APPENDIX A

A

:

A t

A

History

TIMSS 1999 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 the curricular areas of mathematics, science, language, civics, and reading. IEA conducted its First International Science Study (FISS) in 1970-71, and the Second International Science Study (SISS) in 1983 and 1984. Both studies tested students in both eighth grades, and the final year of se

In 1999, TIMSS again assessed eighth-grade stu
ics and science to measure trends in student ac
TIMSS 19991995, 26AVAILED THEMSELVES OF TI
Exhibit A.1). Twelve additional countries partic
total of 38 countries. Of those taking part in 19
pated in 1995 at the fourth grade.

A.1

[2] Since fourth-grade students in 1995 were in eighth grade in 1999, these countries can compare the eighth-grade performance of this cohort of students with their performance at the fourth grade, as well as with the eighth-grade performance of students in other countries.

[1] Results for 41 countries are reported in the 1995 international reports. Italy also completed the 1995 testing, but too late to be included in the international reports. It is counted as a 1995 country in this report and included in all trend exhibits in the 1999 international reports. Unweighted data for the Philippines were reported in an appendix to the international reports in 1995. These data were not included in trend exhibits in the 1999 international reports.

	TIMSS 1999	TIMSS 1995 (Grade 8)	TIMSS 1995 (Grade 4)
Australia			

Developing the TIMSS 1999 Science Test

The timss curriculum framework underlying the science tests was developed for timss in 1995 by groups of science educators with input from the timss National Research Coordinators (nrc

Content	Performance Expectations	Perspectives
Earth Sciences	Understanding	Attitudes
Life Sciences	Theorizing, Analyzing, and Solving Problems	Careers
Physical Sciences	Using Tools, Routine Procedures and Science Processes	Participation
Science, Technology, and Mathematics	Investigating the Natural World	Increasing Interest
	Communicating	Safety
		Habits of Mind

Exhibit A.3 Distribution of Science Items by Content Reporting Category and Performance Category

TIMSS 1999 8th grade Science

[illegible] 9	[illegible] 9	[illegible]	[illegible] -	[illegible] - 1	[illegible] 2
Earth Science	15	22	17	5	23
Life Science	27	40	28	12	42
Physics	27	39	28	11	39
Chemistry	14	20	15	5	22
Environmental and Resource Issues	9	13	7	6	14
Scientific Inquiry and the Nature of Science	8	12	9	3	13
[illegible]	100	146	104	42	153

[illegible] 9	[illegible] 9	[illegible]	[illegible] -	[illegible] - 1	[illegible] 2
Understanding Simple Information	39	57	56	1	57
Understanding Complex Information	31	45	30	15	47
Theorizing, Analyzing and Solving Problems	19	28	5	23	32
Using Tools, Routine Procedures and Science Processes	7	10	9	1	10
Investigating the Natural World	4	6	4	2	7
[illegible]	100	146	104	42	153

1 Free response items include both short-answer and extended-response types.

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

Translation and Verification

The timss instruments were prepared in English and translated into 33 languages, with 10 of the 38 countries collecting data in two languages. In addition, it sometimes was necessary to modify the international versions for cultural reasons, even in the nine 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, using two or more independent translations; (3) consultation with subject-matter experts on cultural adaptations to ensure that the meaning and difficulty of items did not change; (4) verification of translation quality by professional translators from an independent translation company; (5) corrections by the national centers in accordance with the suggestions made; (6) verification by the International Study Center that corrections were made; and (7) a series of statistical checks after the testing to detect items that did not perform comparably across countries.[7]

Population Definition and Sampling

timss in 1995 had as its target population students enrolled in the two adjacent grades that contained the largest proportion of 13-year-old students at the time of testing, which were seventh- and eighth-grade students in most countries. timss in 1999 used the same definition to identify the target grades, but assessed students in the upper of the two grades only, which was the eighth grade in most countries.[8]

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 and that 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. The sampling documentation was used by the International Study Center, in consultation with Statistics Canada and the sampling referee, to evaluate the quality of the samples.

7 More details about the translation verification procedures can be found in O'Connor, K., and Malak, B. (2000), "Translation and Cultural Adaptation of the TIMSS Instruments" in M.O. Martin, K.D. Gregory and S.E. Stemler (eds.), *TIMSS 1999 Technical Report*, Chestnut Hillry and Sbaptation of the TI.pennicS InstrumC88 -gegory andhn the two

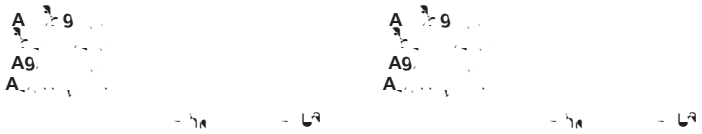
In a few situations where it was not possible to test the entire internationally desired population (all students in the upper of the two adjacent grades with the greatest proportion of 13-year-olds), countries were permitted to define a national desired population that excluded part of the internationally desired population. Exhibit A.5 shows any

	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

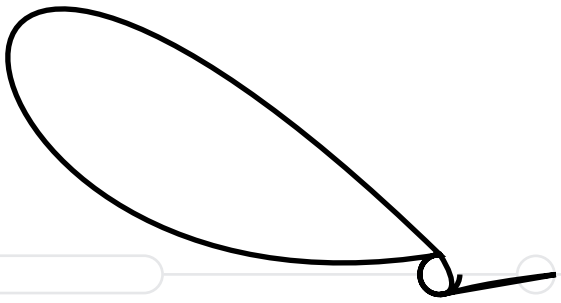
Data Collection

Each participating country was responsible for carrying out all aspects of the data collection, using standardized procedures developed for the study. Training manuals were created 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. These 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 in the nrc



A dash (–) indicates data are not available.



IRT Scaling and Data Analysis

The general approach to reporting the TIMSS achievement data was based primarily on item response theory (IRT) scaling methods.[11] The science results were summarized using a family of 2-parameter and 3-parameter IRT models for dichotomously-scored items (right or wrong), and generalized partial credit models for items with 0, 1, or 2 available score points. The IRT scaling method produces a score by averaging the responses of each student to the items that he or she took in a way that takes into account the difficulty and discriminating power of each item. The methodology used in TIMSS includes refinements that enable reliable scores to be produced even though individual students responded to relatively small subsets of the total science item pool. Achievement scales were produced for each of the six science content areas (earth science, life science, physics, chemistry, environmental and resource issues, and scientific inquiry and the nature of science) as well as for science overall.

The IRT methodology 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. To provide a reliable measure of student achievement in both 1999 and 1995, the overall science scale was calibrated using students from the countries that participated in both years. When all countries participating in 1995 at the eighth grade are treated equally, the TIMSS scale average over those countries is 500 and the standard deviation is 100. Since the countries varied in size, each country was weighted to contribute equally to the mean and standard deviation of the scale. The average and standard deviation of the scale scores are arbitrary and do not affect scale interpretation. When the metric of the scale had been established, students from the countries that tested in 1999 but not 1995 were assigned scores on the basis of the new scale.

IRT scales were also created for each of the six science content areas for the 1999 data. However, insufficient items were used both in 1995 and in 1999 to establish reliable IRT content area scales for trend purposes. The trend exhibits presented in Chapter 3 were based on the average percentage of students responding correctly to the common items in each content area.

[11] For a detailed description of the TIMSS scaling, see Yamamoto, K., and Kulick, E. (2000), "Scaling Methods and Procedures for the TIMSS Mathematics and Science Scales" in M.O. Martin, K.D. Gregory and S.E. Stemler (eds.), *TIMSS 1999 Technical Report*, Chestnut Hill, MA: Boston College.

To allow more accurate estimation of summary statistics for student subpopulations, the

multiple comparison charts (Exhibit 1.2 and those in Appendix B), the Bonferroni procedure adjusts for the number of countries in the chart, minus one. In exhibits where a country statistic is compared to the international average, the adjustment is for the number of countries.[13]

Setting International Benchmarks of Student Achievement

International benchmarks of student achievement were computed at each grade level for both mathematics and science. The benchmarks are points in the weighted international distribution of achievement scores that separate the 10 percent of students located on top of the distribution, the top 25 percent of students, the top 50 percent, and the bottom 25 percent. The percentage of students in each country meeting or exceeding the international benchmarks is reported. The benchmarks correspond to the 90th, 75th, 50th, and 25th percentiles of the international distribution of achievement. When computing these percentiles, each country contributed as many students to the distribution as there were students in the target population in the country. That is, each country's contribution to setting the international benchmarks was proportional to the estimated population enrolled at the eighth grade.

In order to interpret the timss scale scores and analyze achievement at the international benchmarks, timss conducted a scale anchoring analysis to describe achievement of students at those four points on the scale. Scale anchoring is a way of describing students' performance at different points on a scale in terms of what they know and can do. It involves a statistical component, in which items that discriminate between successive points on the scale are identified, and a judgmental component in which subject-matter experts examine the items and generalize to students' knowledge and understandings.

Science Curriculum Questionnaire

In an effort to collect information about the content of the intended curriculum in science, timss asked National Research Coordinators to complete a questionnaire about the structure, organization, and content coverage of their national curricula. nrcs reviewed 42 science topics and reported the percentage of their eighth-grade students for which each topic was intended in their curriculum. Although most topic descriptions were used without modification, there were occasions when nrcs found it necessary to expand on or qualify the topic description to describe their situation accurately. These country-specific adaptations to the science curriculum questionnaire are presented in Exhibit A.11.

Exhibit A.11

Australia	t : t t (t t, , t t t ,).	A t t (t t 90%)	t t , , t t t , & t t . 8.
	B . : t t . t . (t , t)	A t t (t t 90%)	t t , t t . 8.
	t : t t tt (t , , , t)	A t t (t t 90%)	. t t t .
	t : A , , t	A t t (t t 90%)	. t t t .
	: , , t	A t t (t t 90%)	. t t .
	: t f(, / , , , t)	A t t (t t 90%)	A t t t . 8.
Belgium	B . : t , . t , t , t	A t t (t t 90%)	t , t , t t t . 8.
	t	N t t . 8	t t t tt . t t . 8, t t .
	: t tt (. t , , t t tt , . , .)	A t t (t t 90%)	t . t t t j t t t .
	: . t , , (, t , t , . t ;)	A t t (t t 90%)	t t . 8.
	: . t (t , t , . t)	A t t (t t 90%)	t . t t t j t t t .
Chile	t : t , t (, , t , ,)	A t t (t t 90%)	& t t . 8.
	t : t t (t t, , t t t ,).	A t t (t t 90%)	t t t & t t . 8.
	B . : t , . t , t , t	A t t (t t 90%)	t , t , t t t . 8.
	t : t t tt (t , , , t)	A t t (t t 90%)	A , , t t t . 8.
	t : t t t t (t . , t , t)	A t t (t t 90%)	t t t . 8.
Chinese Taipei	B . : t t , t ,	N t t . 8	t t , t , t t , t t t t .
	: . t , , (, t , t , . t ;)	N t t . 8	, t , t . t t t t t . 9. t t . t
	: (t t t t t / /)	A t t (t t 90%)	t t t . 8.
	: t t . t (t , t t , . t)	A t t (t t 90%)	. t t t . 8.
	t t N t : t t (t . t , . t ,	A t t (t t 90%)	t . t , . , . t t . 8.
	t t N t : t t (t , t , t ,	A t t (t t 90%)	t t t . 8.
Cyprus	t : t , t (, , t , ,)	A t t (t t 90%)	, , & t t . 8.
	t : t t (t t, , t t t ,).	A t t (t t 90%)	t t t & t t . 8.
	B . : B . t (t , t t , ,)	A t t (t t 90%)	t t t . 8.
	B . : t , . t , t , t	A t t (t t 90%)	t , t , t t t . 8.
	t : t t tt (t , , , t)	A t t (t t 90%)	t t t . 8.
	: . t , , (, t , t , . t ;)	A t t (t t 90%)	t t . 8.

Iran	t : . (t t , t t)	A t t t (t t 90%)	t . 8.
	: t (, / , , , t)	N t . . 8.	/ t . 8.
Israel	B . : (t , t , . t)	A t t t (t t 90%)	t & . t t . . 8.
	B . : t , . t , t , t	N t . . 8	t t . . 8.
Japan	: t (, / , , , t)	A t t t (t t 90%)	t t . . 8.
	t t N t : t t (t , t , t , ,)	N t . . 8.	t t . 8.
Korea	B . : B . t (t , t t , ,)	A t t t (t t 90%)	t t . . 8.
	t : t t tt (t , , , t)	A t t t (t t 90%)	t t t . . 8.
	t :A , , t	A t t t (t t 90%)	t t t . . 8.
	: t t (t , t , t)	N t . . 8.	t t . . 8.
	: . t , , (, t , t , . t ;)	A t t t (t t 90%)	t t . 8.
	: . t (t , t , t)	A t t t (t t 90%)	. t t t . . 8.
	: t t . t (t , t t , . t)	A t t t (t t 90%)	t t t t . . 8.
	: t (, / , , , t)	A t t t (t t 90%)	t t . . 8.
	t t N t : t t (t . t , . t , . .)	N t . . 8.	. t t . . 8.
	t t N t : t t (t , t , t , ,)	N t . . 8.	t . . 8.
New Zealand	B . : t t . t . (t , t)	A t t t (t t 90%)	B t t . . 8.
	B . : t , . t , t , t	A t t t t t	t t t . . 8.
	t : t t tt (t , , , t)	A t t t t t	t t . . 8.
	t : t t (t , t , / t)	A t t t (t t 90%)	t t t .
	t : t t t t t (t . , t , t)	A t t t (t t 90%)	t . t . . 8.
	t : . (t t , t t)	A t t t t t	t t t t t . . 8.
	: . t , , (, t , t , . t ;)	A t t t (t t 90%)	t . t t . . 8 (6).
	: t (, / , , , t)	A t t t (t t 90%)	t t . . 8.
	t t N t : t t (t , t , t , ,)	A t t t t t	t t t . 8.

Russian Federation	B . : t t . t . (t , t)	N t t t . . 8.	
	B . : t , t , t , t	N t t t . . 8.	
	t : t t tt (t , , , t)	A t t t t t . . 8.	
	t : t t (t , t , / t)	A t t t t t t . t (t t 90%) . 8.	
	: . t , , (, t , t , t . ;)	A t t . t . t t . . 8. t (t t 90%)	
	: t t t	A t t t t t . . 8. t (t t 90%)	
	: t (,		