94
Austria	72	15	87
Canada	46	48	94
Cyprus	35	63	98
Czech Republic	75	15	91
England	58	41	99
Greece	11	88	99
Hong Kong	43	50	93
Hungary	70	19	89
Iceland	15	84	99
Iran, Islamic Rep.	51	32	83
Ireland	68	23	92
Israel	-	-	-
Japan	91	9	99
Korea	67	24	91
Kuwait	-	-	-
Latvia (LSS)	55	21	76
Netherlands	63	30	93
New Zealand	50	49	99
Norway	38	62	100
Portugal	45	48	93
Scotland	23	76	99
Singapore	80	17	98
Slovenia	60	0	60
Thailand	60	11	71
United States	61	34	95

*Third and fourth grades in most countries; see Table 2 for more information about the grades tested in each country.
A dash (–) indicates data are unavailable. Israel and Kuwait did not test the lower grade.
Because results are rounded to the nearest whole number some totals may appear inconsistent.

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

Table A.4

School Participation Rates and Sample Sizes

Upper Grade (Fourth Grade*)

Country	School Participation Before Replacement (Weighted Percentage)	School Participation After Replacement (Weighted Percentage)	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample That Participated	Number of Replacement Schools That Participated[1]		Total Number of Schools That Participated
						Procedural	Other	
Australia	66	69	268	268	169	9	0	178
Austria	51	72	150	150	71	31	31	133
Canada	90	90	423	420	390	0	0	390
Cyprus	97	97	150	150	146	0	0	146
Czech Republic	91	94	215	215	181	7	0	188
England	63	88	150	145	92	35	0	127
Greece	93	93	187	187	174	0	0	174
Hong Kong	84	84	156	148	124	0	0	124
Hungary	100	100	150	150	150	0	0	150
Iceland	95	95	153	151	144	0	0	144
Iran, Islamic Rep.	100	100	180	180	180	0	0	180
Ireland	94	96	175	173	161	4	0	165
Israel	40	40	100	100	40	0	47	87
Japan	93	96	150	150	137	4	0	141
Korea	100	100	150	150	150	0	0	150
Kuwait	100	100	150	150	150	0	0	150
Latvia (LSS)	74	74	169	169	125	0	0	125
Netherlands	31	62	196	196	63	67	0	130
New Zealand	80	99	150	150	120	29	0	149
Norway	85	94	150	148	126	13	0	139
Portugal	95	95	150	150	143	0	0	143
Scotland	78	83	184	184	143	9	0	152
Singapore	100	100	191	191	191	0	0	191
Slovenia	81	81	150	150	121	0	0	121
Thailand	96	96	155	155	154	0	0	154
United States	85	85	220	213	182	0	0	182

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

[1] Replacement schools selected in accordance with the TIMSS sampling procedures are listed in the "procedural" column. Those selected using unapproved methods are listed in the "other" column and were not included in the computation of school participation rates.

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

Table A.5

Student Participation Rates and Sample Sizes
Upper Grade (Fourth Grade*)

Country	Within School Student Participation (Weighted Percentage)	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
Australia	96	6930	37	104	6789	282	6507
Austria	96	2779	12	6	2761	116	2645
Canada	96	9193	81	268	8844	436	8408
Cyprus	86	3972	4	3	3965	589	3376
Czech Republic	92	3555	7	0	3548	280	3268
England	95	3489	73	122	3294	168	3126
Greece	95	3358	6	116	3236	183	3053
Hong Kong	98	4475	0	1	4474	63	4411
Hungary	92	3272	0	0	3272	266	3006
Iceland	90	2149	23	101	2025	216	1809
Iran, Islamic Rep.	97	3521	5	36	3480	95	3385
Ireland	93	3134	14	40	3080	207	2873
Israel	94	2486	0	3	2483	132	2351
Japan	97	4453	0	0	4453	147	4306
Korea	95	2971	133	0	2838	26	2812
Kuwait	95	4578	34	0	4544	226	4318
Latvia (LSS)	93	2390	12	1	2377	161	2216
Netherlands	96	2639	0	4	2635	111	2524
New Zealand	96	2627	82	20	2525	104	2421
Norway	97	2391	16	42	2333	76	2257
Portugal	96	2994	15	16	2963	110	2853
Scotland	92	3735	0	139	3596	295	3301
Singapore	98	7274	14	0	7260	121	7139
Slovenia	94	2720	3	0	2717	151	2566
Thailand	100	3042	0	50	2992	0	2992
United States	94	8224	61	412	7751	455	7296

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

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

Table A.6

School Participation Rates and Sample Sizes

Lower Grade (Third Grade*)

Country	School Participation Before Replacement (Weighted Percentage)	School Participation After Replacement (Weighted Percentage)	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample That Participated	Number of Replacement Schools That Participated		Total Number of Schools That Participated
						Proce-dural	Other	
Australia	66	69	268	264	166	9	0	175
Austria	49	70	150	149	68	29	31	128
Canada	88	88	423	418	375	0	0	375
Cyprus	98	98	150	150	147	0	0	147
Czech Republic	91	93	215	215	180	7	0	187
England	64	88	150	145	93	35	0	128
Greece	91	91	187	187	171	0	0	171
Hong Kong	84	84	156	147	123	0	0	123
Hungary	99	99	150	150	149	0	0	149
Iceland	95	95	153	152	144	0	0	144
Iran, Islamic Rep.	99	99	180	180	178	0	0	178
Ireland	94	96	175	173	160	4	0	164
Israel	-	-	-	-	-	-	-	-
Japan	93	95	150	150	137	5	0	142
Korea	100	100	150	150	150	0	0	150
Kuwait	-	-	-	-	-	-	-	-
Latvia (LSS)	73	73	169	168	123	0	0	123
Netherlands	29	62	196	195	60	69	0	129
New Zealand	80	99	150	150	120	29	0	149
Norway	83	92	150	148	124	12	0	136
Portugal	95	95	150	150	143	0	0	143
Scotland	77	81	184	184	142	8	0	150
Singapore	100	100	191	191	191	0	0	191
Slovenia	81	81	150	149	122	0	0	122
Thailand	96	96	155	154	153	0	0	153
United States	86	86	220	217	186	0	0	186

*Third grade in most countries; see Table 2 for more information about the grades tested in each country.
A dash (–) indicates data are unavailable. Israel and Kuwait did not test the lower grade.
[1] Replacement schools selected in accordance with the TIMSS sampling procedures are listed in the "procedural" column. Those selected using unapproved methods are listed in the "other" column and were not included in the computation of school participation rates.

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

Table A.7

Student Participation Rates and Sample Sizes
Lower Grade (Third Grade*)

Country	Within School Student Participation (Weighted Percentage)	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
Australia	95	5138	31	92	5015	274	4741
Austria	96	2655	10	6	2639	113	2526
Canada	96	8433	77	307	8049	455	7594
Cyprus	85	3913	5	2	3906	598	3308
Czech Republic	93	3484	8	0	3476	220	3256
England	94	3468	70	158	3240	184	3056
Greece	94	3263	4	133	3126	171	2955
Hong Kong	99	4455	0	2	4453	57	4396
Hungary	94	3270	0	0	3270	232	3038
Iceland	91	2017	19	89	1909	211	1698
Iran, Islamic Rep.	98	3504	12	49	3443	82	3361
Ireland	94	3127	14	39	3074	185	2889
Israel	-	-	-	-	-	-	-
Japan	97	4433	0	0	4433	127	4306
Korea	94	2969	138	2	2829	52	2777
Kuwait	-	-	-	-	-	-	-
Latvia (LSS)	94	2218	8	0	2210	156	2054
Netherlands	96	2923	0	14	2909	119	2790
New Zealand	95	2733	91	9	2633	129	2504
Norway	97	2362	8	59	2295	76	2219
Portugal	97	2790	13	31	2746	96	2650
Scotland	90	3663	0	187	3476	344	3132
Singapore	98	7223	14	0	7209	179	7030
Slovenia	95	2659	5	0	2654	133	2521
Thailand	100	2945	0	74	2871	1	2870
United States	95	4280	40	201	4039	220	3819

*Third grade in most countries; see Table 2 for more information about the grades tested in each country.
A dash (–) indicates data are unavailable. Israel and Kuwait did not test the lower grade.

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

Table A.8

Overall Participation Rates

Lower and Upper Grades (Third and Fourth Grades*)

Country	Upper Grade		Lower Grade	
	Overall Participation Before Replacement (Weighted Percentage)	Overall Participation After Replacement (Weighted Percentage)	Overall Participation Before Replacement (Weighted Percentage)	Overall Participation After Replacement (Weighted Percentage)
Australia	63	66	62	65
Austria	49	69	46	67
Canada	86	86	84	84
Cyprus	83	83	83	83
Czech Republic	84	86	85	87
England	60	83	61	83
Greece	88	88	86	86
Hong Kong	83	83	83	83
Hungary	92	92	93	93
Iceland	86	86	86	86
Iran, Islamic Rep.	97	97	97	97
Ireland	88	90	88	91
Israel	38	38	-	-
Japan	90	92	90	93
Korea	95	95	94	94
Kuwait	95	95	-	-
Latvia (LSS)	69	69	69	69
Netherlands	29	59	28	60
New Zealand	77	95	76	95
Norway	82	91	81	89
Portugal	92	92	92	92
Scotland	71	76	69	73
Singapore	98	98	98	98
Slovenia	76	76	77	77
Thailand	96	96	96	96
United States	80	80	81	81

*Third and Fourth grades in most countries; see Table 2 for information about the grades tested in each country.
A dash (–) indicates data are unavailable. Israel and Kuwait did not test the lower grade.

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

Indicating Compliance with Sampling Guidelines in the Report

Figure A.3 shows how countries have been grouped in tables reporting achievement results. Countries that complied with the TIMSS guidelines for grade selection and classroom sampling, and that achieved acceptable participation rates, are shown in both the schools and students or a combined rate (the product of school and student participation) of 75% with or without replacement schools. Countries that met the guidelines only after including replacement schools are annotated. These countries (17 at the fourth grade and 16 at the third grade) appear in the tables in Chapters 1, 2, and 3 ordered by achievement.

Countries that did not reach at least 50% school participation without the use of replacements 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.3. These countries are presented in a separate section of the achievement tables in Chapters 1, 2, and 3 in alphabetical order, and are shown in tables in Chapters 4 and 5 in italics.

To provide a better curricular match, Slovenia elected to test its third- and fourth-grade students even though that meant not testing the two grades with the most 9-year-olds and resulted in its students being somewhat older than those in the other countries. Slovenia is also presented in a separate section of the achievement tables in Chapters 1, 2, and 3 and is shown in tables in Chapters 4 and 5 in italics. Table A.3 shows the percentage of 9-year-olds for each country in the grades tested.

Hungary did not completely comply with the guidelines for sampling classrooms at the fourth grade and thus its results are also presented in a separate section of the achievement tables in Chapters 1, 2, and 3 in alphabetical order, and are italicized in tables in Chapters 4 and 5. At the fourth grade, Israel, Kuwait, and Thailand also had difficulty complying with the classroom selection guidelines, but in addition had other difficulties (Kuwait tested a single grade with relatively few 9-year-olds; Israel had low sampling participation rates; Thailand had a high percentage of older students), and so these countries are also presented in separate sections in tables in Chapters 1, 2, and 3, and are italicized in tables in Chapters 4 and 5. Israel and Kuwait did not test at the lower grade.

Figure A.3

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

	Fourth Grade	Third Grade
Countries satisfying guidelines for sample participation rates, grade selection and sampling procedures	Canada Cyprus Czech Republic †2 England Greece Hong Kong Iceland Iran, Islamic Rep. Ireland Japan Korea New Zealand Norway Portugal † Scotland Singapore United States	Canada Cyprus Czech Republic †2 England Greece Hong Kong Iceland Iran, Islamic Rep. Ireland Japan Korea New Zealand Norway Portugal Singapore United States
Countries not satisfying guidelines for sample participation	Australia Austria 1 Latvia (LSS) Netherlands	Australia Austria 1 Latvia (LSS) Netherlands Scotland
Countries not meeting age/grade specifications (high percentage of older students)	Slovenia	Slovenia
Countries with unapproved sampling procedures at the classroom level	Hungary	Hungary
Countries with unapproved sampling procedures at classroom level and not meeting other guidelines	1 Israel Kuwait Thailand	Thailand

†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 1). Because coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.
2 National Defined Population covers less than 90 percent of National Desired Population (see Table 1).

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

Data Collection

Each participating country was responsible for carrying out all aspects of the data collection, using standardized procedures developed for the study. Training manuals were developed for school coordinators and test administrators that explained procedures for receipt and distribution of materials as well as for the activities related to the testing sessions. The test administrator manuals covered procedures for test security, standardized scripts to regulate directions and timing, rules for answering students' questions, and steps to ensure that identification on the test booklets and questionnaires corresponded to the information on the forms used to track students.

Each country was responsible for conducting quality control procedures and describing this effort as part of the NRC's report documenting procedures used in the study. In addition, the International Study Center considered it essential to establish some method to monitor compliance with standardized procedures. NRCs were each asked to nominate a person, such as a retired school teacher, to serve as the quality control monitor for his or her own country, and in almost all cases, the International Study Center adopted the NRC's first suggestion. The International Study Center developed manuals for the quality control monitors and briefed them in two-day training sessions about TIMSS, the responsibilities of the national centers in conducting the study, and their own roles and responsibilities.

The quality control monitors interviewed the NRCs about data collection plans and procedures. They also selected a sample of approximately 10 schools to visit, where they observed testing sessions and interviewed school coordinators.[10] Quality control monitors observed test administrations and interviewed school coordinators in 37 countries, and interviewed school coordinators or test administrators in 3 additional countries.

The results of the interviews indicate that, in general, NRCs had prepared well for data collection and, despite the heavy demands of the schedule and shortages of resources, were in a position to conduct the data collection in an efficient and professional manner. Similarly, the TIMSS tests appeared to have been administered in compliance with international procedures, including the activities preliminary to the testing session, the activities during the testing sessions, and the school-level activities related to receiving, distributing, and returning materials from the national centers.

[10] The results of the interviews and observations by the quality control monitors are presented in Martin, M.O., Hoyle, C.D., and Gregory, K.D. (1996). "Monitoring the TIMSS Data Collection" and "Observing the TIMSS Test Administration," both 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.

Scoring the Free-Response Items

Because approximately one-third of the written test time was devoted to free-response items, TIMSS needed to develop procedures for reliably evaluating student responses within and across countries. Scoring utilized two-digit codes with rubrics specific to each item. Development of the rubrics was led by the Norwegian TIMSS national center. The first digit designates the correctness level of the response. The second digit, combined with the first digit, represents a diagnostic code used to identify specific types of approaches, strategies, or common errors and misconceptions. Although not specifically used in this report, analyses of responses based on the second digit should provide insight into ways to help students better understand mathematics concepts and problem-solving approaches.

To meet the goal of implementing reliable scoring procedures based on the TIMSS rubrics, the International Study Center prepared guides containing the rubrics and explanations of how to implement them, together with example student responses for the various rubric categories. These guides, together with more examples of student responses for practice in applying the rubrics, were used as a basis for an ambitious series of regional training sessions. The training sessions 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.[11]

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 were to be coded independently by two different readers. Table A.9 shows the average and range of the within-country percentage of exact agreement between scorers on the free-response items in the Population 1 mathematics test for 16 countries. Unfortunately, lack of resources precluded several countries from providing this information. A very high percentage of exact agreement was observed, with averages across the items for the correctness score ranging from 94% to 99% and an overall average of 97% across the 16 countries.

To provide information about the cross-country agreement among scorers, TIMSS conducted a special study at Population 2, where 39 scorers from 21 of the participating countries evaluated common sets of students' responses to more than half of the free-response items. Unfortunately, resources did not allow an international reliability study to be conducted for Population 1. However, the results of the international reliability study at Population 2 demonstrated a very high percentage of exact agreement on the correctness and diagnostic scores. The TIMSS data from the reliability studies indicate that scoring procedures were extremely robust for the mathematics items, especially for the correctness score used for the analyses in this report.[12]

[11] 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.

[12] Details about the reliability studies can be found in Mullis, I.V.S., and Smith, T.A. (1996). "Quality Control Steps for Free-Response Scoring" 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.

Table A.9

TIMSS Within-Country Free-Response Coding Reliability Data for Population 1 Mathematics Items*

Country	Correctness Score Agreement			Diagnostic Code Agreement		
	Average Percent of Exact Agreement Across Items	Range of Percent of Exact Agreement		Average Percent of Exact Agreement Across Items	Range of Percent of Exact Agreement	
		Min	Max		Min	Max
Australia	96	76	100	90	69	100
Canada	95	81	100	90	67	99
Czech Republic	98	93	100	95	85	100
England	99	93	100	97	87	100
Hong Kong	96	85	99	91	73	98
Ireland	98	89	100	94	85	99
Iran, Islamic Rep.	94	84	99	88	74	96
Israel	96	86	100	92	65	100
Japan	99	98	100	99	96	100
Netherlands	96	84	100	92	78	100
Norway	99	95	100	96	80	100
New Zealand	99	96	100	96	88	100
Portugal	97	89	99	95	82	98
Scotland	94	79	99	86	62	97
Singapore	98	90	100	96	89	100
United States	99	93	100	96	81	100
AVERAGE	97	88	100	93	79	99

*Based on 23 mathematics items, including 4 multiple-part items.
Note: Percent agreement was computed separately for each part, and each part was treated as a separate item in computing averages and ranges.

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

Test Reliability

Table A.10 displays the mathematics test reliability coefficient for each country for the lower and upper grades (usually third and fourth grades). This coefficient is the median KR-20 reliability across the eight test booklets. Median reliabilities in the lower grade ranged from .72 to .87, and in the upper grade from .74 to .88. 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 .82 for the lower grade and .84 for the upper grade.

Data Processing

To ensure the availability of comparable, high-quality data for analysis, TIMSS engaged in a rigorous set of quality control steps to create the international database.[13] 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 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 International Study Center conducted a review of item statistics for each of the cognitive items in each of the countries to identify poorly performing items. Six countries had one or more mathematics items deleted (in most cases, one). Usually the poor statistics (negative point-biserials for the key, large item-by-country interactions, and statistics indicating lack of fit with the model) were a result of translation, adaptation, or printing deviations.

[13] 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.

Table A.10

Cronbach's Alpha Reliability Coefficients[1]
TIMSS Mathematics Test
Lower and Upper Grades (Third and Fourth Grades*)

Country	Lower Grade	Upper Grade
Australia	0.85	0.86
Austria	0.79	0.79
Canada	0.82	0.85
Cyprus	0.79	0.85
Czech Republic	0.83	0.84
England	0.84	0.86
Greece	0.84	0.86
Hong Kong	0.80	0.84
Hungary	0.84	0.84
Iceland	0.73	0.83
Iran, Islamic Rep.	0.72	0.79
Ireland	0.83	0.84
Israel	-	0.83
Japan	0.82	0.82
Korea	0.79	0.82
Kuwait	-	0.74
Latvia (LSS)	0.80	0.82
Netherlands	0.76	0.79
New Zealand	0.83	0.86
Norway	0.77	0.81
Portugal	0.83	0.82
Scotland	0.81	0.86
Singapore	0.87	0.88
Slovenia	0.82	0.82
Thailand	0.81	0.81
United States	0.83	0.86
International Median	0.82	0.84

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

[1] The reliability coefficient for each country is the median KR-20 reliability across the eight test booklets. The international median is the median of the reliability coefficients for all countries.

A dash (-) indicates data are unavailable. Israel and Kuwait did not test the lower grade.

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

IRT Scaling and Data Analysis

Two general analysis approaches were used for this report – item response theory scaling methods and average percent correct technology. The overall mathematics results were summarized using an item response theory (IRT) scaling method (Rasch model). This scaling method produces a mathematics score by averaging the responses of each student to the items that student took in a way that takes into account the difficulty of each item. The methods used in TIMSS include refinements that enable reliable scores to be produced even though individual students responded to relatively small subsets of the total mathematics item pool. Analyses of the response patterns of students from participating countries indicated that, although the items in the test address a wide range of mathematical content, the performance of the students across the items was sufficiently consistent to be usefully summarized in a single mathematics score.

An IRT approach was preferred for developing comparable estimates of performance for all students, since students answered different test items depending upon which of the eight test booklets they received. The IRT analysis provides a common scale on which performance can be compared across countries. In addition to providing a basis for estimating mean achievement, scale scores permit estimates of how students within countries vary and provide information on percentiles of performance. The scale was standardized using students from both the grades tested. When all participating countries and grades are treated equally, the TIMSS scale average is 500 and the standard deviation is 100. Since the countries varied in size, each country was reweighted to contribute equally to the mean and standard deviation of the scale. The average of the scale scores was constructed to be the average of the 26 means of participants that were available at the fourth grade and the 24 means at the third grade. The average and standard deviation of the scale scores are arbitrary and do not affect scale interpretations.

The analytic approach underlying the results in Chapters 2 and 3 of this report involved calculating the percentage of correct answers for each item for each participating country (as well as the percentages of different types of incorrect responses). The percentages of correct responses were averaged to summarize mathematics performance overall and in each of the content areas for each country as a whole and by gender. For items with more than one part, each part was analyzed separately in calculating the percentage of correct responses. Also, for items with more than one point awarded for full credit, the percentage of correct responses reflects an average of the points received by students in each country. This was achieved by including the percentage of students receiving one score point as well as the percentage receiving two score points in the calculations. Thus, the average percent correct is based on the number of score points rather than the number of items, per se. An exception to this is the international average percent correct reported for example items, where the values reflect the percentage of students receiving full credit.

Estimating the Link Between Fourth- and Eighth-Grade Performance

Fifteen of the items in mathematics (15%) and 18 in science (19%) were included in the tests at both Populations 1 and 2. The difference in performance between the populations on these items was used to estimate the link between the third and fourth grades on one hand and the seventh and eighth grades on the other.

For each of the link items, the international item difficulty level from the IRT analyses for Population 1 was subtracted from the international difficulty level at Population 2. Investigations of the results indicated that the increases between the two populations were relatively stable across items, especially in mathematics. It also was determined that between-grade increases between the third and fourth grades and between the seventh and eighth grades on the link items were consistent with the between-grade increases observed on the entire pool of items for Populations 1 and 2, respectively. Thus, the average difference across items was used to estimate the difference in performance between the two populations.

In making the link, results for the third- and fourth-grade students were placed on the scale used to report seventh- and eighth-grade performance. Because of the difference in variances between the scales for Populations 1 and 2, it first was necessary to transform the Population 1 scales. The adjustment factor for mathematics was .96 and for science was 1.25. Next, a constant (121 scale points for mathematics and 283 for science) was subtracted from the Population 1 results for each country.

The country means for the third and fourth grades transformed to the seventh- and eighth-grade scale are shown in Table A.11. The results shown in Table A.11 are based on all items administered to the third and fourth graders. The relative standings of the countries are identical to those presented in Chapter 1. Since there were relatively few items in common, the size of the link is approximate. The standard errors for the third- and fourth-grade estimates incorporate an added component to account for the uncertainty of this approximation. Because the link is very approximate, the achievement increases between the third/fourth grades and the seventh/eighth grades must be interpreted with extreme caution.

Table A.11

Mathematics Performance at the Third, Fourth, Seventh, and Eighth Grades* Based on the Population 2 (Seventh- and Eighth-Grade) Scale

Country	Third Grade Mean	Fourth Grade Mean	Seventh Grade Mean	Eighth Grade Mean
Australia	347 (8.8)	408 (8.4)	498 (3.8)	530 (4.0)
Austria	351 (9.4)	421 (8.4)	509 (3.0)	539 (3.0)
Canada	334 (8.3)	395 (8.5)	494 (2.2)	527 (2.4)
Cyprus	296 (8.3)	366 (8.4)	446 (1.9)	474 (1.9)
Czech Republic	361 (8.5)	428 (8.5)	523 (4.9)	564 (4.9)
England	321 (8.4)	376 (8.5)	476 (3.7)	506 (2.6)
Greece	294 (8.8)	356 (8.9)	440 (2.8)	484 (3.1)
Hong Kong	387 (8.4)	447 (8.9)	564 (7.8)	588 (6.5)
Hungary	340 (8.9)	410 (8.7)	502 (3.7)	537 (3.2)
Iceland	276 (8.3)	338 (8.3)	459 (2.6)	487 (4.5)
Iran, Islamic Rep.	245 (8.6)	294 (8.8)	401 (2.0)	428 (2.2)
Ireland	340 (8.6)	412 (8.6)	500 (4.1)	527 (5.1)
Israel	- -	394 (8.6)	- -	522 (6.2)
Japan	400 (8.0)	457 (8.1)	571 (1.9)	605 (1.9)
Korea	422 (8.2)	471 (8.1)	577 (2.5)	607 (2.4)
Kuwait	- -	267 (8.3)	- -	392 (2.5)
Latvia (LSS)	328 (8.9)	388 (9.2)	462 (2.8)	493 (3.1)
Netherlands	357 (8.3)	438 (8.5)	516 (4.1)	541 (6.7)
New Zealand	305 (8.8)	362 (8.9)	472 (3.8)	508 (4.5)
Norway	287 (8.4)	365 (8.4)	461 (2.8)	503 (2.2)
Portugal	291 (8.7)	340 (8.6)	423 (2.2)	454 (2.5)
Scotland	323 (8.5)	383 (8.7)	463 (3.7)	498 (5.5)
Singapore	414 (9.1)	484 (9.4)	601 (6.3)	643 (4.9)
Slovenia	351 (8.4)	414 (8.5)	498 (3.0)	541 (3.1)
Thailand	309 (9.3)	354 (9.1)	495 (4.9)	522 (5.7)
United States	344 (8.5)	407 (8.4)	476 (5.5)	500 (4.6)
International Averages	334 (1.8)	391 (1.7)	493 (0.8)	520 (0.8)

*Third, fourth, seventh, and eighth grades in most countries; see Table 2 for more information about the grades tested in each country.
( ) Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.
Because population coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.
A dash (-) indicates data are unavailable. Israel and Kuwait did not test the third or seventh grades.
Note: Since there are only 15 mathematics items in common in the tests given to the two grades, the estimate of the relationship is approximate. The standard errors for the third- and fourth-grade estimates incorporate an added component to account for the uncertainty of this approximation. The seventh- and eighth-grade means are the same as those reported in *Mathematics Achievement in the Middle School Years: IEA's Third and Science Study.*

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

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.