Appendix A

TIMSS DESIGN AND PROCEDURES

TEST DESIGN

The TIMSS tests were developed through an international consensus involving input from experts in mathematics, science, and educational measurement. The TIMSS Subject Matter Advisory Committee ensured that the tests reflected current thinking and priorities within the fields of mathematics and science. Every effort was made to help ensure that the tests represented the curricula of the participating countries and that the items exhibited no bias toward or against particular countries. This involved 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 analyses of data collected in the pilot testing. The final forms of the tests were endorsed by the National Research Coordinators (NRCs) of the participating countries.

TIMSS tested primary-school (third and fourth grades) and middle-school (seventh and eighth grades) students in mathematics and science. In mathematics, the third- and fourth-grade tests included items from six content areas: whole numbers; fractions and proportionality; measurement, estimation, and number sense; data representation, analysis, and probability; geometry; and algebra. For the seventh and eighth grades, the mathematics test included items from six content areas: fractions and number sense; proportionality; measurement; data representation, analysis, and probability; geometry; and algebra. In science, the primary-school test included items from four content areas: earth science; life science; physical science; and environmental issues and the nature of science. For the seventh and eighth grades, the science test included items from five content areas: earth science; life science; chemistry; physics; and environmental issues and the nature of science.

The mathematics and science literacy test for final-year students was designed to test students' general knowledge and understanding of mathematical and scientific principles. The mathematics items cover number sense, including fractions, percentages, and proportionality. Algebraic sense, measurement, and estimation are also covered, as are data representation and analysis. Reasoning and social utility were emphasized in several items. A general criterion in selecting the items was that they should involve the types of mathematics questions that could arise in real-life situations and that they be contextualized accordingly. Similarly, the science items selected for use in the TIMSS literacy test were organized according to three areas of science – earth science, life science, and physical science – and included a reasoning and social utility component. The emphasis was on measuring how well students can use their knowledge in addressing real-world problems having a science

component. The test was designed to enable reporting for mathematics literacy and science literacy separately as well as overall.

To maximize the content coverage of the TIMSS tests, yet minimize the burden on individual students, TIMSS used a multiple matrix sampling design whereby subsets of items from the total item pool were administered to sub-samples of students.[1] Each student responded to a subset of the total item pool; by aggregating data across booklets, TIMSS was able to derive population estimates of mathematics and science achievement. TIMSS does not provide individual proficiency estimates. The design was nearly identical for the primary and middle school assessments, but different for the assessment of final-year students.

For the primary and middle school tests, items were assigned to 26 mutually exclusive groups or "clusters." The clusters were then assigned to eight test booklets so that one cluster appeared in all test booklets, some clusters appeared in several test booklets, and some clusters appeared in one test booklet. Each test booklet contained mathematics and science test items. The test booklets were systematically distributed to students and each student completed one. Primary-school students had 64 minutes to complete their test booklets, and middle-school students had 90 minutes.

For the final year of secondary-school assessment, there were nine test booklets containing the assessment material for mathematics and science literacy, advanced mathematics, and physics. Two of these booklets contained exclusively mathematics and science literacy items, and one booklet contained some mathematics and science literacy items. Students were assigned one of nine booklets depending upon their academic preparation; all students were eligible to receive the two mathematics and science literacy booklets. Final-year students had 90 minutes to complete their booklets.

In each test, approximately one-quarter of the items were in the free-response format, requiring students to generate and write their own answers. Designed to take up about one-third of students' response time, some of these questions asked for short answers while others required extended responses in which students needed to show their work. The remaining questions were in multiple-choice format. In scoring the tests, correct answers to most questions were worth one point. Consistent with the approach of allotting longer response times for constructed-response questions than for multiple-choice questions, responses to some of these questions (particularly those requiring extended responses) could earn partial credit, with a fully correct answer being awarded two or three points.

[1] The TIMSS test design is fully described in Adams, R.J. and Gonzalez, E.J. (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.

SAMPLING

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 physics.

Countries participating in the study were required to test the students in the two grades at Population 2, but could choose whether or not to participate at the other levels.

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 sampling information available and on the quality of the sampling activities themselves. For TIMSS, NRCs worked on all phases of sampling with staff from Statistics Canada. NRCs were trained 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), 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 the achievement tables presented in Chapter 1 of this report, countries are grouped according to the extent to which they met the TIMSS sampling requirements. In the remaining tables, the names of countries that did not meet the TIMSS standards for sampling are presented in italics.

COVERAGE OF THE TARGET POPULATIONS

In a few situations where it was not possible to implement TIMSS testing for the entire internationally desired population, countries were permitted to define a national desired population that did not include part of the internationally desired population. Tables A.1, A.2, and A.3 show any differences in coverage between the international and national desired populations for countries participating in each assessment. Most participants

achieved 100% coverage. In some instances, countries, as a matter of practicality, needed to define their tested population according to the structure of school systems; in others, parts of the country were simply unwilling to take part in TIMSS. 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 rural areas).

Table A.1

Coverage of TIMSS Target Population - Population 1

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%

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

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.

Table A.2

Coverage of TIMSS Target Population - Population 2

The International Desired Population is defined as follows:
Population 2 - All students enrolled in the two adjacent grades with the largest proportion of

Country	International Desired Population		National Desired Population		
	Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
Australia	100%		0.2%	0.7%	0.8%
Austria	100%		2.9%	0.2%	3.1%
Belgium (Fl)	100%		3.8%	0.0%	3.8%
Belgium (Fr)	100%		4.5%	0.0%	4.5%
Bulgaria	100%		0.6%	0.0%	0.6%
Canada	100%		2.4%	2.1%	4.5%
Colombia	100%		3.8%	0.0%	3.8%
Cyprus	100%		0.0%	0.0%	0.0%
Czech Republic	100%		4.9%	0.0%	4.9%
Denmark	100%		0.0%	0.0%	0.0%
[2] England	100%		8.4%	2.9%	11.3%
France	100%		2.0%	0.0%	2.0%
[1] Germany	88%	One region (B-W) excluded	8.8%	0.9%	9.7%
Greece	100%		1.5%	1.3%	2.8%
Hong Kong	100%		2.0%	0.0%	2.0%
Hungary	100%		3.8%	0.0%	3.8%
Iceland	100%		1.7%	2.9%	4.5%
Iran, Islamic Rep.	100%		0.3%	0.0%	0.3%
Ireland	100%		0.0%	0.4%	0.4%
[1] Israel	74%	Hebrew Public Education System	3.1%	0.0%	3.1%
Japan	100%		0.6%	0.0%	0.6%
Korea	100%		2.2%	1.6%	3.8%
Kuwait	100%		0.0%	0.0%	0.0%
[1] Latvia (LSS)	51%	Latvian-speaking schools	2.9%	0.0%	2.9%
[1] Lithuania	84%	Lithuanian-speaking schools	6.6%	0.0%	6.6%
Netherlands	100%		1.2%	0.0%	1.2%
New Zealand	100%		1.3%	0.4%	1.7%
Norway	100%		0.3%	1.9%	2.2%
Portugal	100%		0.0%	0.3%	0.3%
Romania	100%		2.8%	0.0%	2.8%
Russian Federation	100%		6.1%	0.2%	6.3%
Scotland	100%		0.3%	1.9%	2.2%
Singapore	100%		4.6%	0.0%	4.6%
Slovak Republic	100%		7.4%	0.1%	7.4%
Slovenia	100%		2.4%	0.2%	2.6%
South Africa	100%		9.6%	0.0%	9.6%
Spain	100%		6.0%	2.7%	8.7%
Sweden	100%		0.0%	0.9%	0.9%
[1] Switzerland	86%	22 of 26 cantons	4.4%	0.8%	5.3%
Thailand	100%		6.2%	0.0%	6.2%
United States	100%		0.4%	1.7%	2.1%

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

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.

*   One region (Baden-Wuerttemberg) did not participate.

Table A.3

Coverage of TIMSS Target Population - Population 3

The International Desired Population is defined as follows:
Population 3 - All students in final year of secondary school*

Country	International Desired Population		National Desired Population	
	Coverage	Notes on Coverage	Sample Exclusions	Notes on Exclusions
Australia	100%		5.5%	
Austria	100%		18.2%	Colleges and courses lasting less than 3 years excluded
Canada	100%		8.9%	
Cyprus	100%		22.0%	Private and vocational schools excluded
Czech Republic	100%		6.0%	
Denmark	100%		2.3%	
France	100%		1.0%	
Germany	100%		11.3%	
Hungary	100%		0.2%	
Iceland	100%		0.1%	
Italy	70%	Four regions did not participate	0.9%	
Lithuania	84%	Lithuanian speaking students	0.0%	
Netherlands	100%		21.6%	Apprenticeship programs excluded
New Zealand	100%		0.0%	
Norway	100%		3.8%	
Russian Federation	100%		43.0%	Vocational schools and non-Russian speaking students excluded
Slovenia	100%		6.0%	
South Africa	100%		0.0%	
Sweden	100%		0.2%	
Switzerland	100%		2.5%	
United States	100%		3.7%	

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

* See Table 1.2 for more information about the grades tested in each country.

TIMSS Coverage Index for Final-Year Assessment

A further difficulty in defining the desired population for the final-year assessment is that many students drop out before the final year of any track. Thus a TIMSS Coverage Index (TCI) was calculated that quantifies the proportion of the entire school-leaving age cohort that is covered by the TIMSS final-year sample in each country. The TCI was defined as follows:

$$TCI = \frac{Total\ Enrollment\ in\ TIMSS\ Grades\ 1995}{(Total\ National\ Population\ Aged\ 15\text{-}19\ in\ 1995)/5}$$

The *numerator* in this expression is the total enrollment in the grades tested by TIMSS, estimated from the weighted sample data. This estimate corresponds to the size of the population to which the TIMSS results generalize and makes appropriate provision for student non-response. It does not include students who are no longer attending school or students who were excluded from the sample on grounds of physical or other disability. It also does not include students who were repeating the final grade. Because some students repeat the final year of a track, or take the final year in more than one track at different times, they may be in the final year of a track without completing their secondary education that year. On the one hand, students who are not completing their education still have the potential to gain further knowledge in additional years of schooling, and thus will not have attained their full yield at the time of the TIMSS assessment. On the other hand, and of more serious concern, the presence both of students who are repeating the final track and of those who will repeat that track can contribute a substantial downward bias to the estimated achievement of the population. Repeating students would be represented twice in the population, and are likely to be lower-achieving on average than those who do not repeat. The only practical way for TIMSS to deal with this problem was to exclude students who were repeating the final year. Thus, the population of final-year students is formally defined as those students taking the final year of one track of the secondary system for the first time.

The *denominator* in the expression is an estimate of the school-leaving age cohort size. Since the age at which students in upper-secondary school may leave school varies, TIMSS estimated the size of the school-leaving age cohort by taking the average of the size of the 1995 age cohorts for 15-, 16-, 17-, 18-, and 19-year-olds in each country. (Although the estimate was generally based on the 15-19 age group, there were exceptions; for example, in Germany it was based on the 17-19 age group.) This information was provided by NRCs from official population census figures in their countries. This approach reflects the fact that students in the final year of secondary school are likely to be almost entirely a subset of the population of 15- to 19-year-olds in most countries. Table A.4 presents the computation of the TCI for each country.

Table A.4

Computation of TCI: Estimated Percentage of School-Leaving Age Cohort Covered by TIMSS Sample - Final Year of Secondary School*

Country	Estimated School-Leaving Age Cohort Size (A)	Estimated Number of Students Represented by Sample (B)	Estimated Number of Students Excluded from Sample (C)	Estimated Number of Other Students Not Represented by Sample (D)	TIMSS Coverage Index (TCI)[†] (B/A)
Australia	250,852	170,849	9,944	70,059	68%
Austria	93,168	70,721	15,682	6,765	76%
Canada	374,499	263,241	25,559	85,699	70%
Cyprus	9,464	4,535	1,279	3,650	48%
Czech Republic	177,180	137,467	8,821	30,892	78%
Denmark	65,683	37,872	872	26,939	58%
France	760,452	637,935	6,509	116,008	84%
Germany	870,857	655,916	83,514	131,427	75%
Hungary	170,524	111,281	201	59,042	65%
Iceland	4,231	2,308	2	1,921	55%
Italy	739,268	380,834	3,459	354,975	52%
Lithuania	52,140	22,160	0	29,980	43%
Netherlands	187,087	145,916	40,293	878	78%
New Zealand	53,284	37,549	4	15,731	70%
Norway	52,180	43,806	1,747	6,627	84%
Russian Federation	2,145,918	1,031,187	777,913	336,818	48%
Slovenia	30,354	26,636	1,706	2,012	88%
South Africa	766,334	374,618	0	391,716	49%
Sweden	101,058	71,333	168	29,557	71%
Switzerland	79,547	65,174	1,671	12,702	82%
United States	3,612,800	2,278,564	88,642	1,245,594	63%

*   See Table 1.2 for more information about the grades tested in each country.

†   TIMSS Coverage Index (TCI): Estimated percentage of school-leaving age cohort covered by TIMSS sample.

SCHOOL AND STUDENT SAMPLING AND PARTICIPATION RATES

Within countries, TIMSS used a two-stage sample design at Population 1 and Population 2, where the first stage involved selecting 150 public and private schools within each country.[2] Within each school, the basic approach required countries to use random procedures to select, for the Population 1 assessment, one mathematics class at the fourth grade and one at the third grade, and for the Population 2 assessment, one mathematics class at the eighth grade and one at the seventh 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, for each population, a representative sample of 7,500 students per country, with approximately 3,750 students at each grade. Tables A.5 and A.6 present the school and student samples sizes for fourth grade, and Tables A.7 and A.8 those for eighth grade.

TIMSS also used a two-stage sample design for Population 3, the final year of secondary school. The first stage involved sampling 120 public and private schools in each country. Within each school, the basic approach required countries to use random procedures to select 40 students. The actual number of schools and students selected depended in part on the structure of the education system – tracked or untracked – and on where the student subpopulations were in the system. Within each sampled school, eligible students were classified as being one of four types (not having taken advanced mathematics or physics, having taken advanced mathematics but not physics, having taken physics but not advanced mathematics, or having taken both advanced mathematics and physics), and a sample of each group was drawn. Test booklets were assigned to students based on their classification. Tables A.9 and A.10 present the school and student samples sizes for the final year of secondary school mathematics and science literacy assessment.

For each assessment, countries were required to achieve a participation rate of at least 85% of both schools and students, or a combined rate of 75% (the product of school and student participation with or without replacement schools). Tables A.11 through A.13 present the school, student, and overall participation rates for fourth grade, eighth grade, and the final year of secondary school.

Table A.5

**School Sample Sizes
Fourth Grade***

Country	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	268	268	169	9	0	178
Austria	150	150	71	31	31	133
Canada	423	420	390	0	0	390
Cyprus	150	150	146	0	0	146
Czech Republic	215	215	181	7	0	188
England	150	145	92	35	0	127
Greece	187	187	174	0	0	174
Hong Kong	156	148	124	0	0	124
Hungary	150	150	150	0	0	150
Iceland	153	151	144	0	0	144
Iran, Islamic Rep.	180	180	180	0	0	180
Ireland	175	173	161	4	0	165
Israel	100	100	40	0	47	87
Japan	150	150	137	4	0	141
Korea	150	150	150	0	0	150
Kuwait	150	150	150	0	0	150
Latvia (LSS)	169	169	125	0	0	125
Netherlands	196	196	63	67	0	130
New Zealand	150	150	120	29	0	149
Norway	150	148	126	13	0	139
Portugal	150	150	143	0	0	143
Scotland	184	184	143	9	0	152
Singapore	191	191	191	0	0	191
Slovenia	150	150	121	0	0	121
Thailand	155	155	154	0	0	154
United States	220	213	182	0	0	182

*   See Table 1.2 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.

Because population coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.

Table A.6

Student Sample Sizes
Fourth Grade*

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

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

*   See Table 1.2 for more information about the grades tested in each country.

Because population coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.

Table A.7

School Sample Sizes
Eighth Grade*

Country	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
Australia	214	214	158	3	161
Austria	159	159	62	62	124
Belgium (Fl)	150	150	92	49	141
Belgium (Fr)	150	150	85	34	119
Bulgaria	167	167	111	4	115
Canada	413	388	363	1	364
Colombia	150	150	136	4	140
Cyprus	55	55	55	0	55
Czech Republic	150	149	143	6	149
Denmark	158	157	144	0	144
England	150	144	80	41	121
France	151	151	127	0	127
Germany	153	150	102	32	134
Greece	180	180	156	0	156
Hong Kong	105	104	85	0	85
Hungary	150	150	150	0	150
Iceland	161	132	129	0	129
Iran, Islamic Rep.	192	191	191	0	191
Ireland	150	149	125	7	132
Israel	100	100	45	1	46
Japan	158	158	146	5	151
Korea	150	150	150	0	150
Kuwait	69	69	69	0	69
Latvia (LSS)	170	169	140	1	141
Lithuania	151	151	145	0	145
Netherlands	150	150	36	59	95
New Zealand	150	150	137	12	149
Norway	150	150	136	10	146
Portugal	150	150	142	0	142
Romania	176	176	163	0	163
Russian Federation	175	175	170	4	174
Scotland	153	153	119	8	127
Singapore	137	137	137	0	137
Slovak Republic	150	150	136	9	145
Slovenia	150	150	121	0	121
South Africa	180	180	107	7	114
Spain	155	154	147	6	153
Sweden	120	120	116	0	116
Switzerland	259	258	247	3	250
Thailand	150	150	147	0	147
United States	220	217	169	14	183

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

* See Table 1.2 for more information about the grades tested in each country.

Because population coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.

Table A.8

Student Sample Sizes
Eighth Grade*

Country	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	8027	63	61	7903	650	7253
Austria	2969	14	4	2951	178	2773
Belgium (Fl)	2979	1	0	2978	84	2894
Belgium (Fr)	2824	0	1	2823	232	2591
Bulgaria	2300	0	0	2300	327	1973
Canada	9240	134	206	8900	538	8362
Colombia	2843	6	0	2837	188	2649
Cyprus	3045	15	0	3030	107	2923
Czech Republic	3608	6	0	3602	275	3327
Denmark	2487	0	0	2487	190	2297
England	2015	37	60	1918	142	1776
France	3141	0	0	3141	143	2998
Germany	3318	0	35	3283	413	2870
Greece	4154	27	23	4104	114	3990
Hong Kong	3415	12	0	3403	64	3339
Hungary	3339	0	0	3339	427	2912
Iceland	2025	10	65	1950	177	1773
Iran, Islamic Rep.	3770	20	0	3750	56	3694
Ireland	3411	28	10	3373	297	3076
Israel	1453	6	0	1447	32	1415
Japan	5441	0	0	5441	300	5141
Korea	2998	31	0	2967	47	2920
Kuwait	1980	3	0	1977	322	1655
Latvia (LSS)	2705	19	0	2686	277	2409
Lithuania	2915	2	0	2913	388	2525
Netherlands	2112	14	1	2097	110	1987
New Zealand	4038	121	12	3905	222	3683
Norway	3482	26	49	3407	140	3267
Portugal	3589	70	13	3506	115	3391
Romania	3899	0	0	3899	174	3725
Russian Federation	4311	42	10	4259	237	4022
Scotland	3289	0	46	3243	380	2863
Singapore	4910	18	0	4892	248	4644
Slovak Republic	3718	5	3	3710	209	3501
Slovenia	2869	15	8	2846	138	2708
South Africa	4793	0	0	4793	302	4491
Spain	4198	27	102	4069	214	3855
Sweden	4483	71	28	4384	309	4075
Switzerland	4989	16	24	4949	94	4855
Thailand	5850	0	0	5850	0	5850
United States	8026	104	108	7814	727	7087

*   See Table 1.2 for more information about the grades tested in each country.

Because population coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.

Table A.9

School Sample Sizes - Mathematics and Science Literacy
Final Year of Secondary School*

Country	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
Australia	132	132	71	16	87
Austria	182	182	74	95	169
Canada	389	389	333	4	337
Cyprus	29	28	28	0	28
Czech Republic	150	150	150	0	150
Denmark	130	130	122	0	122
France	71	71	56	0	56
Germany	174	174	121	31	152
Hungary	204	204	204	0	204
Iceland	30	30	30	0	30
Italy	150	150	93	8	101
Lithuania	168	142	142	0	142
Netherlands	141	141	52	27	79
New Zealand	79	79	68	11	79
Norway	171	171	122	9	131
Russian Federation	175	165	159	4	163
Slovenia	172	172	79	0	79
South Africa	185	140	90	0	90
Sweden	157	157	145	0	145
Switzerland	401	401	378	5	383
United States	250	250	190	21	211

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

* See Table 1.2 for more information about the grades tested in each country.

Table A.10

Student Sample Sizes - Mathematics and Science Literacy
Final Year of Secondary School*

Country	Number of Students Sampled in Participating Schools	Number of Students Withdrawn[†]	Number of Students Excluded	Number of Students Eligible	Number of Students Absent	Number of Participating Students
Australia	4130	37	0	4093	1040	1844
Austria	3693	140	0	3553	398	1779
Canada	11782	732	0	11050	1470	4832
Cyprus	1224	15	0	1209	38	473
Czech Republic	4188	43	0	4145	326	1899
Denmark	5208	0	0	5208	672	2604
France	4096	275	0	3821	600	1590
Germany	6971	94	117	6760	1666	2182
Hungary	5493	265	0	5228	137	5091
Iceland	2500	132	2	2366	663	1703
Italy	2426	148	3	2275	192	1578
Lithuania	4196		0	4195	574	2887
Netherlands	1882		20	1681	211	1470
New Zealand	2687	580	1	2106	343	1763
Norway	4056	76	65	3915	349	2518
Russian Federation	5356	536	0	4820	182	2289
Slovenia	3755	37	1	3717	282	1387
South Africa	3695	906	0	2789	32	2757
Sweden	5362	184	12	5166	589	2816
Switzerland	5939	258	0	5681	262	2976
United States	14812	603	293	13916	3082	5371

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

* See Table 1.2 for more information about the grades tested in each country.

† Sampled students who reported that they were repeating the final year, were incorrectly classified, or were otherwise ineligible.

Table A.11

Participation Rates
Fourth Grade*

Country	School Participation		Student Participation (Weighted Percentage)	Overall Participation	
	School Participation Before Replacement (Weighted Percentage)	School Participation After Replacement (Weighted Percentage)		Overall Participation Before Replacement (Weighted Percentage)	Overall Participation After Replacement (Weighted Percentage)
Australia	66	69	96	63	66
Austria	51	72	96	49	69
Canada	90	90	96	86	86
Cyprus	97	97	86	83	83
Czech Republic	91	94	92	84	86
England	63	88	95	60	83
Greece	93	93	95	88	88
Hong Kong	84	84	98	83	83
Hungary	100	100	92	92	92
Iceland	95	95	90	86	86
Iran, Islamic Rep.	100	100	97	97	97
Ireland	94	96	93	88	90
Israel	40	40	94	38	38
Japan	93	96	97	90	92
Korea	100	100	95	95	95
Kuwait	100	100	95	95	95
Latvia (LSS)	74	74	93	69	69
Netherlands	31	62	96	29	59
New Zealand	80	99	96	77	95
Norway	85	94	97	82	91
Portugal	95	95	96	92	92
Scotland	78	83	92	71	76
Singapore	100	100	98	98	98
Slovenia	81	81	94	76	76
Thailand	96	96	100	96	96
United States	85	85	94	80	80

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

*   See Table 1.2 for more information about the grades tested in each country.

 Because population coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.

Table A.12

Participation Rates
Eighth Grade*

Country	School Participation		Student Participation (Weighted Percentage)	Overall Participation	
	School Participation Before Replacement (Weighted Percentage)	School Participation After Replacement (Weighted Percentage)		Overall Participation Before Replacement (Weighted Percentage)	Overall Participation After Replacement (Weighted Percentage)
Australia	75	77	92	69	70
Austria	41	84	95	39	80
Belgium (Fl)	61	94	97	59	91
Belgium (Fr)	57	79	91	52	72
Bulgaria	72	74	86	62	63
Canada	90	91	93	84	84
Colombia	91	93	94	85	87
Cyprus	100	100	97	97	97
Czech Republic	96	100	92	89	92
Denmark	93	93	93	86	86
England	56	85	91	51	77
France	86	86	95	82	82
Germany	72	93	87	63	81
Greece	87	87	97	84	84
Hong Kong	82	82	98	81	81
Hungary	100	100	87	87	87
Iceland	98	98	90	88	88
Iran, Islamic Rep.	100	100	98	98	98
Ireland	84	89	91	76	81
Israel	45	46	98	44	45
Japan	92	95	95	87	90
Korea	100	100	95	95	95
Kuwait	100	100	83	83	83
Latvia (LSS)	83	83	90	75	75
Lithuania	96	96	87	83	83
Netherlands	24	63	95	23	60
New Zealand	91	99	94	86	94
Norway	91	97	96	87	93
Portugal	95	95	97	92	92
Romania	94	94	96	89	89
Russian Federation	97	100	95	93	95
Scotland	79	83	88	69	73
Singapore	100	100	95	95	95
Slovak Republic	91	97	95	86	91
Slovenia	81	81	95	77	77
South Africa	60	64	97	58	62
Spain	96	100	95	91	94
Sweden	97	97	93	90	90
Switzerland	93	95	98	92	94
Thailand	99	99	100	99	99
United States	77	85	92	71	78

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

*   See Table 1.2 for more information about the grades tested in each country.

Because population coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.

Table A.13

Participation Rates - Mathematics and Science Literacy
Final Year of Secondary School*

Country	School Participation		Student Participation (Weighted Percentage)	Overall Participation	
	School Participation Before Replacement (Weighted Percentage)	School Participation After Replacement (Weighted Percentage)		Overall Participation Before Replacement (Weighted Percentage)	Overall Participation After Replacement (Weighted Percentage)
Australia	48.8	66.2	78.1	38.1	51.8
Austria	35.9	90.9	79.7	28.6	72.5
Canada	82.2	82.6	82.7	68.0	68.3
Cyprus	100.0	100.0	98.2	98.2	98.2
Czech Republic	100.0	100.0	92.2	92.2	92.2
Denmark	54.9	54.9	88.9	48.8	48.8
France	80.3	80.3	85.6	68.7	68.7
Germany	88.7	100.0	80.1	71.0	80.1
Hungary	100.0	100.0	97.7	97.7	97.7
Iceland	100.0	100.0	73.6	73.6	73.6
Italy	59.9	65.0	94.8	56.8	61.6
Lithuania	97.1	97.1	87.9	85.4	85.4
Netherlands	35.8	56.3	87.6	31.3	49.3
New Zealand	87.0	100.0	80.6	70.1	80.6
Norway	74.1	80.0	88.9	65.9	71.1
Russian Federation	93.0	99.3	90.9	84.6	90.3
Slovenia	45.6	45.6	92.8	42.3	42.3
South Africa	65.0	65.0	99.4	64.6	64.6
Sweden	95.3	95.3	86.5	82.4	82.4
Switzerland	87.0	89.1	95.0	82.6	84.6
United States	77.1	85.1	74.6	57.6	63.5

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

* See Table 1.2 for more information about the grades tested in each country.

INDICATING COMPLIANCE WITH SAMPLING GUIDELINES

In Figures A.1, A.2, and A.3, countries are grouped by how they met the TIMSS sampling requirements. Countries 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 – and that complied with the TIMSS guidelines for grade selection and classroom sampling are shown in the first panel of each figure. Countries that met the guidelines only after including replacement schools are annotated.

Countries not reaching at least 50% school participation without the use of replacement schools, or that failed to reach the participation standard even with the inclusion of replacement schools, are shown in the second panel of the figures.

For the Population 1 and Population 2 assessments, some countries sought to provide a better curricular match and thus did not test the two grades required by the TIMSS population definition (for Population 1, the two grades with the most 9-year-olds at the time of testing, and for Population 2 the two grades with the most 13-year-olds at the time of testing). This led to their students being somewhat older than those in the other countries. These countries are grouped together in Figures A.1 and A.2.

For a variety of reasons, some countries did not comply with the guidelines for within-school sampling or had difficulty meeting several sampling guidelines; these are grouped together in the figures.

---

[2] The sample design for TIMSS is described in detail in Foy, P., Schleicher, A., and Rust, K. (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.

Figure A.1

Countries Grouped According to Their Compliance with Guidelines for Sample Implementation and Participation Rates - Fourth Grade*

Fourth Grade

Countries satisfying guidelines for sample participation rates, grade selection, and sampling procedures

Canada	Japan
Cyprus	Korea
Czech Republic	New Zealand
[1,2] England	Norway
Greece	Portugal
Hong Kong	[†] Scotland
Iceland	Singapore
Iran, Islamic Rep.	United States
Ireland	

Countries not satisfying guidelines for sample participation

Australia	[1] Latvia (LSS)
Austria	Netherlands

Countries not meeting age/grade specifications (high percentage of older students)

Slovenia

Countries with unapproved sampling procedures at the classroom level

Hungary

Countries with unapproved sampling procedures at classroom level and not meeting other guidelines

[1] Israel	Thailand
Kuwait	

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

1 National Desired Population does not cover all of International Desired Popoulation (see Table A.1).
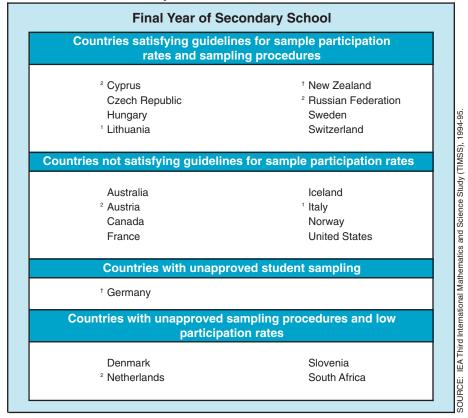
2 National Defined Population covers less than 90 percent of National Desired Population (see Table A.1).

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

* See Table 1.2 for more information about the grades tested in each country.

Because population coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.

Figure A.2

Countries Grouped According to Their Compliance with Guidelines for Sample Implementation and Participation Rates - Eighth Grade*

<table>
<tr><td colspan="2" align="center">Eighth Grade</td></tr>
<tr><td colspan="2" align="center">Countries satisfying guidelines for sample participation rates, grade selection, and sampling procedures</td></tr>
<tr>
<td>
† Belgium (Fl)
Canada
Cyprus
Czech Republic
†2 England
France
Hong Kong
Hungary
Iceland
Iran, Islamic Rep.
Ireland
Japan
Korea
</td>
<td>
1 Latvia (LSS)
1 Lithuania
New Zealand
Norway
Portugal
Russian Federation
Singapore
Slovak Republic
Spain
Sweden
1 Switzerland
† United States
</td>
</tr>
<tr><td colspan="2" align="center">Countries not satisfying guidelines for sample participation</td></tr>
<tr>
<td>
Australia
Austria
Belgium (Fr)
</td>
<td>
Bulgaria
Netherlands
† Scotland
</td>
</tr>
<tr><td colspan="2" align="center">Countries not meeting age/grade specifications
(high percentage of older students)</td></tr>
<tr>
<td>
Colombia
†1 Germany
</td>
<td>
Romania
Slovenia
</td>
</tr>
<tr><td colspan="2" align="center">Countries with unapproved sampling
procedures at the classroom level</td></tr>
<tr>
<td>
Denmark
Greece
</td>
<td>
Thailand
</td>
</tr>
<tr><td colspan="2" align="center">Countries with unapproved sampling procedures at
classroom level and not meeting other guidelines</td></tr>
<tr>
<td>
1 Israel
Kuwait
</td>
<td>
South Africa
</td>
</tr>
</table>

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

1   National Desired Population does not cover all of International Desired Popoulation (see Table A.2).

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

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

*   See Table 1.2 for more information about the grades tested in each country.

Figure A.3

Countries Grouped According to Their Compliance with Guidelines for Sample Implementation and Participation Rates – Mathematics and Science Literacy* Final Year of Secondary School

<table>
<tr><td colspan="2" align="center">Final Year of Secondary School</td></tr>
<tr><td colspan="2" align="center">Countries satisfying guidelines for sample participation rates and sampling procedures</td></tr>
<tr>
<td>
² Cyprus
Czech Republic
Hungary
¹ Lithuania
</td>
<td>
† New Zealand
² Russian Federation
Sweden
Switzerland
</td>
</tr>
<tr><td colspan="2" align="center">Countries not satisfying guidelines for sample participation rates</td></tr>
<tr>
<td>
Australia
² Austria
Canada
France
</td>
<td>
Iceland
¹ Italy
Norway
United States
</td>
</tr>
<tr><td colspan="2" align="center">Countries with unapproved student sampling</td></tr>
<tr><td colspan="2">† Germany</td></tr>
<tr><td colspan="2" align="center">Countries with unapproved sampling procedures and low participation rates</td></tr>
<tr>
<td>
Denmark
² Netherlands
</td>
<td>
Slovenia
South Africa
</td>
</tr>
</table>

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

1 National Desired Population does not cover all of International Desired Popoulation (see Table A.3).

2 National Defined Population covers less than 90 percent of National Desired Population (see Table A.3).

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

* See Table 1.2 for more information about the grades tested in each country.

DATA COLLECTION PROCEDURES

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 detailed procedures for receipt and distribution of materials as well as for the activities related to the testing sessions. The test administrator manuals covered 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 for describing these in their NRC's reports. In addition, the International Study Center considered it essential to establish some method to monitor compliance with standard procedures. NRCs were asked to nominate a person, such as a retired school teacher, to serve as quality control monitor for their countries, and in almost all cases the International Study Center adopted the NRCs' 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 about 10 schools to visit, where they observed testing sessions and interviewed school coordinators.[3] 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 collect the data in an efficient and professional manner. Similarly, the TIMSS tests appeared to have been administered in compliance with international procedures throughout the activities preliminary to the testing session, those during testing, and the school-level activities related to receiving, distributing, and returning materials from the national centers.

SCORING THE FREE-RESPONSE ITEMS

Because about 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 used 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

---

[3] 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.

level of the response. The second digit, combined with the first, represents a diagnostic code used to identify specific types of approaches, strategies, or common errors and misconceptions. Although not specifically used to estimate overall proficiency in mathematics and science, 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 ensure reliable scoring procedures based on the TIMSS rubrics, the International Study Center prepared guides containing the rubrics and explaining 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. These sessions were designed to assist representatives of national centers who would then be responsible for training personnel in their countries to apply the two-digit codes reliably.[4]

To gather and document empirical information about the within-country agreement among scorers, TIMSS developed a procedure whereby systematic subsamples of some 10% of the students' responses were coded independently by two scorers. The percentage of exact agreement between the scorers was computed for each free-response item based on both the score level (first digit) and the diagnostic code (second digit) level. A very high percentage of exact agreement at the score level was observed for the free-response items on all TIMSS tests.[5]

DATA PROCESSING

To ensure the availability of comparable, high-quality data for analysis, TIMSS undertook a rigorous set of quality control steps to create the international database.[6] TIMSS prepared manuals and software for countries to use in entering their data so that the information would be in a standard international format before being forwarded to the IEA Data Processing Center in Hamburg. Upon arrival at the Center, the data from each country underwent an exhaustive cleaning process. That process involved several iterative steps and procedures designed to identify, document, and correct

---

[4] 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,*

[5] Summaries of the scoring reliability data for each test are included in the appendices of the international reports (see references in Chapter 1).

[6] 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.

deviations from the international instruments, file structures, and coding schemes. The 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 reviewed the item statistics of each cognitive item in each country to identify poorly performing items. 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 deviations in translation, adaptation, or printing.

IRT Scaling and Data Analysis

The mathematics and science achievement results were summarized using an item response theory (IRT) scaling method (Rasch model).[7] This scaling method produces a test score by averaging the responses of each student to the items they took in a way that takes into account the difficulty of each item. The method used in TIMSS includes 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 each TIMSS test address a wide range of mathematics or science content, the performance of the students across the items was sufficiently consistent to be usefully summarized in a single score per test.

The IRT method was preferred for developing comparable estimates of performance for all students, since students answered different test items depending upon which test booklet 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. For Population 1 and Population 2, each 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 vary in size, each country was reweighted to contribute equally to the mean and standard deviation of the scale. The international averages of the Population 1

---

[7] The TIMSS scaling model is fully documented in Adams, R.J., Wu, M.L., and Macaskill, G. (1997). "Scaling Methodology and Procedures for the Mathematics and Science Scales" in M.O. Martin and D.L. Kelly (eds.), *Third International Mathematics and Science Study Techni-*

scale scores (mathematics and science) were constructed to be the averages of the 26 means of countries that were available at fourth grade and the 24 means of those at third grade. The international averages of the Population 2 scale scores (mathematics and science) were constructed to be the averages of the 41 means of countries that were available at eighth grade and the 39 means of those at seventh grade. For the Population 3 mathematics and science literacy assessment, the mathematics literacy scale and the science literacy scale were constructed using data from the 21 countries that participated in the assessment and have an average of 500 and a standard deviation of 100.

ESTIMATING SAMPLING ERROR

Because the statistics presented in this report are national estimates based on samples of schools and students rather than the values that could be calculated if every school and student in a country 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.[8] The use of confidence intervals, based on the standard errors, allows inferences to be made 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.

[8] The jackknife repeated replication technique for estimating sampling errors is documented in Gonzalez, E.J. and Foy, P. (1997). "Estimation of Sampling Variability, Design Effects, and Effective Sample Sizes" in M.O. Martin and D.L. Kelly (eds.), *Third International Mathematics and Science Study Technical Report, Volume II*. Chestnut Hill, MA: Boston College.