NCES
National Center for Education Statistics

TIMSS USA

U.S. Department of Education
Institute of Education Sciences
NCES 2005–005

Highlights From the Trends in International Mathematics and Science Study (TIMSS) 2003

December 2004

Patrick Gonzales
National Center for Education Statistics

Juan Carlos Guzmán
Lisette Partelow
Erin Pahlke
Education Statistics Services Institute

Leslie Jocelyn
David Kastberg
Trevor Williams
Westat

Patrick Gonzales
Project Officer
National Center for Education Statistics

U.S. Department of Education
Rod Paige
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Robert Lerner
Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

December 2004

The NCES World Wide Web Home Page is http://nces.ed.gov
The NCES World Wide Electronic Catalog is http://nces.ed.gov

Suggested Citation

Patrick Gonzales, Juan Carlos Guzmán, Lisette Partelow, Erin Pahlke, Leslie Jocelyn, David Kastberg, and Trevor Williams. (2004). *Highlights From the Trends in International Mathematics and Science Study (TIMSS) 2003* (NCES 2005–005). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

For ordering information on this report, write:

U.S. Department of Education
ED Pubs
P.O. Box 1398
Jessup, MD 20794-1398

Call toll free 1-877-4ED-PUBS or order online at http://www.edpubs.org

Content Contact:
Patrick Gonzales, (202) 502-7421
E-mail: TIMSS@ed.gov

Acknowledgments

The authors wish to thank the students, teachers, and school officials who participated in TIMSS 2003. Their time and effort have provided invaluable data to the nation. The authors also wish to thank all of those who contributed to the design, writing, production, and review of this report for their thoughtful critique, insightful suggestions, and creativity. In particular, the authors acknowledge the contributions of Arnold Goldstein, Eugene Owen, Val Plisko, Taslima Rahman, Elois Scott, and Marilyn Seastrom of the National Center for Education Statistics; Deven Carlson, Fraser Ireland, Angie KewalRamani, Mike Planty, and Robert Stillwell of the Education Statistics Services Institute; Stephen Roey of Westat; Tom Loveless of the Brookings Institution; and the members of the TIMSS 2003 Expert Panel: Rolf Blank, Council of Chief State School Officers; Betsy Brand, American Youth Policy Forum; Nancy Bunt, Alleghany (PA) Intermediate Unit, Math and Science Collaborative; Rodger Bybee, Biological Sciences Curriculum Study; Joan Ferrini-Mundy, Michigan State University; Ramesh Gangolli, University of Washington; Patricia Harvey, St. Paul Public Schools; Kati Haycock, the Education Trust; and Jack Jennings, Center on Education Policy. In addition, the authors acknowledge Brian Henigin and Michael Stock of Westat for the design and layout of the report.

Table of Contents

List of Tables

List of Figures

List of Exhibits

Introduction

The Trends in International Mathematics and Science Study (TIMSS) 2003 is the third comparison of mathematics and science achievement carried out since 1995 by the International Association for the Evaluation of Educational Achievement (IEA), an international organization of national research institutions and governmental research agencies. TIMSS can be used to track changes in achievement over time. Moreover, TIMSS is closely linked to the curricula of the participating countries, providing an indication of the degree to which students have learned concepts in mathematics and science they have encountered in school. In 2003, some 46 countries participated in TIMSS, at either the fourth- or eighth-grade level, or both.

This summary highlights initial findings on the performance of U.S. fourth- and eighth-grade students relative to their peers in other countries on the TIMSS assessment. The summary is based on the findings presented in two reports published by IEA:

- *TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Eighth and Fourth-Grades* (Martin et al. 2004) and
- *TIMSS 2003 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Eighth and Fourth-Grades* (Mullis et al. 2004).

These two IEA reports were simultaneously published with this summary report and are available online at http://www.timss.org.

This summary report describes the mathematics and science performance of fourth- and eighth-graders in participating countries over time. For a number of the participating countries, changes in mathematics and science achievement can be documented over 8 years, from 1995 to 2003. For others, changes can be documented over a shorter period of time, 4 years from 1999 to 2003. Table 1 shows the countries that participated in TIMSS 2003, and their participation in earlier TIMSS data collections. The fourth-grade assessment was offered in 1995 and 2003, while the eighth-grade assessment was offered in 1995, 1999, and 2003.

Table 1. Participation in the TIMSS fourth-grade and eighth-grade assessments, by country: 1995, 1999, and 2003

Country	Fourth grade		Eighth grade		
	1995	2003	1995	1999	2003
Total	15	25	22	29	45
Armenia		✓			✓
Australia[1]	✓	✓	✓		✓
Bahrain					✓
Belgium-Flemish[2]		✓	✓	✓	✓
Botswana					✓
Bulgaria			✓	✓	✓
Chile				✓	✓
Chinese Taipei		✓		✓	✓
Cyprus	✓	✓	✓	✓	✓
England[3]	✓	✓			
Egypt					✓
Estonia					✓
Ghana					✓
Hong Kong SAR[4]	✓	✓	✓	✓	✓
Hungary	✓	✓	✓	✓	✓
Indonesia				✓	✓
Iran, Islamic Republic of	✓	✓	✓	✓	✓
Israel[5]				✓	✓
Italy[5]		✓		✓	✓
Japan	✓	✓	✓	✓	✓
Jordan				✓	✓
Korea, Republic of			✓	✓	✓
Latvia[6]	✓	✓	✓	✓	✓
Lebanon					✓
Lithuania		✓	✓	✓	✓

See notes at end of table.

[1]Table A7 in appendix A groups the participating countries by continent and membership in the Organization for Economic Cooperation and Development (OECD), an intergovernmental organization of 30 industrialized countries that serves as a forum for members to cooperate in research and policy development on social and economic topics of common interest.

Table 1. Participation in the TIMSS fourth-grade and eighth-grade assessments, by country: 1995, 1999, and 2003—Continued

Country	Fourth grade		Eighth grade		
	1995	2003	1995	1999	2003
Macedonia, Republic of				✓	✓
Malaysia				✓	✓
Moldova, Republic of		✓		✓	✓
Morocco[5]		✓			✓
Netherlands	✓	✓	✓	✓	✓
New Zealand	✓	✓	✓	✓	✓
Norway	✓	✓	✓		✓
Palestinian National Authority					✓
Philippines		✓		✓	✓
Romania			✓	✓	✓
Russian Federation		✓	✓	✓	✓
Saudi Arabia					✓
Scotland	✓	✓	✓		✓
Serbia					✓
Singapore	✓	✓	✓	✓	✓
Slovak Republic			✓	✓	✓
Slovenia[1]	✓	✓	✓		✓
South Africa[7]				✓	✓
Sweden			✓		✓
Tunisia		✓		✓	✓
United States	✓	✓	✓	✓	✓

[1]Because of national-level changes in the starting age/date for school, 1999 data for Australia and Slovenia cannot be compared to 2003.
[2]Only the Flemish education system in Belgium participated in TIMSS in 2003.
[3]England collected data at grade 8 in 1995, 1999 and 2003, but due to problems with meeting the minimum sampling requirements for 2003, its eighth-grade data are not shown in this report.
[4]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[5]Because of changes in the population tested, 1995 data for Israel and Italy, and 1999 data for Morocco are not shown.
[6]Only Latvian-speaking schools were included in 1995 and 1999. For trend analyses, only Latvian-speaking schools are included in the estimates.
[7]Because within classroom sampling was not accounted for, 1995 data are not shown for South Africa.
NOTE: Countries that participated in 1995 and 1999 but did not participate in 2003 are not shown. Only countries that completed the necessary steps for their data to appear in the reports from the International Study Center are listed. In addition to the countries listed above, four separate jurisdictions participated in TIMSS 2003: the provinces of Ontario and Quebec in Canada; the Basque region of Spain; and the state of Indiana. Information on these four jurisdictions can be found in the international TIMSS 2003 reports. The Syrian Arab Republic participated in TIMSS 2003 at the eighth-grade level, but due to sampling difficulties, it is not shown in this report. Yemen participated in TIMSS 2003 at the fourth-grade level, but because it did not comply with the minimum sample requirements, it is not shown in this report. Countries could participate at either grade level. Countries were required to sample students in the upper of the two grades that contained the largest number of 9-year-olds and 13-year-olds, respectively. In the United States and most countries, this corresponds to grade 4 and grade 8. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995, 1999, and 2003.

Average student performance in the United States is compared to that of students in other countries that participated in each assessment.

- At fourth grade, comparisons are made among students in the 25 countries that participated in TIMSS 2003, and in the 15 countries that participated in TIMSS 2003 and TIMSS 1995.
- At eighth grade, comparisons are made among students in the 45 countries that participated in TIMSS 2003, and in the 34 countries that participated in TIMSS 2003 and at least one earlier data collection, either TIMSS 1995 or 1999, or both.
- Results are presented first for mathematics and then for science at both grade levels.
- All estimates for the United States are based on the performance of students from both public and private schools, unless otherwise indicated.

All countries were required to draw random, nationally representative samples of students and schools. The U.S. fourth-grade sample achieved an initial school response rate of 70 percent (weighted); with a school response rate of 82 percent, after replacement schools were added. From the schools that agreed to participate, students were sampled in intact classes. A total of 10,795 fourth-grade students were sampled for the assessment and 9,829 participated, for a 95 percent student response rate. The resulting fourth-grade overall response rate, with replacements included, was 78 percent. The U.S. eighth-grade sample achieved an initial school response rate of 71 percent; with a school response rate of 78 percent, after replacement schools were added. A total of 9,891 students were sampled for the eighth-grade assessment and 8,912 completed the assessment, for a 94 percent student response rate. The resulting eighth-grade overall response rate, with replacements included, was 73 percent.

In addition to the assessments, students, their teachers, and principals were asked to complete questionnaires related to their school and learning experiences. At fourth grade, the assessment took approximately 72 minutes to complete. At eighth grade, the assessment took approximately 90 minutes. Detailed information on data collection, sampling, response rates, test development and design, weighting, and scaling is included in appendix A. Example items from the fourth- and eighth-grade assessments are included in appendix B.

Comparisons made in this report have been tested for statistical significance at the .05 level. Differences between averages or percentages that are statistically significant are discussed using comparative terms such as "higher" and "lower." Differences that are not statistically significant are either not discussed or referred to as "no measurable differences found" or "not statistically significant." In this latter case, failure to find a statistically significant difference should not be interpreted to mean that the estimates are the same or similar; rather, failure to find a difference may also be due to measurement error or sampling. In addition, because the results of tests of statistical significance are, in part, influenced by sample sizes, when subgroup comparisons are drawn within the United States, effect sizes are also included in order to provide the reader an increased understanding of the importance of the significant difference. These effect sizes use standard deviations, rather than standard errors, and thus are not influenced by the size of the subgroup samples. In social sciences, as used here, an effect size of .2 is considered small, one of .5 is of medium importance, and one of .8 or larger is considered large. Information on the technical aspects of the study can be found in appendix A, as well as in the *TIMSS 2003 Technical Report* (Martin, Mullis, and Chrostowski 2004).

Detailed tables with estimates and standard errors for all analyses included in this report are provided in appendix C. A list of TIMSS publications and resources published by NCES and the IEA is provided in appendix E.

Mathematics

How did U.S. fourth- and eighth-graders perform in mathematics in 2003?

Fourth Grade:

- In 2003, U.S. fourth-grade students scored 518, on average, in mathematics, exceeding the international average of 495 (table 2 and table C1 in appendix C). U.S. fourth-graders outperformed their peers in 13 of the other 24 participating countries, and performed lower than their peers in 11 countries.

- In comparison to students in the other 10 OECD-member countries participating in the fourth-grade TIMSS assessment, U.S. fourth-graders outperformed their peers in mathematics in 5 countries (Australia, Italy, New Zealand, Norway, and Scotland) and were outperformed by their peers in the other 5 countries (Belgium-Flemish, England, Hungary, Japan, and the Netherlands) (table 2).

Table 2. Average mathematics scale scores of fourth-grade students, by country: 2003

Country	Average score
International average	495
Singapore	594
Hong Kong SAR[1,2]	575
Japan	565
Chinese Taipei	564
Belgium-Flemish	551
Netherlands[2]	540
Latvia	536
Lithuania[3]	534
Russian Federation	532
England[2]	531
Hungary	529
United States[2]	**518**
Cyprus	510
Moldova, Republic of	504
Italy	503
Australia[2]	499
New Zealand	493
Scotland[2]	490
Slovenia	479
Armenia	456
Norway	451
Iran, Islamic Republic of	389
Philippines	358
Morocco	347
Tunisia	339

■ Average is higher than the U.S. average
□ Average is not measurably different from the U.S. average
■ Average is lower than the U.S. average

[1]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[2]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[3]National desired population does not cover all of the international desired population.
NOTE: Countries are ordered by 2003 average score. The test for significance between the United States and the international average was adjusted to account for the U.S. contribution to the international average. Countries were required to sample students in the upper of the two grades that contained the largest number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Table 3. Average mathematics scale scores of eighth-grade students, by country: 2003

Country	Average score
International average[1]	466
Singapore	605
Korea, Republic of	589
Hong Kong SAR[2,3]	586
Chinese Taipei	585
Japan	570
Belgium-Flemish	537
Netherlands[2]	536
Estonia	531
Hungary	529
Malaysia	508
Latvia	508
Russian Federation	508
Slovak Republic	508
Australia	505
(United States)	**504**
Lithuania[4]	502
Sweden	499
Scotland[2]	498
(Israel)	496
New Zealand	494
Slovenia	493
Italy	484
Armenia	478
Serbia[4]	477
Bulgaria	476
Romania	475
Norway	461
Moldova, Republic of	460
Cyprus	459
(Macedonia, Republic of)	435
Lebanon	433
Jordan	424
Iran, Islamic Republic of	411
Indonesia[4]	411
Tunisia	410
Egypt	406
Bahrain	401
Palestinian National Authority	390
Chile	387
(Morocco)	387
Philippines	378
Botswana	366
Saudi Arabia	332
Ghana	276
South Africa	264

Eighth Grade:

- In 2003, U.S. eighth-graders scored 504, on average, in mathematics. This average score exceeded the international average as well as the average scores of their peers in 25 of the 44 other participating countries (table 3 and table C2 in appendix C). U.S. eighth-graders were outperformed by students in nine countries: five Asian countries (Chinese Taipei, Hong Kong SAR, Japan, Korea, and Singapore) and four European countries (Belgium-Flemish, Estonia, Hungary, and the Netherlands).

- In comparison to their peers in the other 12 OECD-member countries participating in the eighth-grade TIMSS assessment, U.S. eighth-graders outperformed students in mathematics in 2 countries (Italy and Norway) and were outperformed by their peers in 5 countries (Belgium-Flemish, Hungary, Korea, Japan, and the Netherlands) (table 3).

■ Average is higher than the U.S. average
□ Average is not measurably different from the U.S. average
■ Average is lower than the U.S. average

[1]The international average reported here differs from that reported in Mullis et al. (2004) due to the deletion of England. In Mullis et al., the reported international average is 467.
[2]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[3]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[4]National desired population does not cover all of the international desired population.
NOTE: Countries are ordered by 2003 average score. The test for significance between the United States and the international average was adjusted to account for the U.S. contribution to the international average. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one country may be significant while a large difference between the United States and another country may not be significant. Parentheses indicate countries that did not meet international sampling or other guidelines in 2003. Countries were required to sample students in the upper of the two grades that contained the largest number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Did the mathematics performance of U.S. fourth- and eighth-graders change between 1995 and 2003?

Fourth Grade:

- In both 1995 and 2003, U.S. fourth-graders had an average score of 518 in mathematics (table 4 and table C3 in appendix C). Fourth-graders in six other countries also showed no measurable change in average mathematics performance over the same time period.

- In contrast, fourth-graders in 6 of the 15 participating countries showed an increase in average mathematics achievement scores between 1995 and 2003: Cyprus, England, Hong Kong SAR, Latvia-LSS,[2] New Zealand, and Slovenia (table 4). Fourth-graders in two countries—the Netherlands and Norway—experienced a decrease in average mathematics achievement scores over the same period of time.

Eighth Grade:

- U.S. eighth-graders showed significant improvement in average mathematics performance over the 8-year period between 1995 and 2003 (table 5 and table C4 in appendix C). In 1995, U.S. eighth-graders had an average score of 492. In 2003, U.S. eighth-graders improved their average mathematics score by 12 points, to 504. No measurable change was detected in the average U.S. mathematics performance between 1999 and 2003, thus indicating that the increase in average mathematics performance in the United States occurred primarily between 1995 and 1999.

- In addition to the United States, eighth-graders in six other countries improved their average mathematics performance between 1995 and 2003 or between 1999 and 2003: Hong Kong SAR, Israel, Korea, Latvia-LSS, Lithuania, and the Philippines (table 5).

- Eighth-graders in 11 countries showed significant declines in their average mathematics achievement between 1995 and 2003 or between 1999 and 2003: Belgium-Flemish, Bulgaria, Cyprus, Iran, Japan, Macedonia, Norway, Russian Federation, Slovak Republic, Sweden, and Tunisia. The remaining 16 countries showed no measurable difference in the average mathematics scores of their students (table 5).

Table 4. Differences in average mathematics scale scores of fourth-grade students, by country: 1995 and 2003

Country	1995	2003	Difference[1]
Singapore	590	594	4
Hong Kong SAR[2,3]	557	575	18 ▲
Japan	567	565	-3
(Netherlands)[3]	549	540	-9 ▼
(Latvia-LSS)[4]	499	533	34 ▲
England[3]	484	531	47 ▲
(Hungary)	521	529	7
United States[3]	**518**	**518**	**#**
Cyprus	475	510	35 ▲
(Australia)[3]	495	499	4
New Zealand[5]	469	496	26 ▲
Scotland[3]	493	490	-3
(Slovenia)	462	479	17 ▲
Norway	476	451	-25 ▼
Iran, Islamic Republic of	387	389	2

#Rounds to zero.
▲$p < .05$, denotes a significant increase.
▼$p < .05$, denotes a significant decrease.
[1]Difference calculated by subtracting 1995 from 2003 estimate using unrounded numbers.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[4]Designated LSS because only Latvian-speaking schools were included in 1995. For this analysis, only Latvian-speaking schools are included in the 2003 average.
[5]In 1995, Maori-speaking students did not participate. Estimates in this table are computed for students taught in English only, which represents between 98-99 percent of the student population in both years.
NOTE: Countries are ordered based on the 2003 average scores. Parentheses indicate countries that did not meet international sampling or other guidelines in 1995. All countries met international sampling and other guidelines in 2003, except as noted. See NCES (1997) for details regarding the 1995 data. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between averages for one country may be significant while a large difference for another country may not be significant. Countries were required to sample students in the upper of the two grades that contained the largest number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details. Detail may not sum to totals because of rounding.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2003.

[2]Designated LSS because only Latvian-speaking schools were included in 1995. For this analysis, only Latvian-speaking schools are included in the 2003 average.

Table 5. Differences in average mathematics scale scores of eighth-grade students, by country: 1995, 1999, and 2003

Country	1995	1999	2003	Difference[1] (2003-1995)	Difference[1] (2003-1999)
Singapore	609	604	605	-3	1
Korea, Republic of	581	587	589	8 ▲	2
Hong Kong SAR[2,3]	569	582	586	17 ▲	4
Chinese Taipei	—	585	585	†	#
Japan	581	579	570	-11 ▼	-9 ▼
Belgium-Flemish	550	558	537	-13 ▼	-21 ▼
(Netherlands)[2]	529	540	536	7	-4
Hungary	527	532	529	3	-2
Malaysia	—	519	508	†	-11
Russian Federation	524	526	508	-16 ▼	-18 ▼
Slovak Republic	534	534	508	-26 ▼	-26 ▼
(Latvia-LSS)[4]	488	505	505	17 ▲	#
(Australia)[5]	509	—	505	-4	†
(United States)	**492**	**502**	**504**	**12 ▲**	**3**
Lithuania[6]	472	482	502	30 ▲	20 ▲
Sweden	540	—	499	-41 ▼	†
(Scotland)[2]	493	—	498	4	†
(Israel)[7]	—	466	496	†	29 ▲
New Zealand	501	491	494	-7	3
(Slovenia)[5]	494	—	493	-2	†
Italy[7]	—	479	484	†	4
(Bulgaria)	527	511	476	-51 ▼	-34 ▼
(Romania)	474	472	475	2	3
Norway	498	—	461	-37 ▼	†
Moldova, Republic of	—	469	460	†	-9
Cyprus	468	476	459	-8 ▼	-17 ▼
(Macedonia, Republic of)	—	447	435	†	-12 ▼
Jordan	—	428	424	†	-3
Iran, Islamic Republic of	418	422	411	-7	-11 ▼
Indonesia[6]	—	403	411	†	8
Tunisia	—	448	410	†	-38 ▼
Chile	—	392	387	†	-6
Philippines	—	345	378	†	33 ▲
South Africa[8]	—	275	264	†	-11

—Not available.
†Not applicable.
#Rounds to zero.
▲p<.05, denotes a significant increase.
▼p<.05, denotes a significant decrease.
[1]Difference calculated by subtracting 1995 or 1999 from 2003 estimate using unrounded numbers.
[2]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[3]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[4]Designated LSS because only Latvian-speaking schools were included in 1995 and 1999. For this analysis, only Latvian-speaking schools are included in the 2003 average.
[5]Because of national-level changes in the starting age/date for school, 1999 data for Australia and Slovenia cannot be compared to 2003.
[6]National desired population does not cover all of the international desired population in all years for Lithuania, and in 2003 for Indonesia.
[7]Because of changes in the population tested, 1995 data for Israel and Italy are not shown.
[8]Because within classroom sampling was not accounted for, 1995 data are not shown for South Africa.
NOTE: Countries are sorted by 2003 average scores. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between averages for one country may be significant while a large difference for another country may not be significant. Parentheses indicate countries that did not meet international sampling or other guidelines in 1995, 1999, or 2003. See appendix A for details regarding 2003 data. See Gonzales et al. (2000) for details regarding 1995 and 1999 data. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries this corresponds to grade 8. See table A1 in appendix A for details. Detail may not sum to totals because of rounding.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995, 1999, and 2003.

Has the relative mathematics performance of U.S. fourth- and eighth-grade students changed since 1995?

Fourth Grade:

- Although the average mathematics score of U.S. fourth-graders was 518 in both 1995 and 2003 (table 4), the data suggest that the standing of U.S. fourth-graders relative to their peers in 14 other countries was lower in 2003 than in 1995 in mathematics (table 6 and table C3 in appendix C). In 1995, on average, U.S. fourth-graders were outperformed in mathematics by fourth-graders in 4 of these countries and outperformed fourth-graders in 9 of these countries. In 2003, on average, U.S. fourth-graders were outperformed by fourth-graders in seven of these countries and outperformed fourth-graders in 7 of these countries.

Table 6. Average mathematics scale scores of fourth-grade students, by country: 1995 and 2003

Country	1995	Country	2003
Singapore	590	Singapore	594
Japan	567	Hong Kong SAR[1, 2]	575
Hong Kong SAR[1,2]	557	Japan	565
(Netherlands)	549	Netherlands[1]	540
(Hungary)	521	Latvia-LSS[3]	533
United States	**518**	England[1]	531
(Latvia-LSS)[3]	499	Hungary	529
(Australia)	495	**United States[1]**	**518**
Scotland	493	Cyprus	510
England	484	Australia[1]	499
Norway	476	New Zealand[4]	496
Cyprus	475	Scotland[1]	490
New Zealand[4]	469	Slovenia	479
(Slovenia)	462	Norway	451
Iran, Islamic Republic of	387	Iran, Islamic Republic of	389

■ Average is higher than the U.S. average
□ Average is not measurably different from the U.S. average
■ Average is lower than the U.S. average

[1]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]Designated LSS because only Latvian-speaking schools were included in 1995. For this analysis, only Latvian-speaking schools are included in the 2003 average.
[4]In 1995, Maori-speaking students did not participate. Estimates in this table are computed for students taught in English only, which represents between 98-99 percent of the student population in both years.
NOTE: Countries are ordered based on the average score. Parentheses indicate countries that did not meet international sampling or other guidelines in 1995. All countries met international sampling and other guidelines in 2003, except as noted. See NCES (1997) for details regarding 1995 data. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one country may be significant while a large difference between the United States and another country may not be significant. Countries were required to sample students in the upper of the two grades that contained the most number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2003.

Eighth Grade:

- The average mathematics score for U.S. eighth-graders increased from 492 in 1995 to 504 in 2003. Over the same period, several countries whose eighth-graders outperformed the U.S. eighth-graders in 1995 experienced decreases in their average scores. The net effect resulted in a higher relative standing for U.S. eighth-graders in 2003 (table 7 and table C5 in appendix C). Among this group of countries, U.S. eighth-graders were outperformed by eighth-graders in 12 countries, on average, and outperformed eighth-graders in 4 countries in mathematics in 1995. In 2003, U.S. eighth-graders were outperformed by eighth-graders in seven of these countries, on average, and outperformed eighth-graders in six of these countries in mathematics.

Table 7. Average mathematics scale scores of eighth-grade students, by country: 1995 and 2003

Country	1995	Country	2003
Singapore	609	Singapore	605
Japan	581	Korea, Republic of	589
Korea, Republic of	581	Hong Kong SAR[1,2]	586
Hong Kong SAR[1]	569	Japan	570
Belgium-Flemish	550	Belgium-Flemish	537
Sweden	540	Netherlands[2]	536
Slovak Republic	534	Hungary	529
(Netherlands)	529	Russian Federation	508
Hungary	527	Slovak Republic	508
(Bulgaria)	527	Latvia-LSS[3]	505
Russian Federation	524	Australia	505
(Australia)	509	**(United States)**	**504**
New Zealand	501	Lithuania[4]	502
Norway	498	Sweden	499
(Slovenia)	494	Scotland[2]	498
(Scotland)	493	New Zealand	494
United States	**492**	Slovenia	493
(Latvia-LSS)[3]	488	Bulgaria	476
(Romania)	474	Romania	475
Lithuania[4]	472	Norway	461
Cyprus	468	Cyprus	459
Iran, Islamic Republic of	418	Iran, Islamic Republic of	411

■ Average is higher than the U.S. average
□ Average is not measurably different from the U.S. average
■ Average is lower than the U.S. average

[1]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[2]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[3]Designated LSS because only Latvian-speaking schools were included in 1995. For this analysis, only Latvian-speaking schools are included in the 2003 average.
[4]National desired population does not cover all of the international desired population.
NOTE: Countries are ordered by average score. Parentheses indicate countries that did not meet international sampling or other guidelines in 1995 or 2003. See appendix A for details regarding 2003 data. See NCES (1997) for details regarding 1995 data. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one country may be significant while a large difference between the United States and another country may not be significant. Countries were required to sample students in the upper of the two grades that contained the largest number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2003.

Did the performance of U.S. fourth- and eighth-graders in the mathematics content areas change between 1995 and 2003?

Fourth Grade:

- Changes in performance between 1995 and 2003 on the five mathematics content areas measured in TIMSS (Number; Patterns, Equations, and Relationships; Measurement; Geometry; and Data) could not be calculated due to a limited number of items in common between the two assessments.

Eighth Grade:

- Between 1999 and 2003, U.S. eighth-graders showed significant improvement in correctly answering items associated with two of the five content areas: Algebra (i.e., patterns, equations, and relationships) and Data.[3] There were no measurable differences detected in the average percentage of U.S. students who correctly answered items in Geometry, Measurement, and Number between 1999 and 2003 (table C6 in appendix C).
- The United States was among 17 countries that showed significant change—either an increase or decrease—in the average percentage of eighth-grade students who were able to correctly respond to items in at least one of the five eighth-grade mathematics content areas between 1999 and 2003 (table C6 in appendix C).

Did the mathematics performance of U.S. population groups change between 1995 and 2003?

Fourth Grade:

- No measurable change was detected in the average mathematics achievement of U.S. fourth-grade boys or girls between 1995 and 2003 (figure 1). Nonetheless, U.S. boys outperformed girls in mathematics in 2003, which differs from 1995 when no measurable difference was detected.[4,5]
- Fourth-grade boys and girls in 6 of the 14 other countries showed an improvement in average mathematics achievement: Cyprus, England, Hong Kong SAR, Latvia-LSS, New Zealand, and Slovenia (table C7 in appendix C).
- Black fourth-grade students in the United States demonstrated an improvement in average mathematics achievement between 1995 and 2003 (figure 1 and table C8 in appendix C). In 1995, U.S. Black fourth-graders scored 457 in mathematics, on average, compared to 472 in 2003. As a result, over these 8 years, the gap in average scores between White and Black fourth-grade students narrowed, from 84 score points in 1995 to 69 score points in 2003.[6] White and Hispanic fourth-graders showed no measurable change in their average mathematics scores over this time period.
- In 2003, U.S. fourth-graders in U.S. public schools with the highest poverty level (75 percent or more of students eligible for free or reduced-price lunch) had lower average mathematics scores compared to their counterparts in public schools with lower poverty levels (figure 1). Fourth-graders in public schools with the lowest poverty level (10 percent or less eligible students) had higher average mathematics scores than students in schools with higher poverty levels. The difference in the average mathematics scores of students in schools with the lowest and highest poverty levels was 96 score points in 2003.[7]

[3]Although many of the participating countries collected data in all three years, analyses of changes in the mathematics content areas at eighth grade are limited to 1999 and 2003 due to the limited number of in common items from year to year.

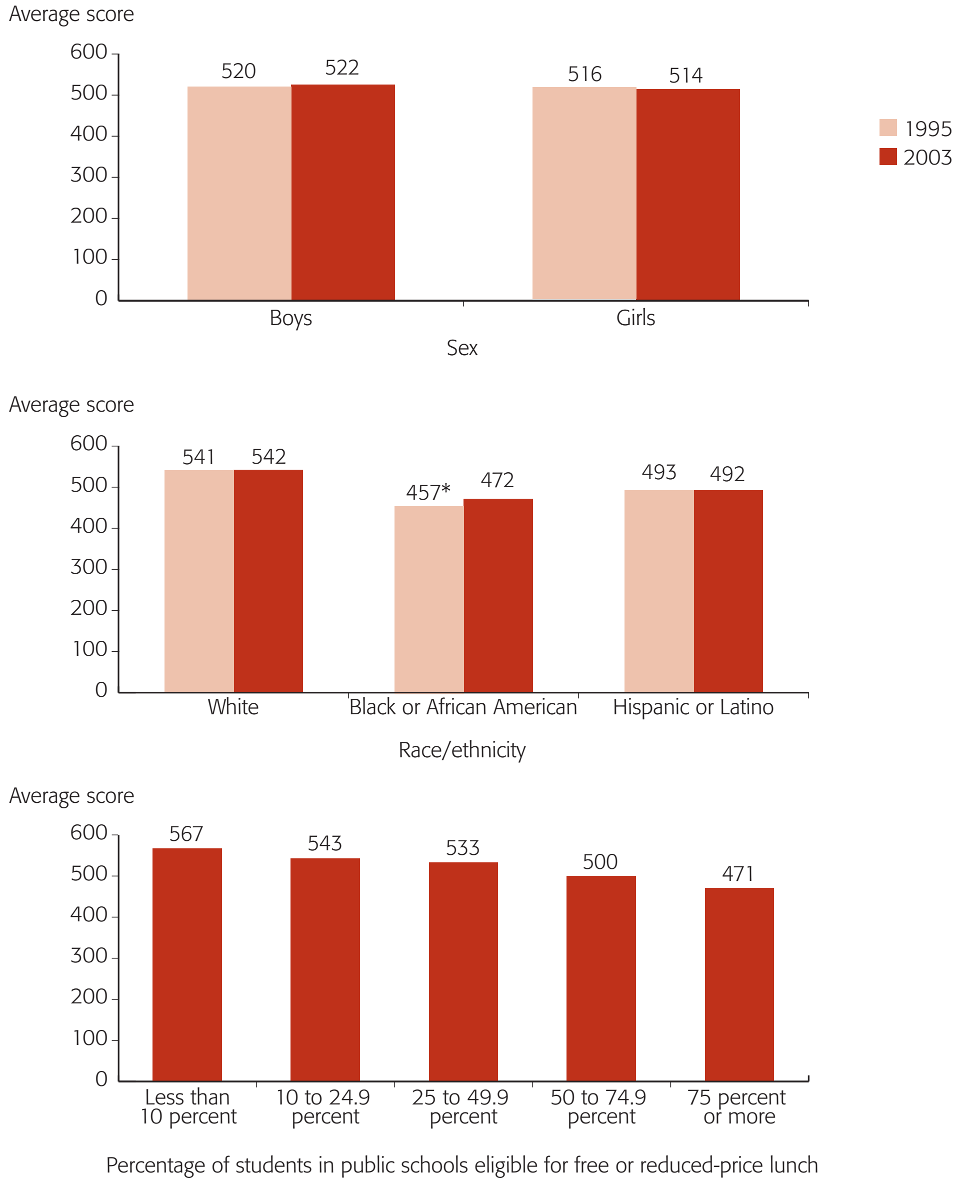
[4]The effect size of the difference between two means can be calculated by dividing the raw difference in means by the pooled standard deviation of the comparison groups (appendix A for an explanation). The effect size of the difference in mathematics achievement between U.S. boys and girls in 2003 is .11 (table C21 in appendix C for standard deviations of U.S. student population groups).

[5]See NCES (1997) for details on U.S. fourth-grade results for TIMSS 1995.

[6]The effect sizes of the differences in mathematics achievement between White and Black and between White and Hispanic fourth-graders in 2003 are 1.07 and .73, respectively (table C21 in appendix C for standard deviations of U.S. student population groups).

[7]The effect size of the difference in mathematics achievement between fourth-graders in public schools with the lowest and highest levels of poverty in 2003 is 1.55 (table C21 in appendix C for standard deviations of U.S. student population groups).

Figure 1. Average mathematics scale scores of U.S. fourth-grade students, by sex, race/ethnicity, and poverty level: 1995 and 2003



*p<.05, denotes a significant difference from 2003 average score.
NOTE: Reporting standards not met for Asian category in 1995 and American Indian or Alaska Native and Native Hawaiian or Other Pacific Islander for both years. Racial categories exclude Hispanic origin. Other races/ethnicities are included in U.S. totals shown throughout the report. Analyses by poverty level are limited to students in public schools only. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between averages for one student group may be significant while a large difference for another student group may not be significant. The United States met international guidelines for participation rates in 2003 only after replacement schools were included. See appendix A for more information.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2003.

Eighth Grade:

- In 2003, U.S. eighth-grade boys and girls both showed improvement in mathematics compared to 1995 (figure 2 and table C10 in appendix C). U.S. eighth-grade boys outperformed girls in 2003.[8] In 2003, U.S. eighth-grade boys' average score in mathematics was 507. This is 12 score points higher than in 1995, when U.S. boys scored 495. U.S. girls' average mathematics score was 502 in 2003. This is also 12 score points higher than in 1995, when U.S. girls scored 490.

- The United States is one of four countries in which both eighth-grade boys and girls improved their average mathematics performance in 2003 over previous assessment years (table C10 in appendix C). In addition to the United States, both eighth-grade boys and girls improved their average mathematics performance in Israel, Lithuania, and the Philippines.

- Both Black and Hispanic eighth-grade students in the United States demonstrated improvement in mathematics achievement between 1995 and 2003 (figure 2 and table C11 in appendix C). In 1995, U.S. Black eighth-grade students scored 419 in mathematics, on average. This improved to 448, on average, in 2003. Likewise, in 1995, U.S. Hispanic eighth-grade students scored 443 in mathematics, on average, improving to an average score of 465 in 2003.

- As a result of the improvement in the average mathematics achievement of Black eighth-grade students between 1995 and 2003, the gap in average scores between White and Black eighth-grade students narrowed, from 97 score points in 1995 to 77 score points in 2003 (figure 2).[9] Although Hispanic eighth-grade students showed improvement in their mathematics performance between 1995 and 2003, there was no measurable change detected in the gap in average scores between White and Hispanic eighth-grade students.

- In 2003, U.S. eighth-graders in U.S. public schools with the highest poverty level (75 percent or more of students eligible for free or reduced-price lunch) had lower average mathematics scores compared to their counterparts in public schools with lower poverty levels (figure 2). In contrast, students in schools with the lowest level (10 percent or less eligible students) had higher average mathematics scores than students in schools with poverty levels of 25 percent or more eligible. The difference in the average mathematics scores of students in schools with the lowest and highest poverty levels was 103 score points in 2003.[10]

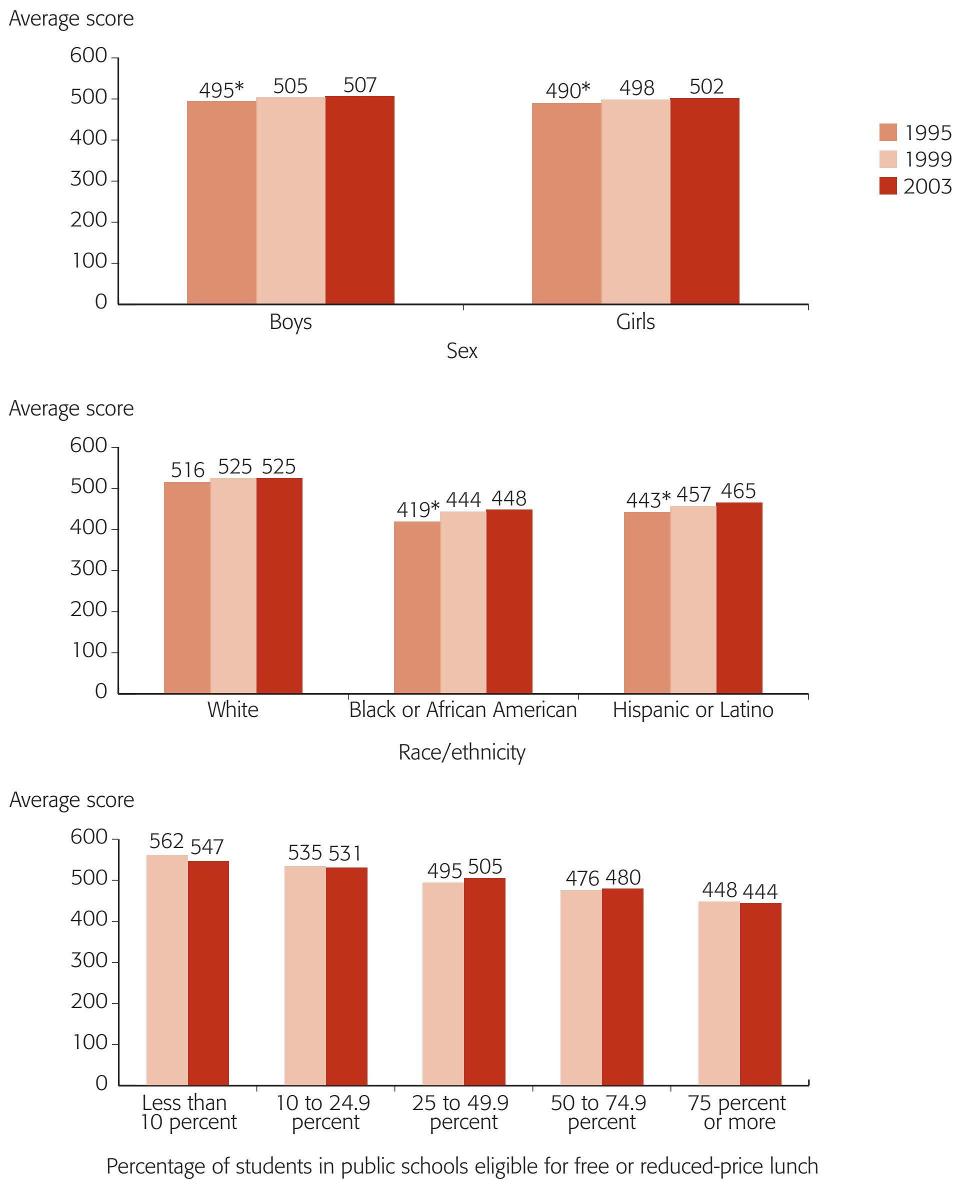
- As was the case in the aggregate, results by poverty level showed no measurable change in average mathematics achievement between 1999 and 2003, the two years for which data are available (figure 2 and table C11 in appendix C).

[8]The effect size of the difference in mathematics achievement between U.S. eighth-grade boys and girls in 2003 is .07 (table C21 in appendix C for standard deviations of U.S. student population groups).

[9]The effect sizes of the differences in mathematics achievement between White and Black and between White and Hispanic eighth-graders in 2003 are 1.11 and .83, respectively (table C21 in appendix C for standard deviations of U.S. student population groups).

[10]The effect size of the difference in mathematics achievement between eighth-graders in public schools with the lowest and highest levels of poverty in 2003 is 1.57 (table C21 in appendix C for standard deviations of U.S. student population groups).

Figure 2. Average mathematics scale scores of U.S. eighth-grade students, by sex, race/ethnicity, and poverty level: 1995, 1999, and 2003



*p<.05, denotes a significant difference from 2003 average score.
NOTE: Reporting standards not met for Asian category in 1995 or 1999. Reporting standards not met for American Indian or Alaska Native and Native Hawaiian or Other Pacific Islander in 1995, 1999, and 2003. Racial categories exclude Hispanic origin. Other races/ethnicities are included in U.S. totals shown throughout the report. Analyses by poverty level are limited to students in public schools only. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between averages for one student group may be significant while a large difference for another student group may not be significant. The United States met international guidelines for participation rates in 2003 only after replacement schools were included. See appendix A for more information.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2003.

Science

How did U.S. fourth- and eighth-graders perform in science in 2003?

Fourth Grade:

- In 2003, fourth-graders in the United States scored 536, on average, on the TIMSS science assessment, which was higher than the international average of 489 (table 8 and table C1 in appendix C). Of the 24 other participating countries, fourth-graders in 16 countries demonstrated lower science scores, on average, than fourth-graders in the United States, while students in 3 countries—Chinese Taipei, Japan, and Singapore—outperformed their peers in the United States.

- In comparison to the other 10 OECD-member countries in science, U.S. fourth-grade students outperformed their peers in seven countries in 2003 (Australia, Belgium-Flemish, Italy, the Netherlands, New Zealand, Norway, and Scotland; table 8). Japanese fourth-grade students were the only group of students to outperform U.S. fourth-grade students among the participating OECD-member countries.

Table 8. Average science scale scores of fourth-grade students, by country: 2003

Country	Average score
International average	489
Singapore	565
Chinese Taipei	551
Japan	543
Hong Kong SAR[1,2]	542
England[2]	540
United States[2]	**536**
Latvia	532
Hungary	530
Russian Federation	526
Netherlands[2]	525
Australia[2]	521
New Zealand	520
Belgium-Flemish	518
Italy	516
Lithuania[3]	512
Scotland[2]	502
Moldova, Republic of	496
Slovenia	490
Cyprus	480
Norway	466
Armenia	437
Iran, Islamic Republic of	414
Philippines	332
Tunisia	314
Morocco	304

■ Average is higher than the U.S. average
□ Average is not measurably different from the U.S. average
▪ Average is lower than the U.S. average

[1]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[2]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[3]National desired population does not cover all of the international desired population.
NOTE: The test for significance between the United States and the international average was adjusted to account for the U.S. contribution to the international average. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one country may be significant while a large difference between the United States and another country may not be significant. Countries were required to sample students in the upper of the two grades that contained the largest number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Table 9. Average science scale scores of eighth-grade students, by country: 2003

Country	Average score
International average[1]	473
Singapore	578
Chinese Taipei	571
Korea, Republic of	558
Hong Kong SAR[2,3]	556
Estonia	552
Japan	552
Hungary	543
Netherlands[2]	536
(United States)	**527**
Australia	527
Sweden	524
Slovenia	520
New Zealand	520
Lithuania[4]	519
Slovak Republic	517
Belgium-Flemish	516
Russian Federation	514
Latvia	512
Scotland[2]	512
Malaysia	510
Norway	494
Italy	491
(Israel)	488
Bulgaria	479
Jordan	475
Moldova, Republic of	472
Romania	470
Serbia[4]	468
Armenia	461
Iran, Islamic Republic of	453
(Macedonia, Republic of)	449
Cyprus	441
Bahrain	438
Palestinian National Authority	435
Egypt	421
Indonesia[4]	420
Chile	413
Tunisia	404
Saudi Arabia	398
(Morocco)	396
Lebanon	393
Philippines	377
Botswana	365
Ghana	255
South Africa	244

Eighth Grade:

- In science, U.S. eighth-graders exceeded the international average and outperformed their peers in 32 of the 44 other participating countries (table 9 and table C2 in appendix C). U.S. eighth-graders performed lower, on average, than their peers in seven countries and were not found to perform measurably different from students in five countries.
- An examination of the performance of students from the other 12 OECD-member countries shows that U.S. eighth-grade students outperformed their peers in science in 5 of the countries (Belgium-Flemish, Italy, Norway, Scotland, and the Slovak Republic) and were outperformed by their peers in 3 of the countries (Hungary, Japan, and Korea; table 9).

■ Average is higher than the U.S. average
□ Average is not measurably different from the U.S. average
▪ Average is lower than the U.S. average

[1]The international average reported here differs from that reported in Martin et al. (2004) due to the deletion of England. In Martin et al., the reported international average is 474.
[2]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[3]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[4]National desired population does not cover all of the international desired population.
NOTE: Countries are ordered by 2003 average score. The test for significance between the United States and the international average was adjusted to account for the U.S. contribution to the international average. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one country may be significant while a large difference between the United States and another country may not be significant. Parentheses indicate countries that did not meet international sampling or other guidelines in 2003. Countries were required to sample students in the upper of the two grades that contained the largest number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Did the science performance of U.S. fourth- and eighth-graders change between 1995 and 2003?

Fourth Grade:

- There was no measurable difference detected in the average science performance of U.S. fourth-graders between 1995 and 2003 (table 10 and table C13 in appendix C). Fourth-graders in two other countries also showed no measurable change in science performance over the same time period.
- Fourth-graders in 9 of the 15 participating countries showed an increase in average science achievement scores between 1995 and 2003: Cyprus, England, Hong Kong SAR, Hungary, Iran, Latvia-LSS, New Zealand, Singapore, and Slovenia (table 10). Fourth-graders in three countries—Japan, Norway, and Scotland—experienced a decrease in average science achievement scores over the same period.

Eighth Grade:

- In 2003, U.S. eighth-graders improved in science compared to 1995 and 1999. U.S. eighth-graders scored 527, on average, in science in 2003, which was 12 score points higher than in 1999 and 14 score points higher than in 1995 (table 11 and table C14 in appendix C). The data indicate that the increase in average science performance in the United States occurred primarily between 1999 and 2003.
- Eighth-graders in 11 other countries demonstrated a significant increase in their average science achievement between 1995 and 2003 or between 1999 and 2003: Australia, Hong Kong SAR, Israel, Jordan, Korea, Latvia-LSS, Lithuania, Malaysia, Moldova, the Philippines, and Slovenia (table 11).
- Eighth-graders in 11 countries showed significant declines in their average science achievement between 1995 and 2003 or between 1999 and 2003 (table 11). The remaining 11 countries showed no measurable difference in the average mathematics scores of their students between 1995 and 2003, or between 1999 and 2003 (table 11).

Table 10. Differences in average science scale scores of fourth-grade students, by country: 1995 and 2003

Country	1995	2003	Difference[1]
Singapore	523	565	42 ▲
Japan	553	543	-10 ▼
Hong Kong SAR[2,3]	508	542	35 ▲
England[3]	528	540	13 ▲
United States[3]	**542**	**536**	**-6**
(Hungary)	508	530	22 ▲
(Latvia-LSS)[4]	486	530	43 ▲
(Netherlands)[3]	530	525	-5
New Zealand[5]	505	523	18 ▲
(Australia)[3]	521	521	-1
Scotland[2]	514	502	-12 ▼
(Slovenia)	464	490	26 ▲
Cyprus	450	480	30 ▲
Norway	504	466	-38 ▼
Iran, Islamic Republic of	380	414	34 ▲

▲p<.05, denotes a significant increase.
▼p<.05, denotes a significant decrease.
[1]Difference calculated by subtracting 1995 from 2003 estimate using unrounded numbers.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]Met international guidelines for participation rates only after replacement schools were included.
[4]Designated LSS because only Latvian-speaking schools were included in 1995. For this analysis, only Latvian-speaking schools are included in the 2003 average.
[5]In 1995, Maori-speaking students did not participate. Estimates in this table are computed for students taught in English only, which represents between 98-99 percent of the student population in both years.
NOTE: Countries are ordered based on the 2003 average scores. Parentheses indicate countries that did not meet international sampling or other guidelines in 1995. All countries met international sampling and other guidelines in 2003, except as noted. See NCES (1997) for details regarding 1995 data. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between averages for one country may be significant while a large difference for another country may not be significant. Countries were required to sample students in the upper of the two grades that contained the largest number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details. Detail may not sum to totals because of rounding.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2003.

Table 11. Differences in average science scale scores of eighth-grade students, by country: 1995, 1999, and 2003

Country	1995	1999	2003	Difference[1] (2003-1995)	(2003-1999)
Singapore	580	568	578	-3	10
Chinese Taipei	—	569	571	†	2
Korea, Republic of	546	549	558	13 ▲	10 ▲
Hong Kong SAR[2,3]	510	530	556	46 ▲	27 ▲
Japan	554	550	552	-2	3
Hungary	537	552	543	6	-10 ▼
(Netherlands)[2]	541	545	536	-6	-9
(United States)	**513**	**515**	**527**	**15 ▲**	**12 ▲**
(Australia)[4]	514	—	527	13 ▲	†
Sweden	553	—	524	-28 ▼	†
(Slovenia)[4]	514	—	520	7 ▲	†
New Zealand	511	510	520	9	10
(Lithuania)[5]	464	488	519	56 ▲	31 ▲
Slovak Republic	532	535	517	-15 ▼	-18 ▼
Belgium-Flemish	533	535	516	-17 ▼	-19 ▼
Russian Federation	523	529	514	-9	-16 ▼
(Latvia-LSS)[6]	476	503	513	37 ▲	11
(Scotland)[2]	501	—	512	10	†
Malaysia	—	492	510	†	18 ▲
Norway	514	—	494	-21 ▼	†
Italy[7]	—	493	491	†	-2
(Israel)[7]	—	468	488	†	20 ▲
(Bulgaria)	545	518	479	-66 ▼	-39 ▼
Jordan	—	450	475	†	25 ▲
Moldova, Republic of	—	459	472	†	13 ▲
(Romania)	471	472	470	-1	-2
Iran, Islamic Republic of	463	448	453	-9 ▼	5
(Macedonia, Republic of)	—	458	449	†	-9
Cyprus	452	460	441	-11 ▼	-19 ▼
Indonesia[5]	—	435	420	†	-15 ▼
Chile	—	420	413	†	-8
Tunisia	—	430	404	†	-26 ▼
Philippines	—	345	377	†	32 ▲
South Africa[8]	—	243	244	†	1

—Not available.
†Not applicable.
▲p<.05, denotes a significant increase.
▼p<.05, denotes a significant decrease.
[1]Difference calculated by subtracting 1995 or 1999 from 2003 estimate using unrounded numbers.
[2]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[3]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[4]Because of national-level changes in the starting age/date for school, 1999 data for Australia and Slovenia cannot be compared to 2003.
[5]National desired population does not cover all of the international desired population in all years for Lithuania, and in 2003 for Indonesia.
[6]Designated LSS because only Latvian-speaking schools were included in 1995 and 1999. For this analysis, only Latvian-speaking schools are included in the 2003 average.
[7]Because of changes in the population tested, 1995 data for Israel and Italy are not shown.
[8]Because within classroom sampling was not accounted for, 1995 data are not shown for South Africa.
NOTE: Countries are sorted by 2003 average scores. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between averages for one country may be significant while a large difference for another country may not be significant. Parentheses indicate countries that did not meet international sampling and/or other guidelines in 1995, 1999, and/or 2003. See appendix A for details regarding 2003 data. See Gonzales et al. (2000) for details regarding 1995 and 1999 data. Countries were required to sample students in the upper of the two grades that contained the largest number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details. Detail may not sum to totals because of rounding.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995, 1999, and 2003.

Has the relative science performance of U.S. fourth- and eighth-grade students changed since 1995?

Fourth Grade:

- The available data suggest that, as in mathematics, though there was no measurable difference detected in the average science performance of U.S. fourth-graders between 1995 and 2003 (table 10), the standing of U.S. fourth-graders in science relative to their peers in 14 other countries appears lower in 2003 than in 1995 (table 12 and table C13 in appendix C). In 1995, fourth-graders in one country, Japan, outperformed U.S. fourth-graders in science, while U.S. fourth-graders outperformed students in 13 countries. In 2003, U.S. fourth-graders were outperformed by students in two countries, on average, had average scores that were not measurably different from those of fourth-graders in four other countries, and outperformed students in eight countries.

Table 12. Average science scale scores of fourth-grade students, by country: 1995 and 2003

Country	1995	Country	2003
Japan	553	Singapore	565
United States	**542**	Japan	543
(Netherlands)	530	Hong Kong SAR[1,2]	542
England	528	England[1]	540
Singapore	523	**United States[1]**	**536**
(Australia)	521	Hungary	530
Scotland	514	Latvia-LSS[3]	530
Hong Kong SAR[2]	508	Netherlands[1]	525
(Hungary)	508	New Zealand[4]	523
New Zealand[4]	505	Australia[1]	521
Norway	504	Scotland[1]	502
(Latvia-LSS)[3]	486	Slovenia	490
(Slovenia)	464	Cyprus	480
Cyprus	450	Norway	466
Iran, Islamic Republic of	380	Iran, Islamic Republic of	414

■ Average is higher than the U.S. average
□ Average is not measurably different from the U.S. average
▪ Average is lower than the U.S. average

[1]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]Designated LSS because only Latvian-speaking schools were included in 1995. For this analysis, only Latvian-speaking schools are included in the 2003 average.
[4]In 1995, Maori-speaking students did not participate. Estimates in this table are computed for students taught in English only, which represents between 98-99 percent of the student population in both years.
NOTE: Countries are ordered based on the average score. Parentheses indicate countries that did not meet international sampling or other guidelines in 1995. All countries met international sampling and other guidelines in 2003, except as noted. See NCES (1997) for details for 1995 data. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one country may be significant while a large difference between the United States and another country may not be significant. Countries were required to sample students in the upper of the two grades that contained the most number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2003.

Eighth Grade:

- As was observed for mathematics, the available data show that not only did U.S. eighth-graders show significant improvement in science between 1995 and 2003 (table 11), but the relative standing of U.S. students also improved in science relative to students in the 21 other countries with data from 1995 and 2003 (table 13 and table C15 in appendix C). In 1995, U.S. eighth-graders were outperformed in science by eighth-graders in nine of these countries, and outperformed eighth-graders in five of these countries. In 2003, U.S. eighth-graders were outperformed by students in 5 of these countries, and outperformed students in 11 of these countries.

Table 13. Average science scales scores of eighth-grade students, by country: 1995 and 2003

Country	1995	Country	2003
Singapore	580	Singapore	578
Japan	554	Korea, Republic of	558
Sweden	553	Hong Kong SAR[1,2]	556
Korea, Republic of	546	Japan	552
(Bulgaria)	545	Hungary	543
(Netherlands)	541	Netherlands[2]	536
Hungary	537	**(United States)**	**527**
Belgium-Flemish	533	Australia	527
Slovak Republic	532	Sweden	524
Russian Federation	523	Slovenia	520
Norway	514	New Zealand	520
(Australia)	514	Lithuania[3]	519
(Slovenia)	514	Slovak Republic	517
United States	**513**	Belgium-Flemish	516
New Zealand	511	Russian Federation	514
Hong Kong SAR[1]	510	Latvia-LSS[4]	513
(Scotland)	501	Scotland[2]	512
(Latvia-LSS)[4]	476	Norway	494
(Romania)	471	Bulgaria	479
Lithuania[3]	464	Romania	470
Iran, Islamic Republic of	463	Iran, Islamic Republic of	453
Cyprus	452	Cyprus	441

■ Average is higher than the U.S. average
□ Average is not measurably different from the U.S. average
■ Average is lower than the U.S. average

[1]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[2]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[3]National desired population does not cover all of the international desired population.
[4]Designated LSS because only Latvian-speaking schools were included in 1995. For this analysis, only Latvian-speaking schools are included in the 2003 average.
NOTE: Countries are ordered by average score. Parentheses indicate countries that did not meet international sampling or other guidelines in 1995 or 2003. See appendix A for details regarding 2003 data. See NCES (1997) for details regarding 1995 data. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one country may be significant while a large difference between the United States and another country may not be significant. Countries were required to sample students in the upper of the two grades that contained the largest number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2003.

Did the performance of U.S. fourth- and eighth-graders in the science content areas change between 1995 and 2003?

Fourth Grade:

- Changes in average performance between 1995 and 2003 on the three science content areas measured in TIMSS in the fourth grade (Life Science; Physical Science; and Earth Science) could not be calculated due to a limited number of items in common between the two assessments.

Eighth Grade:

- Between 1999 and 2003, there was an increase in the average percentage of U.S. eighth-graders who correctly answered items in two of the five eighth-grade content areas in science: Earth Science and Physics (table C16 in appendix C).[11] There were no measurable differences detected in the average percentage of U.S. eighth-graders who correctly answered items in Chemistry, Environmental Science, and Life Science between 1999 and 2003.

Did the science performance of U.S. population groups change between 1995 and 2003?

Fourth Grade:

- The United States is one of four countries in which fourth-grade boys turned in a lower average science performance in 2003 than in 1995 (figure 3 and table C17 in appendix C). U.S. fourth-grade girls showed no measurable change in their average science performance. Fourth-grade girls in three countries showed a decline in their average science performance.
- As a result of the lower performance of U.S. boys in science, the gap in the average science achievement of U.S. fourth-grade boys and girls narrowed between 1995 and 2003, from 12 points in 1995 to 5 points in 2003 (figure 3).[12] Nonetheless, on average, U.S. boys outperformed girls in science in 2003, which was the case in 1995 as well.[13]
- As observed for mathematics, Black fourth-grade students in the United States showed improvement in their average science performance, scoring 487 in 2003 compared to 462 in 1995 (figure 3 and table C18 in appendix C).
- White fourth-grade students in the United States demonstrated a decline in average science performance during the same period (figure 3). U.S. White fourth-grade students scored 572, on average, in science in 1995, declining to an average of 565 in 2003. No measurable change was detected in the average science performance of U.S. Hispanic fourth-graders.
- As a result of significant changes in the average science scores of White and Black fourth-grade students, the average achievement gap between White and Black fourth-grade students narrowed from 110 score points in 1995 to 78 score points in 2003 (figure 3). Moreover, the gap in science achievement between Black and Hispanic fourth-graders also narrowed, from 41 score points in 1995 to 12 score points in 2003. There was no measurable difference in the score gap between White and Hispanic fourth-grade students over the same period of time.[14]
- In 2003, U.S. fourth-graders in U.S. public schools with the highest poverty level (75 percent or more of students eligible for free or reduced-price lunch) had lower average science scores compared to their counterparts in public schools with lower levels (figure 3). Fourth-graders in public schools with the lowest poverty level (10 percent or less eligible students) had higher average science scores than students in schools with poverty levels of 25 percent or more. The difference in the average science scores of students in schools with the lowest and highest poverty levels was 99 score points in 2003.[15]

[11]Although many of the participating countries collected data in all three years, analyses of changes in the science content areas at eighth grade are limited to 1999 and 2003 due to the limited number of in common items from year to year.

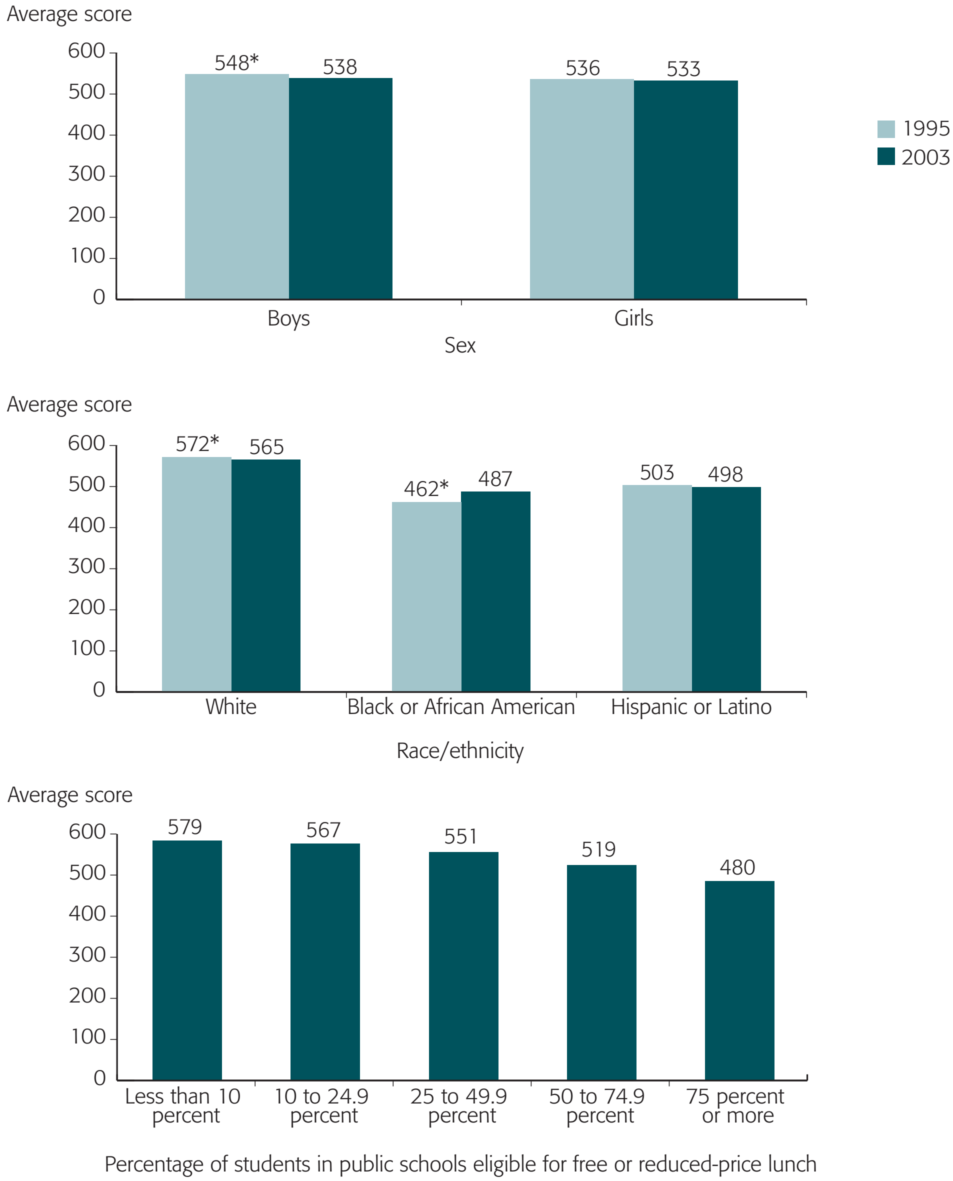
[12]The effect size of the difference in science achievement between U.S. fourth-grade boys and girls in 2003 is .07 (table C21 in appendix C for standard deviations of U.S. student population groups).

[13]See NCES (1997) for details on U.S. fourth-grade results for TIMSS 1995.

[14]The effect size of the differences between the average science scores of White and Black, and between White and Hispanic fourth-graders in the United States in 2003 are 1.15 and .94, respectively (table C21 in appendix C for standard deviations of U.S. student population groups).

[15]The effect size of the difference in science achievement between U.S. fourth-grade students in public schools with the lowest and highest levels of poverty in 2003 is 1.51 (table C21 in appendix C for standard deviations of U.S. student population groups).

Figure 3. Average science scale scores of U.S. fourth-grade students, by sex, race/ethnicity, and poverty level: 1995 and 2003



*p<.05, denotes a significant difference from 2003 average score.
NOTE: Reporting standards not met for Asian category in 1995 and American Indian or Alaska Native and Native Hawaiian or Other Pacific Islander for both years. Racial categories exclude Hispanic origin. Other races/ethnicities are included in U.S. totals shown throughout the report. Analyses by poverty level are limited to students in public schools only. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between averages for one student group may be significant while a large difference for another student group may not be significant. The United States met international guidelines for participation rates in 2003 only after replacement schools were included. See appendix A for more information.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2003.

Eighth Grade:

- In 2003, both U.S. eighth-grade boys and girls showed improvement in their average science performance compared to 1995 (figure 4 and table C19 in appendix C).[16] In 2003, U.S. eighth-grade boys scored 536 in science, on average. This was 16 score points higher than in 1995, when U.S. boys scored 520, on average. U.S. girls scored 519 in science, on average, in 2003. This was 14 score points higher than in 1995 and 1999, when U.S. girls scored 505, on average.

- In 2003, U.S. eighth-grade boys outperformed girls in science, on average, which was also the case in 1995 and 1999 (figure 4).

- Both Black and Hispanic eighth-grade students in the United States demonstrated improvement in their average science achievement between 1995 and 2003, and between 1999 and 2003 (figure 4 and table C20 in appendix C). In 1995, U.S. Black eighth-grade students scored 422 in science, on average. This improved to an average of 463 in 2003. U.S. Hispanic eighth-grade students scored 446 in science in 1995, on average, improving to an average score of 482 in 2003.

- As a result of improvements in the average science achievement of Black and Hispanic eighth-graders, the achievement gap between White and Black eighth-graders narrowed from 122 score points in 1995 to 89 score points in 2003, and the achievement gap between White and Hispanic eighth-grade students narrowed from 98 points in 1995 to 70 points in 2003 (figure 4).[17]

- In 2003, U.S. eighth-graders in U.S. public schools with the highest poverty level (75 percent or more of students eligible for free or reduced-price lunch) had lower average science scores compared to their counterparts in public schools with lower poverty levels (figure 4). In contrast, students in schools with the lowest poverty level (10 percent or less eligible students) had higher average science scores than students in schools with poverty levels of 25 percent or more eligible. The difference in the average science scores of students in schools with the lowest and highest poverty levels was 110 score points in 2003.[18]

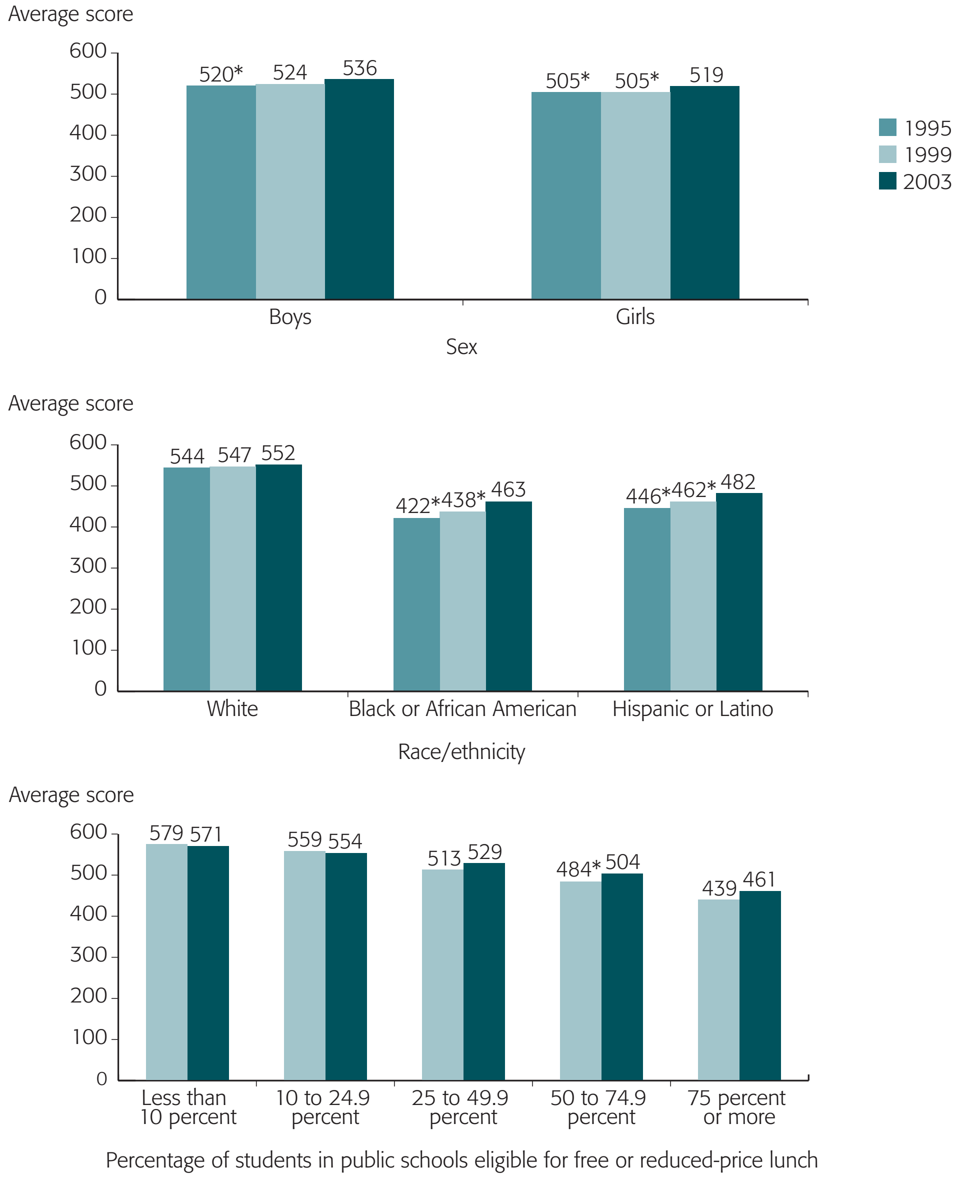
- With a single exception, U.S. eighth-graders who attended schools with varying percentages of students eligible to participate in the federal free or reduced-price lunch program showed no measurable change in their science achievement between 1999 and 2003, the 2 years for which data are available (figure 4 and table C20 in appendix C). U.S. eighth-graders who attended schools in which 50 to almost 75 percent of students were eligible for free or reduced-price lunch did, however, improve their science performance between 1999 and 2003.

[16]See Gonzales et al. (2000) for details on U.S. eighth-grade results for TIMSS 1999.

[17]The effect size of the differences between the average science scores of White and Black, and between White and Hispanic eighth-graders in the United States in 2003 are 1.32 and .99, respectively (table C21 in appendix C for standard deviations of U.S. student population groups).

[18]The effect size of the difference in science achievement between U.S. eighth-grade students in public schools with the lowest and highest levels of poverty in 2003 is 1.67 (table C21 in appendix C for standard deviations of U.S. student population groups).

Figure 4. Average science scale scores of U.S. eighth-grade students, by sex, race/ethnicity, and poverty level: 1995, 1999, and 2003



*p<.05, denotes a significant difference from 2003 average score.
NOTE: Reporting standards not met for Asian category in 1995 or 1999. Reporting standards not met for American Indian or Alaska Native and Native Hawaiian or Other Pacific Islander in 1995, 1999, and 2003. Racial categories exclude Hispanic origin. Other races/ethnicities are included in U.S. totals shown throughout the report. Analyses by poverty level are limited to students in public schools only. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between averages for one student group may be significant while a large difference for another student group may not be significant. The United States met international guidelines for participation rates in 2003 only after replacement schools were included. See appendix A for more information.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2003.

Summary

Looking across the results in mathematics and science, the following points can be made.

- In 2003, fourth-graders in 3 countries—Chinese Taipei, Japan, and Singapore—outperformed U.S. fourth-graders in both mathematics and science, while students in 13 countries turned in lower average mathematics and science scores than U.S. students (tables 2 and 8). U.S. fourth-grade students outperformed their peers in five OECD—member countries (Australia, Italy, New Zealand, Norway and Scotland) of which three are English-speaking countries (Australia, New Zealand and Scotland).

- No measurable changes were detected in the average mathematics and science scores of U.S. fourth-graders between 1995 and 2003 (tables 4 and 10). Moreover, the available data suggest that the performance of U.S. fourth-graders in both mathematics and science was lower in 2003 than in 1995 relative to the 14 other countries that participated in both studies (tables 6 and 12).

- On the other hand, fourth-grade students in six countries showed improvement in both average mathematics and science scores between 1995 and 2003: Cyprus, England, Hong Kong SAR, Latvia-LSS, New Zealand and Slovenia. At the same time, fourth-graders in Norway showed measurable declines in average mathematics and science achievement over the same time period (tables 4 and 10).

- U.S. fourth-grade girls showed no measurable change in their average performance in mathematics and science between 1995 and 2003 (figures 1 and 3). U.S. fourth-grade boys also showed no measurable change in their average mathematics performance, but a measurable decline in science performance over the same time period.

- U.S. Black fourth-graders improved in both mathematics and science between 1995 and 2003 (figures 1 and 3). Hispanic fourth-graders showed no measurable changes in either subject, while White fourth-graders showed no measurable change in mathematics, but declined in science.

- As a result of changes in the performance of Black and White fourth-graders, the gap in achievement between White and Black fourth-grade students in the United States narrowed between 1995 and 2003 in both mathematics and science (figures 1 and 3). In addition, the gap in achievement between Black and Hispanic fourth-graders also narrowed in science over the same time period.

- In 2003, U.S. fourth-graders in U.S. public schools with the highest poverty levels (75 percent or more of students eligible for free or reduced-price lunch) had lower average mathematics and science scores compared to their counterparts in public schools with lower poverty levels (figures 1 and 3).

- Eighth-graders in the five Asian countries that outperformed U.S. eighth-graders in mathematics in 2003—Chinese Taipei, Hong Kong SAR, Japan, Korea, and Singapore—also outperformed U.S. eighth-graders in science in 2003, with eighth-graders in Estonia and Hungary performing better than U.S. students in mathematics and science as well (tables 3 and 9). Students in three of these Asian countries—Chinese Taipei, Japan, and Singapore—outperformed both U.S. fourth- and eighth-graders in mathematics and science on average (tables 2, 3, 8, and 9).

- U.S. eighth-graders improved their average mathematics and science performances in 2003 compared to 1995 (tables 5 and 11). The growth in achievement occurred primarily between 1995 and 1999 in mathematics, and between 1999 and 2003 in science. Moreover, the available data suggest that the performance of U.S. eighth-graders in both mathematics and science was higher in 2003 than it was in 1995 relative to the 21 other countries that participated in the studies (tables 7 and 13).

- In addition to students in the United States, eighth-graders in six other countries showed significant increases in both mathematics and science in 2003 compared to either 1999 or 1995: Hong Kong SAR, Israel, Korea, Latvia-LSS, Lithuania, and the Philippines (tables 5 and 11). On the other hand, eighth-graders in eight countries declined in their mathematics and science performance over this same time period.

- U.S. eighth-grade boys and girls, and U.S. eighth-grade Blacks and Hispanics improved their mathematics and science performances from 1995 (figures 2 and 4). As a result, the gap in achievement between White and Black eighth-graders narrowed in both mathematics and science over this time period.
- In 2003, U.S. eighth-graders in U.S. public schools with the highest poverty levels (75 percent or more of students eligible for free or reduced-price lunch) had lower average mathematics and science scores compared to their counterparts in public schools with lower poverty levels (figures 2 and 4).

References

Braswell, J., Daane, M., and Grigg, W. (2003). *The Nation's Report Card: Mathematics Highlights 2003* (NCES 2004–451). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences,* 2nd ed. Hillsdale, NJ: Erlbaum.

Dossey, J., O'Sullivan, C., and McCrone, S. (forthcoming). *Problem Solving in International Comparative Assessments* (NCES 2005–107). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Ferraro, D., and Rust, K. (2003). *U.S. 2003 TIMSS School Sample: Final Report.* Rockville, MD: Westat.

Gonzales, P., Calsyn, C., Jocelyn, L., Mak, K., Kastberg, D., Arafeh, S., Williams, T., and Tsen, W (2000). *Pursuing Excellence: Comparisons of International Eighth-Grade Mathematics and Science Achievement From a U.S. Perspective, 1995 and 1999* (NCES 2001–028). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Lemke, M., Sen, A., Pahlke, E., Partelow, L., Miller, D., Williams, T., Kastberg, D., Jocelyn, L. (2004). *International Outcomes of Learning in Mathematics and Problem Solving: Results From PISA 2003* (NCES 2005–003). U.S. Department of Education. Washington, DC: U.S. Government Printing Office.

Martin, M.O. and Kelly, D.L. (Eds.). (1997). *Third International Mathematics and Science Study Technical Report, Volume II: Implementation and Analysis.* Chestnut Hill, MA: Boston College.

Martin, M.O., Mullis, I.V.S., and Chrostowski, S.J. (Eds.). (2004). *TIMSS 2003 Technical Report: Findings from IEA's Trends in International Mathematics and Science Study at the Eighth and Fourth Grades.* Chestnut Hill, MA: Boston College.

Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., and Chrostowski, S.J. (2004). *TIMSS 2003 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Eighth and Fourth Grades.* Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., and Chrostowski, S.J. (2004). *TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Eighth and Fourth Grades.* Chestnut Hill, MA: Boston College.

National Center for Education Statistics, U.S. Department of Education. (1997). *Pursuing Excellence: A Study of U.S. Fourth-Grade Mathematics and Science Achievement in International Context* (NCES 97–255). Washington, DC: U.S. Government Printing Office.

Neidorf, T.S., Binkley, M., Gattis, K., and Nohara, D. (forthcoming). *A Content Comparison of the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 Mathematics Assessments.* (NCES 2005–105). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Neidorf, T.S., Binkley, M., and Stephens, M. (forthcoming). *A Content Comparison of the National Assessment of Educational Progress (NAEP) 2000 and Trends in International Mathematics and Science Study (TIMSS) 2003 Science Assessments* (NCES 2005–106). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Peak, L. (1996). *Pursuing Excellence: A Study of U.S. Eighth-Grade Mathematics and Science Teaching, Learning, Curriculum, and Achievement in International Context* (NCES 97–198). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Rosnow, R. L., and Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Pyschological Methods,* 1, 331-340.

Van de Kerckhove, W., and Ferraro, D. (forthcoming). *TIMSS 2003 Non-response Bias Analysis* (NCES 2005–103). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Westat. (2000). *WesVar 4.0 User's Guide*. Rockville, MD: Author.

Appendix A: Technical Notes

Information on the technical aspects of TIMSS 2003 is provided below. More detailed information can be found in the TIMSS 2003 Technical Report (Martin, Mullis, and Chrostowski 2004).

Data Collection

The TIMSS 2003 data were collected by each country, following international guidelines and specifications. TIMSS required that countries select random, nationally representative samples of schools and students. TIMSS countries were asked to identify eligible students based on a common set of criteria, allowing for adaptation to country-specific situations. In IEA studies such as TIMSS, the target population for all countries is called the international desired population. For the fourth-grade assessment, the international desired population consisted of all students in the country who were enrolled in the upper of the two adjacent grades that contained the greatest proportion of 9-year-olds at the time of testing. In the United States and most other countries, this corresponded to fourth grade. For the eighth-grade assessment, the international desired population consisted of all students in the country who were enrolled in the upper of the two adjacent grades that contained the greatest proportion of 13-year-olds at the time of testing. In the United States and most other countries, this corresponded to eighth grade.

TIMSS used a two-stage stratified cluster sampling design. The first stage made use of a systematic probability-proportionate-to-size (PPS) technique to select schools. Although countries participating in TIMSS were strongly encouraged to secure the participation of schools selected in the first stage, it was anticipated that a 100 percent participation rate for schools would not be possible in all countries. Therefore, two replacement schools were identified for each originally sampled school, a priori. As each school was selected, the next school in the sampling frame was designated as a replacement school should the originally sampled school choose not to participate in the study. Should the originally sampled school and the replacement school choose not to participate, a second replacement school was chosen by going to the next school in the sampling frame.

The second stage consisted of selecting classrooms within sampled schools. At the classroom level, TIMSS sampled intact mathematics classes that were offered to students in the target grades. In most countries, one mathematics classroom per school was sampled, although some countries, such as the United States, chose to sample two mathematics classrooms per school.

Exclusions in the TIMSS Sample

All countries were required to define their national desired population to correspond as closely as possible to the definition of the international desired population. In some cases, countries needed to exclude schools and students in remote geographical locations or to exclude a segment of the education system. Any exclusions from the international desired population were clearly documented. Countries were expected to keep the excluded population to no more than 10 percent of the national desired population. Exclusions could take place at the school level, within schools, or both. Participants could exclude schools from the sampling frame for the following reasons:

- Locations were geographically remote;
- Size was extremely small;
- Curriculum or school structure was different from the mainstream education system; or
- Instruction provided was only to students in the categories defined as "within-school exclusions."

Within schools, exclusion decisions were limited to students who, because of some disability, were unable to take part in the TIMSS assessment. The general TIMSS rules for defining within-school exclusion included the following three groups:

- *Intellectually disabled students.* These students were considered, in the professional opinion of the school principal or other qualified staff members, to be intellectually disabled, or had been so diagnosed in psychological tests. This category included students who were emotionally or mentally unable to follow even the general instructions of the TIMSS test. It did not include students who merely exhibited poor academic performance or discipline problems.
- *Functionally disabled students*. These students were permanently physically disabled in such a way that they could not participate in the TIMSS assessment. Functionally disabled students who could perform were included in the testing.

- *Non-native-language speakers*. These students could not read or speak the language of the assessment and so could not overcome the language barrier of testing. Typically, a student who had received less than 1 year of instruction in the language of the assessment was excluded, but this definition was adapted in different countries.

School-level and within-school exclusion rates for TIMSS 2003 are detailed in the next section. Exclusion rates for TIMSS 1995 can be found in chapter 2 of Martin and Kelly (1997); exclusion rates for TIMSS 1999 can be found in appendix 2 of Gonzales et al. (2000).

Response Rates

Based on the sample of schools and students that participated in the assessment, countries were assigned to one of four following categories:

Category 1: met requirements

- An unweighted or weighted school response rate without replacement of at least 85 percent and an unweighted or weighted student response rate of at least 85 percent

or

- The product of the weighted school response rate without replacement and the weighted student response rate of at least 75 percent.

Category 2: met requirements after replacements

- If the requirements for category 1 are not met but the country had either an unweighted or weighted school response rate without replacement of at least 50 percent and had either

- An unweighted or weighted school response rate with replacement of at least 85 percent and a weighted student response rate of at least 85 percent

or

- The product of the weighted school response rate with replacement and the weighted student response rate of at least 75 percent.

Category 3: close to meeting requirements after replacements

- If the requirements for category 1 or 2 are not met but the country had either an unweighted or weighted school response rate without replacement of at least 50 percent and

- The product of the weighted school response rate with replacement and the weighted student response rate near 75 percent.

Category 4: failed to meet requirements

- Unacceptable sampling response rate even when replacement schools are included.

In this report, countries in category 1 appear in the tables and figures without annotation; countries in category 2 are annotated in the tables and figures; countries in category 3 are enclosed with parentheses in the tables and figures, as is the case, for example, of the United States and Morocco at eighth grade. Finally, countries in category 4 are not shown in tables or figures in this report. In addition, annotations are included when the exclusion rate exceeds 10 percent. Latvia is designated as Latvia-LSS (Latvian-speaking schools) in some analyses because data collection in 1995 and 1999 was limited to only those schools in which instruction was in Latvian. Finally, Belgium is annotated as Belgium-Flemish because only the Flemish education system in Belgium participated in TIMSS.

Information on the populations assessed and participation rates is provided in table A1. Details on the number of TIMSS participating schools and students in each of the participating countries are provided in table A2.

Table A1. Coverage of TIMSS grade 4 and 8 target population and participation rates, by country: 2003

Country	Grade 4						
	Years of formal schooling	Percentage of international desired population coverage	National desired population overall exclusion rate	Weighted school participation rate before replacement	Weighted school participation rate after replacement	Weighted student participation rate	Combined weighted school and student participation rate
Armenia	4	100	3	99	99	91	90
Australia	4 or 5	100	3	78	90	94	85
Belgium-Flemish	4	100	6	89	99	98	97
Chinese Taipei	4	100	3	100	100	99	99
Cyprus	4	100	3	100	100	97	97
England	5	100	2	54	82	93	76
Hong Kong SAR[1]	4	100	4	77	88	95	83
Hungary	4	100	8	98	99	94	93
Iran, Islamic Republic of	4	100	6	100	100	98	98
Italy	4	100	4	97	100	97	97
Japan	4	100	1	100	100	97	97
Latvia	4	100	4	91	94	94	88
Lithuania	4	92	5	92	96	92	87
Moldova, Republic of	4	100	4	97	100	97	97
Morocco	4	100	2	87	87	93	81
Netherlands	4	100	5	52	87	96	84
New Zealand	4.5 - 5.5	100	4	87	98	95	93
Norway[2]	4	100	4	89	93	95	88
Philippines	4	100	5	78	85	95	81
Russian Federation	3 or 4	100	7	99	100	97	97
Scotland	5	100	1	64	83	92	77
Singapore	4	100	0	100	100	98	98
Slovenia	3 or 4	100	1	95	99	92	91
Tunisia	4	100	1	100	100	99	99
United States	4	100	5	70	82	95	78

See notes at end of table.

Table A1. Coverage of TIMSS grade 4 and 8 target population and participation rates, by country: 2003—Continued

Country	Grade 8						
	Years of formal schooling	Percentage of international desired population coverage	National desired population overall exclusion rate	Weighted school participation rate before replacement	Weighted school participation rate after replacement	Weighted student participation rate	Combined weighted school and student participation rate
Armenia	8	100	3	99	99	90	89
Australia	8 or 9	100	1	81	90	93	83
Bahrain	8	100	0	100	100	98	98
Belgium-Flemish	8	100	3	82	99	97	94
Botswana	8	100	3	98	98	98	96
Bulgaria	8	100	0	97	97	96	92
Chile	8	100	2	98	100	99	99
Chinese Taipei	8	100	5	100	100	99	99
Cyprus	8	100	3	100	100	96	96
Egypt	8	100	3	99	100	97	97
Estonia	8	100	3	99	99	96	95
Ghana	8	100	1	100	100	93	93
Hong Kong SAR[1]	8	100	3	74	83	97	80
Hungary	8	100	9	98	99	95	94
Indonesia	8	80	0	98	100	99	99
Iran, Islamic Republic of	8	100	6	100	100	98	98
Israel	8	100	23	98	99	95	94
Italy	8	100	4	96	100	97	97
Japan	8	100	1	97	97	96	93
Jordan	8	100	1	100	100	96	96
Korea, Republic of	8	100	5	99	99	99	98
Latvia	8	100	4	92	94	89	83
Lebanon	8	100	1	93	95	96	91
Lithuania	8	89	3	92	95	89	84

See notes at end of table.

Table A1. Coverage of TIMSS grade 4 and 8 target population and participation rates, by country: 2003—Continued

Country	Grade 8						
	Years of formal schooling	Percentage of international desired population coverage	National desired population overall exclusion rate	Weighted school participation rate before replacement	Weighted school participation rate after replacement	Weighted student participation rate	Combined weighted school and student participation rate
Macedonia, Republic of	8	100	12	94	99	97	96
Malaysia	8	100	4	100	100	98	98
Moldova, Republic of	8	100	1	99	100	96	96
Morocco	8	69	1	79	79	91	71
Netherlands	8	100	3	79	87	94	81
New Zealand	8.5 - 9.5	100	4	86	97	93	90
Norway	7	100	2	92	92	92	85
Palestinian National Authority	8	100	0	100	100	99	99
Philippines	8	100	1	81	86	96	82
Romania	8	100	1	99	99	98	98
Russian Federation	7 or 8	100	6	99	99	97	96
Saudi Arabia	8	100	1	95	97	97	94
Scotland	9	100	0	76	85	89	76
Serbia	8	81	3	99	99	96	96
Singapore	8	100	0	100	100	97	97
Slovak Republic	8	100	5	96	100	95	95
Slovenia	7 or 8	100	1	94	99	93	91
South Africa	8	100	1	89	96	92	88
Sweden	8	100	3	97	99	89	87
Tunisia	8	100	2	100	100	98	98
United States	8	100	5	71	78	94	73

[1]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[2]Norway Grade 4: 4 years of formal schooling, but first grade is called "first grade/preschool."
NOTE: Only countries that completed the necessary steps for their data to appear in the reports from the International Study Center are listed. In addition to the countries listed above, four separate jurisdictions participated in TIMSS 2003: the provinces of Ontario and Quebec in Canada; the Basque region of Spain; and the state of Indiana. Yemen participated in TIMSS 2003 but due to difficulties with the data, does not appear in this report. England participated in TIMSS 2003 but did not meet the minimum sampling requirements at grade 8. Information on these jurisdictions can be found in the international *TIMSS 2003 Technical report* (Martin, Mullis, and Chrostowski 2004).
SOURCE: Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., and Chrostowski, S.J. (2004). *TIMSS 2003 International Mathematics Report: Findings from the IEA's Trends in International Mathematics and Science Study at the Eighth and Fourth Grades.* Chestnut Hill, MA: Boston College.

Table A2. TIMSS grade 4 and 8 student and school samples, by country: 2003

Country	Grade 4						
	Schools in original sample	Eligible schools in sample	Schools in original sample that participated	Replacement schools	Total schools that participated	Sampled students in participating schools	Students assessed
Armenia	150	150	148	0	148	6,275	5,674
Australia	230	227	178	26	204	4,675	4,321
Belgium-Flemish	150	150	133	16	149	4,866	4,712
Chinese Taipei	150	150	150	0	150	4,793	4,661
Cyprus	150	150	150	0	150	4,536	4,328
England	150	150	79	44	123	3,917	3,585
Hong Kong SAR[1]	150	150	116	16	132	4,901	4,608
Hungary	160	159	156	1	157	3,603	3,319
Iran, Islamic Republic of	176	171	171	0	171	4,587	4,352
Italy	172	171	165	6	171	4,641	4,282
Japan	150	150	150	0	150	4,690	4,535
Latvia	150	149	137	3	140	3,980	3,687
Lithuania	160	160	147	6	153	5,701	4,422
Moldova, Republic of	153	151	147	4	151	4,162	3,981
Morocco	227	225	197	0	197	4,546	4,264
Netherlands	150	149	77	53	130	3,080	2,937
New Zealand	228	228	194	26	220	4,785	4,308
Norway	150	150	134	5	139	4,706	4,342
Philippines	160	160	122	13	135	5,225	4,572
Russian Federation	206	205	204	1	205	4,229	3,963
Scotland	150	150	94	31	125	4,283	3,936
Singapore	182	182	182	0	182	6,851	6,668
Slovenia	177	177	169	5	174	3,410	3,126
Tunisia	150	150	150	0	150	4,408	4,334
United States	310	300	212	36	248	10,795	9,829

See notes at end of table.

Table A2. TIMSS grade 4 and 8 student and school samples, by country: 2003—Continued

Country	Grade 8						
	Schools in original sample	Eligible schools in sample	Schools in original sample that participated	Replacement schools	Total schools that participated	Sampled students in participating schools	Students assessed
Armenia	150	150	149	0	149	6,388	5,726
Australia	230	226	186	21	207	5,286	4,791
Bahrain	67	67	67	0	67	4,351	4,199
Belgium-Flemish	150	150	122	26	148	5,161	4,970
Botswana	152	150	146	0	146	5,388	5,150
Bulgaria	170	169	163	1	164	4,489	4,117
Chile	195	195	191	4	195	6,528	6,377
Chinese Taipei	150	150	150	0	150	5,525	5,379
Cyprus	59	59	59	0	59	4,314	4,002
Egypt	217	217	215	2	217	7,259	7,095
Estonia	154	152	151	0	151	4,242	4,040
Ghana	150	150	150	0	150	5,690	5,100
Hong Kong SAR[1]	150	150	112	13	125	5,204	4,972
Hungary	160	157	154	1	155	3,506	3,302
Indonesia	150	150	148	2	150	5,884	5,762
Iran, Islamic Republic of	188	181	181	0	181	5,215	4,942
Israel	150	147	143	3	146	4,880	4,318
Italy	172	171	164	7	171	4,628	4,278
Japan	150	150	146	0	146	5,121	4,856
Jordan	150	140	140	0	140	4,871	4,489
Korea, Republic of	151	150	149	0	149	5,451	5,309
Latvia	150	149	137	3	140	4,146	3,630
Lebanon	160	160	148	4	152	4,030	3,814
Lithuania	150	150	137	6	143	6,619	4,964

See notes at end of table.

Table A2. TIMSS grade 4 and 8 student and school samples, by country: 2003—Continued

Country	Grade 8						
	Schools in original sample	Eligible schools in sample	Schools in original sample that participated	Replacement schools	Total schools that participated	Sampled students in participating schools	Students assessed
Macedonia, Republic of	150	150	142	7	149	4,028	3,893
Malaysia	150	150	150	0	150	5,464	5,314
Moldova, Republic of	150	149	147	2	149	4,262	4,033
Morocco	227	165	131	0	131	3,243	2,943
Netherlands	150	150	118	12	130	3,283	3,065
New Zealand	175	174	149	20	169	4,343	3,801
Norway	150	150	138	0	138	4,569	4,133
Palestinian National Authority	150	145	145	0	145	5,543	5,357
Philippines	160	160	132	5	137	7,498	6,917
Romania	150	149	148	0	148	4,249	4,104
Russian Federation	216	216	214	0	214	4,926	4,667
Saudi Arabia	160	160	154	1	155	4,553	4,295
Scotland	150	150	115	13	128	3,962	3,516
Serbia	150	150	149	0	149	4,514	4,296
Singapore	164	164	164	0	164	6,236	6,018
Slovak Republic	180	179	170	9	179	4,428	4,215
Slovenia	177	177	169	5	174	3,883	3,578
South Africa	265	265	241	14	255	9,905	8,952
Sweden	160	160	155	4	159	4,941	4,256
Tunisia	150	150	150	0	150	5,106	4,931
United States	301	296	211	21	232	9,891	8,912

[1]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

NOTE: Only countries that completed the necessary steps for their data to appear in the reports from the International Study Center are listed. In addition to the countries listed above, four separate jurisdictions participated in TIMSS 2003: the provinces of Ontario and Quebec in Canada; the Basque region of Spain; and the state of Indiana. Yemen participated in TIMSS 2003 but due to difficulties with the data, does not appear in this report. England participated in TIMSS 2003 but did not meet the minimum sampling requirements at grade 8. Information on these jurisdictions can be found in the international *TIMSS 2003 Technical report* (Martin, Mullis, and Chrostowski 2004).

SOURCE: Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., and Chrostowski, S.J. (2004). *TIMSS 2003 International Mathematics Report: Findings from the IEA's Trends in International Mathematics and Science Study at the Eighth and Fourth Grades.* Chestnut Hill, MA: Boston College.

Sampling, Data Collection, and Response Rates in the United States

The TIMSS 2003 school sample was drawn for the United States in November 2002. The sample design for this school sample was developed to follow international requirements as given in the TIMSS sampling manual. The U.S. sample for 2003 was a two-stage sampling process with the first stage a sample of schools, and the second stage a sample of students' classrooms from the target grade in sampled schools. Unlike TIMSS 1995 and 1999, the sample was not clustered at the geographic level for TIMSS 2003. This change was made in an effort to reduce the design effects and to spread the respondent burden across schools districts as much as possible.

The sample design for TIMSS was a stratified systematic sample, with sampling probabilities proportional to measures of size. The U.S. TIMSS fourth-grade sample had two explicit strata based on poverty. A high poverty school was defined as one in which 50 percent or more of the students were eligible for participation in the federal free or reduced-price lunch program; high poverty schools were oversampled (Ferraro and Rust 2003) This variable was not available for private schools, so they were all treated as low poverty schools. The target sample sizes were 120 high-poverty and 190 low-poverty schools.

Within the poverty strata, there are four categorical implicit stratification variables: type of school (public or private), region of the country[19] (Northeast, Southeast, Central, West), type of location relative to populous areas (eight levels), minority status (above or below 15 percent). The last sort key within the implicit stratification was by grade enrollment in descending order.

The TIMSS eighth-grade sample had no explicit stratification. The frame was implicitly stratified (i.e., sorted for sampling) by four categorical stratification variables: type of school (public or private), region of the country, type of location relative to populous areas (eight levels), minority status (above or below 15 percent). The last sort key within the implicit stratification was by grade enrollment in descending order.

At the same time that the TIMSS sample was selected, replacement schools were identified following the TIMSS guidelines by assigning the two schools neighboring the sampled school on the frame as replacements. There were several constraints on the assignment of substitutes. One sampled school was not allowed to substitute for another, and a given school could not be assigned to substitute for more than one sampled school. Furthermore, substitutes were required to be in the same implicit stratum as the sampled school. If the sampled school was the first or last school in the stratum, then the second school following or preceding the sampled school was identified as the substitute. One was designated a first replacement and the other a second replacement. If an original school refused to participate, the first replacement was then contacted. If that school also refused to participate, the second school was then contacted.

The schools were selected with probability proportionate to the school's estimated enrollment of fourth- and eighth-grade students from the 2003 NAEP school frame with 2000-01 school data. The data for public schools were from the Common Core of Data (CCD), and the data for private schools was from the Private School Survey (PSS). Any school containing a fourth or an eighth grade as of the school year 2000-01 was included on the school sampling frame. Participating schools provided lists of fourth- or eighth-grade classrooms, and one or two intact mathematics classrooms were selected within each school in an equal probability sample. The overall sample design for the United States was intended to approximate a self-weighting sample of students as much as possible, with each fourth- or eighth-grade student having an equal probability of being selected.

The U.S. TIMSS fourth-grade school sample consisted of 310 schools, of which 300 were eligible schools and 212 agreed to participate. The school response rate before replacement was 70 percent (weighted; 71 percent unweighted). The weighted school response rate before replacement is given by the formula:

$$\text{weighted school response rate before replacement} = \frac{\sum_{i \in Y} W_i E_i}{\sum_{i \in (Y \cup N)} W_i E_i},$$

where Y denotes the set of responding original sample schools with age-eligible students, N denotes the set of eligible non-responding original sample schools, W_i denotes the base weight for school i, $W_i = 1/P_i$, where P_i denotes the school selection probability for school i, and E_i denotes the enrollment size of age-eligible students, as indicated on the sampling frame.

In addition to the 212 participating schools, 36 replacement schools also participated for a total of 248 participating schools at the fourth grade in the United States.

A total of 10,795 students were sampled for the fourth-grade assessment. Of these students, 49 were withdrawn from the school before the assessment was administrated. Of the eligible 10,746 sampled students, an additional 429 students were excluded using the criteria described above, for a weighted exclusion rate of 5 percent. Of the 10,317 remaining sample students, a total of 9,829 students participated in the assessment in the United States, since 488 students were absent. The student participation rate was 95 percent.

The combined school and students weighted and unweighted response rate of 78 percent after replacement schools were included was achieved (66 percent weighted and 67 percent unweighted

[19] Region is the 'state-based' region (NAEPRG_S on the output files). Northeast consists of Connecticut, Delaware, District of Columbia, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont. Central consists of Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin. West consists of Alaska, Arizona, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oklahoma, Texas, Utah, Washington, Oregon, California, and Wyoming. Southeast consists of Alabama, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia.

without replacement). As a result, the U.S. data for fourth-grade students are annotated to indicate that international guidelines for participation rates were met only after replacement schools were included.

The U.S. TIMSS eighth-grade school sample consisted of 301 schools, of which 296 were eligible schools and 211 agreed to participate. The school response rate before replacement was 71 percent (weighted and unweighted). In addition to the 211 participating schools, 21 replacement schools also participated for a total of 232 participating schools at the eighth grade in the United States.

A total of 9,891 students were sampled for the assessment. Of these students, 90 were withdrawn from the school before the assessment was administrated. Of the eligible 9,801 sampled students, an additional 279 students were excluded using the criteria described above, for a weighted exclusion rate of 5 percent. Of the 9,522 remaining sample students, a total of 8,912 students participated in the assessment in the United States, since 610 students were absent. The student participation rate was 94 percent (weighted and unweighted). The combined school and students weighted and unweighted response rate of 73 percent after replacement schools were included was achieved (66 percent without replacement schools). As a result, the U.S. data for eighth-grade students are in parentheses to indicate that United States did not meet international sampling guidelines.

NCES standards require a nonresponse bias analysis if the school level response rate is below 80 percent (using the base weight). Since the U.S. school response rates at the fourth and eighth grades were below 80 percent, even with replacements, NCES required an analysis of the potential magnitude of nonresponse bias at the school level. To accomplish this analysis, two methods were chosen (Van de Kerckhove and Ferraro forthcoming). The first method was focused exclusively on the original sample of schools, treating all those that were substituted as nonrespondents. A second method focused on the final sample of schools (including replacements), treating as nonrespondents those schools from which a final response was not received. Both methods were used to analyze the U.S. TIMSS fourth- and eighth-grade data for potential bias.

In order to compare TIMSS respondents and nonrespondents it was necessary to match the sample of schools back to the sample frame to detect as many characteristics as possible that might provide information about the presence of nonresponse bias. Comparing characteristics for respondents and nonrespondents is not always a good measure of nonresponse bias if the characteristics are unrelated or weakly related to more substantive items in the survey. However, this is often the only approach available. The characteristics that were analyzed based on the sampling frame were taken from the 2000-2001 Common Core of Data (CCD) for public schools, and from the 2000-2001 Private School Survey (PSS) for private schools. For categorical variables, the distribution of the characteristics for respondents was compared with the distribution for all schools. The hypothesis of independence between a given school characteristic and the response status (whether or not participated) was tested using a Rao-Scott modified Chi-square statistic. For continuous variables, summary means were calculated. The 95 percent confidence interval for the difference between the mean for respondents and the mean for all schools was tested to see whether or not it included zero. In addition to these tests, logistic regression models were set up to identify whether any of the school characteristics were significant in predicting response status because logistic regression allows investigation of all variables at the same time.

Public and private schools were modeled together using the following variables: community type; public/religious affiliation; NAEP region; poverty level; number of students enrolled in fourth or eighth grade; total number of students; percentage Asian or Pacific Islander students; percentage Black, non-Hispanic students; percentage Hispanic students; percentage American Indian or Alaska Native students; and percentage White, non-Hispanic students.

The investigation into nonresponse bias at the school level for TIMSS fourth grade generally showed that there was no statistically significant relationship between response status and the majority of school characteristics available for analysis. For the original sample of schools in TIMSS fourth grade, schools in the Northeast were less likely to respond than schools in the West, Southeast or Central regions of the country. However, the regression did not confirm this result. The results for the final sample of schools showed a significant effect on the percentage of Black, non-Hispanic students (responding schools had more Black, non-Hispanic students than non-responding schools). However, the regression did not confirm this result.

The investigation into nonresponse bias at the school level for TIMSS eighth grade showed that, for the original sample of schools, responding schools were more likely to be in rural areas than in central city or urban fringe areas, have fewer students than non-responding schools, have fewer Hispanic students, and were more likely to be Catholic or public schools. However, the regression confirmed only that responding schools in the original sample were more likely to be from rural areas and have fewer students than non-responding schools. The number of Hispanic students in responding schools and their public/religious affiliation were not confirmed by the regression. The results with the final sample of schools were more complicated. The total number of students remained significant, but the additional variable of public/religious affiliation also appeared to be significantly related to response rate according to the logistic regression. Public and Catholic schools were more likely to respond than private, non-sectarian and private-other religious schools. Finally, while the first analysis indicated that schools in rural areas were more like to respond than schools in the central city or urban fringe, this was not confirmed by the logistic regression.

The results of these analyses suggest that there is no statistically significant relationship between response status and the majority of the school characteristics tested, with the exception of the variables noted above at each grade level. The potential for nonresponse bias exists however. It is difficult to assess the amount of any bias in the survey as a result of the associations that exist.

It is also not clear what effect the weighting adjustments for nonresponse have on any bias. In general, these weighting adjustments cannot address all of the potential bias, only some of it. There is no evaluation of how much effect the weighting adjustments have on the bias.

Test Development

TIMSS is a cooperative effort involving representatives from every country participating in the study. For TIMSS 2003, the development effort began with a revision of the frameworks that are used to guide the construction of the assessment (Mullis et al. 2001). The framework was updated to reflect changes in the curriculum and instruction of participating countries. Extensive input from experts in mathematics and science education, assessment, curriculum, and representatives from national educational centers around the world contributed to the final shape of the frameworks. Maintaining the ability to measure change over time was an important factor in revising the frameworks.

As part of the TIMSS dissemination strategy, approximately one-third of the 1995 fourth-grade assessment items and one-half of the 1999 eighth-grade assessment items were released for public use. To replace assessment items that had been released in earlier years, countries submitted items for review by subject-matter specialists, and additional items were written to ensure that the content, as explicated in the frameworks, was covered adequately. Items were reviewed by an international Science and Mathematics Item Review Committee and pilot-tested in most of the participating countries. Results from the field test were used to evaluate item difficulty, how well items discriminated between high- and low-performing students, the effectiveness of distracters in multiple-choice items, scoring suitability and reliability for constructed-response items, and evidence of bias towards or against individual countries or in favor of boys or girls. As a result of this review, 243 of the 435 new fourth-grade items were selected for inclusion in the assessment. In total, there were 313 mathematics and science items included in the fourth-grade TIMSS assessment booklets. At eighth grade, the review of the item statistics from the field test led to the inclusion of 230 of the 386 new eighth-grade items in the assessment. In total, there were 383 mathematics and science items included in the eighth-grade TIMSS assessment booklets. More detail on the distribution of new and trend items is included in table A3.

Table A3. Distribution of new and trend mathematics and science items in the TIMSS grade 4 and 8 assessments, by type: 2003

Response type	Grade 4			Grade 8		
	Total	New items	Trend items	Total	New items	Trend items
Total	313	243	70	383	230	153
Multiple choice	183	115	68	237	125	112
Constructed response	130	128	2	146	105	41
Mathematics	161	124	37	194	115	79
Multiple choice	92	55	37	128	69	59
Constructed response	69	69	0	66	46	20
Science	152	119	33	189	115	74
Multiple choice	91	60	31	109	56	53
Constructed response	61	59	2	80	59	21

SOURCE: Martin, M.O., Mullis, I.V.S., and Chrostowski, S.J. (2004). *TIMSS 2003 Technical Report: Findings from IEA's Trends in International Mathematics and Science Study at the Eighth and Fourth Grades.* Chestnut Hill, MA: Boston College.

The TIMSS 2003 frameworks included specifications for what are termed "problem-solving and inquiry" (PSI) tasks. PSI tasks were developed to assess how well students could draw on and integrate information and processes in mathematics and science as part of an investigation or in order to solve problems. The PSI tasks developed for TIMSS 2003 needed to be self-contained, involve minimal equipment, and be integrated into the main assessment without any special accommodations or additional testing time. While the PSI tasks are not full scientific investigations, the tasks were designed to require a basic understanding of the nature of science and mathematics, and to elicit some of the skills essential to the inquiry process. The tasks were designed to draw on students' understandings of and abilities with formulating questions and hypotheses; designing investigations; collecting, representing, analyzing, and interpreting data; and drawing conclusions and developing explanations based on evidence.

The PSI tasks were assembled as longer blocks or clusters of items that, together, related to an overall theme (e.g., speciation). Nine PSI blocks were field-tested at fourth grade. Of the nine blocks, six blocks were eventually incorporated into the fourth-grade assessment. The six blocks covered both mathematics and science, focusing on geometry, measurement, number, life science, earth science, and physical science.

At eighth grade, 10 PSI blocks were field-tested. Of the 10 blocks, 7 blocks were eventually incorporated into the eighth-grade assessment. The seven blocks covered both mathematics and science, focusing on algebra, data, geometry, measurement, number, chemistry, physics, and life science. The PSI tasks were incorporated into the overall assessments and, thus, not reported separately at either grade level.

Design of Instruments

TIMSS 2003 included booklets containing assessment items as well as questionnaires submitted to principals, teachers, and students for response. The assessment booklets were constructed such that not all of the students responded to all of the items. This is consistent with other large-scale assessments, such as the National Assessment of Educational Progress. To keep the testing burden to a minimum, and to ensure broad subject-matter coverage, TIMSS used a rotated block design that included both mathematics and science items. That is, students encountered both mathematics and science items during the assessment. The 2003 fourth-grade assessment consisted of 12 booklets, each requiring approximately 72 minutes of response time. The 12 booklets were rotated among students, with each participating student completing 1 booklet only. The mathematics and science items were assembled into 14 blocks or clusters of items. Each block contained either mathematics items or science items only. The secure or trend items were included in 3 blocks, with the other 11 blocks containing replacement items. Each of the 12 booklets contained 6 blocks (in total).

The 2003 eighth-grade assessment also consisted of 12 booklets, each requiring approximately 90 minutes of response time. The 12 booklets were rotated among students, with each participating student completing 1 booklet only. The mathematics and science items were assembled into 14 blocks or clusters of items. Each block contained either mathematics items or science items only. The secure or trend items were included in 3 blocks, with the other 11 blocks containing replacement items. Each of the 12 booklets contained 6 blocks (in total).

As part of the design process, it was necessary to ensure that the booklets showed a distribution across the mathematics and science content domains as specified in the frameworks. The number of mathematics and science items in the fourth and eighth-grade TIMSS 2003 assessments is shown in table A4.

Table A4. Number of mathematics and science items in the TIMSS grade 4 and 8 assessments, by type and content domain: 2003

Content domain	Grade 4			Grade 8		
		Response type			Response type	
	Total	Multiple choice	Constructed response	Total	Multiple choice	Constructed response
Total items	313	183	130	383	237	146
Mathematics - Total	161	92	69	194	128	66
Number	63	30	33	57	43	14
Patterns, equations, and relationships	24	16	8	47	29	18
Measurement	33	23	10	31	19	12
Geometry	24	12	12	31	22	9
Data	17	11	6	28	15	13
Science - Total	152	91	61	189	109	80
Life science	65	41	24	54	29	25
Physical science	53	29	24	†	†	†
Earth science	34	21	13	31	22	9
Environmental science	†	†	†	27	10	17
Chemistry	†	†	†	31	20	11
Physics	†	†	†	46	28	18

†Not applicable. Content domain does not apply for the grade shown.
SOURCE: Martin, M.O., Mullis, I.V.S. and Chrostowski, S.J. (2004). *TIMSS 2003 Technical Report: Findings from IEA's Trends in International Mathematics and Science Study at the Eighth and Fourth Grades.* Exhibit 2.21. Chestnut Hill, MA: Boston College.

In addition to the assessment booklets, TIMSS 2003 included questionnaires for principals, teachers, and students. As with prior iterations of TIMSS, the questionnaires used in TIMSS 2003 are based on prior versions of the questionnaires. The questionnaires were reviewed extensively by the national research coordinators from the participating countries as well as a Questionnaire Item Review Committee. Like the assessment booklets, all questionnaire items were field tested, and the results reviewed carefully. As a result, some of the questionnaire items needed to be revised prior to their inclusion in the final questionnaires. The questionnaires requested information to help provide a context for the performance scores, focusing on such topics as students' attitudes and beliefs about learning, student habits and homework, and their lives both in and outside of school; teachers' attitudes and beliefs about teaching and learning, teaching assignments, class size and organization, instructional practices, and participation in professional development activities; and principals' viewpoints on policy and budget responsibilities, curriculum and instruction issues, student behavior, as well as descriptions of the organization of schools and courses.

Calculator Usage

Calculators were not permitted during the TIMSS fourth-grade assessment. However, the TIMSS policy on calculator use at the eighth grade was to give students the best opportunity to operate in settings that mirrored their classroom experiences. Beginning with 2003, calculators were permitted but not required for newly developed eighth-grade assessment materials. Participating countries could decide whether or not their students were allowed to use calculators for the new items; the United States allowed students to use calculators. Since calculators were not permitted at the eighth grade in the 1995 or 1999 assessments, the 2003 eighth-grade test booklets were designed so that trend items from these assessments were placed in the first half and new items in 2003 placed in the second half. Where countries chose to permit eighth-grade students to use calculators, they could use them for the second half of the booklet only.

Translation

Source versions of all instruments (assessment booklets, questionnaires and manuals) were prepared in English and translated into the primary language or languages of instruction in each country. In addition, it was sometimes necessary to adapt the instrument for cultural purposes, even in countries that use English as the primary language of instruction. All adaptations were reviewed and approved by the International Study Center to ensure they did not change the substance or intent of the question or answer choices. For example, proper names were sometimes changed to names that would be more familiar to students (e.g., Marja-leena to Maria).

Each country prepared translations of the instruments according to translation guidelines established by the International Study Center. Adaptations to the instruments were documented by each country, and submitted for review. The goal of the translation guidelines was to produce translated instruments of the highest quality that would provide comparable data across countries.

Translated instruments were verified by an independent, professional translation agency prior to final approval and printing of the instruments. Countries were required to submit copies of the final printed instruments to the International Study Center. Further details on the translation process can be found in the TIMSS 2003 Technical Report (Martin, Mullis, and Chrostowski 2004).

Test Administration and Quality Assurance

TIMSS 2003 emphasized the use of standardized procedures in all countries. Each country collected its own data, based on comprehensive manuals and trainings provided by the international project team to explain the survey's implementation, including precise instructions for the work of school coordinators and scripts for test administrators for use in testing sessions. Test administration in the United States was carried out by professional staff trained according to the international guidelines. School staff were asked only to assist with listings of students, identifying space for testing in the school, and specifying any parental consent procedures needed for sampled students.

Each country was responsible for conducting quality control procedures and describing this effort in the national research coordinators' report documenting procedures used in the study. In addition, the International Study Center considered it essential to monitor compliance with the standardized procedures. National research coordinators were asked to nominate one or more persons unconnected with their national center, such as retired school teachers, to serve as quality control monitors for their countries. The International Study Center developed manuals for the monitors and briefed them in 2-day training sessions about TIMSS, the responsibilities of the national centers in conducting the study, and their own roles and responsibilities.

Scoring Reliability

The TIMSS assessment items included both multiple choice and constructed-response items. A scoring rubric (guide) was created for every item included in the TIMSS assessments. These were carefully written and reviewed by national research coordinators and other experts as part of the field test of items, and revised accordingly.

The national research coordinator in each country was responsible for scoring and coding of data in that country, following established guidelines. The national research coordinator and, sometimes, additional staff, attended scoring training sessions held by the International Study Center. The training sessions focused on the scoring rubrics and coding system employed in TIMSS. Participants were provided extensive practice in scoring example items over several days. Information on within-country agreement among coders was collected and documented by the International Study Center. Information on scoring and coding reliability was also used to calculate cross-country agreement among coders. Scoring reliability for TIMSS 2003 is provided in table A5.

Table A5. Within-country constructed-response scoring reliability for TIMSS grade 4 and 8 mathematics and science items, by exact percent score agreement and country: 2003

Country	Grade 4					
	Mathematics			Science		
	Average across items	Range		Average across items	Range	
		Min	Max		Min	Max
International average	99	92	100	96	85	100
Armenia	99	98	100	99	97	100
Australia	100	98	100	99	94	100
Belgium-Flemish	100	96	100	99	89	100
Chinese Taipei	99	83	100	98	89	100
Cyprus	98	91	100	94	76	100
England	99	91	100	98	87	100
Hong Kong SAR[1]	100	98	100	99	97	100
Hungary	98	91	100	95	80	100
Iran, Islamic Republic of	100	98	100	96	85	100
Italy	98	92	100	94	77	100
Japan	99	95	100	97	86	100
Latvia	98	87	100	96	82	100
Lithuania	97	77	100	93	81	100
Moldova, Republic of	100	100	100	100	100	100
Morocco	98	93	100	97	93	100
Netherlands	97	86	100	91	71	99
New Zealand	99	94	100	97	86	100
Norway	99	95	100	97	85	100
Philippines	99	96	100	97	89	100
Russian Federation	100	97	100	99	98	100
Scotland	99	98	100	98	90	100
Singapore	100	99	100	100	99	100
Slovenia	98	84	100	91	74	100
Tunisia	97	89	100	93	79	100
United States	97	88	100	93	70	100

See notes at end of table.

Table A5. Within-country constructed-response scoring reliability for TIMSS grade 4 and 8 mathematics and science items, by exact percent score agreement and country: 2003—Continued

Country	Grade 8					
	Mathematics			Science		
	Average across items	Range		Average across items	Range	
		Min	Max		Min	Max
International average	99	92	100	97	88	100
Armenia	99	94	100	98	92	100
Australia	100	97	100	99	94	100
Bahrain	99	98	100	98	94	100
Belgium-Flemish	99	96	100	97	89	100
Botswana	99	91	100	95	74	100
Bulgaria	96	70	100	91	72	99
Chile	99	95	100	97	91	100
Chinese Taipei	100	91	100	99	97	100
Cyprus	98	86	100	96	87	100
Egypt	100	97	100	100	98	100
Estonia	100	98	100	99	97	100
Ghana	99	97	100	98	93	100
Hong Kong SAR[1]	100	98	100	99	97	100
Hungary	98	90	100	96	87	100
Indonesia	98	90	100	96	87	100
Iran, Islamic Republic of	99	94	100	98	87	100
Israel	98	93	100	95	89	100
Italy	99	95	100	98	91	100
Japan	99	94	100	97	81	100
Jordan	99	98	100	99	97	100
Korea, Republic of	99	87	100	98	84	100
Latvia	98	90	100	94	78	100
Lebanon	100	94	100	100	98	100
Lithuania	97	71	100	90	69	100

See notes at end of table.

Table A5. Within-country constructed-response scoring reliability for TIMSS grade 4 and 8 mathematics and science items, by exact percent score agreement and country: 2003—Continued

Country	Grade 8					
	Mathematics			Science		
	Average across items	Range		Average across items	Range	
		Min	Max		Min	Max
Macedonia, Republic of	100	97	100	99	96	100
Malaysia	100	98	100	99	98	100
Moldova, Republic of	100	99	100	100	99	100
Morocco	97	89	100	94	86	100
Netherlands	97	84	100	90	70	100
New Zealand	99	96	100	98	92	100
Norway	98	91	100	95	83	100
Palestinian National Authority	99	94	100	95	82	100
Philippines	99	97	100	98	89	100
Romania	100	98	100	99	96	100
Russian Federation	99	95	100	99	92	100
Saudi Arabia	99	94	100	97	87	100
Scotland	99	95	100	97	89	100
Serbia	99	96	100	99	94	100
Singapore	100	98	100	100	99	100
Slovak Republic	100	98	100	99	95	100
Slovenia	97	86	100	90	70	100
South Africa	99	95	100	99	94	100
Sweden	98	89	100	92	76	100
Tunisia	98	89	100	98	90	100
United States	97	86	100	92	72	100

[1]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
NOTE: To gather and document within-country agreement among scorers, systematic subsamples of at least 100 students' responses to each constructed-response item was coded independently by two readers. The agreement score indicates the degree of agreement among coders on marking student responses in the same way. See Mullis et al. (2004) and Martin et al. (2004) for more details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Data Entry and Cleaning

Responsibility for data entry was taken by the national research coordinator from each country. The data collected for TIMSS 2003 were entered into data files with a common international format, as specified in the Manual for Entering the TIMSS 2003 Data. Data entry was facilitated by the use of a common software available to all participating countries (WinDEM). The software facilitated the checking and correction of data by providing various data consistency checks. The data were then sent to the IEA Data Processing Center (DPC) in Hamburg, Germany for cleaning. The DPC checked that the international data structure was followed; checked the identification system within and between files; corrected single case problems manually; and applied standard cleaning procedures to questionnaire files. Results of the data cleaning process were documented by the DPC. This documentation was then shared with the national research coordinator with specific questions to be addressed. The national research coordinator then provided the DPC with revisions to coding or solutions for anomalies. The DPC then compiled background univariate statistics and preliminary classical and Rasch Item Analysis. Detailed information on the entire data entry and cleaning process can be found in the TIMSS 2003 Technical Report (Martin, Mullis, and Chrostowski 2004).

Weighting, Scaling, and Plausible Values

Before the data were analyzed, responses from the groups of students assessed were assigned sampling weights to ensure that their representation in TIMSS 2003 results matched their actual percentage of the school population in the grade assessed. Based on these sampling weights, the analyses of TIMSS 2003 data were conducted in two major phases—scaling and estimation. During the scaling phase, item response theory (IRT) procedures were used to estimate the measurement characteristics of each assessment question. During the estimation phase, the results of the scaling were used to produce estimates of student achievement. Subsequent analyses related these achievement results to the background variables collected by TIMSS 2003.

Weighting

Responses from the groups of students were assigned sampling weights to adjust for over-representation or under-representation from a particular group. The use of sampling weights is necessary for the computation of statistically sound, nationally representative estimators. The weight assigned to a student's responses is the inverse of the probability that the student would be selected for the sample. When responses are weighted, none are discarded, and each contributes to the results for the total number of students represented by the individual student assessed. Weighting also adjusts for various situations such as school and student nonresponse because data cannot be assumed to be randomly missing. The internationally defined weighting specifications for TIMSS require that each assessed student's sampling weight should be the product of (1) the inverse of the school's probability of selection, (2) an adjustment for school-level nonresponse, (3) the inverse of the classroom's probability of selection, and (4) an adjustment for student-level nonresponse. All TIMSS 1995, 1999 and 2003 analyses are conducted using sampling weights.

Scaling

TIMSS 1995, 1999, and 2003 used item response theory (IRT) methods to produce score scales that summarized the achievement results. With this method, the performance of a sample of students in a subject area or sub-area could be summarized on a single scale or a series of scales, even when different students had been administered different items. Because of the reporting requirements for TIMSS and because of the large number of background variables associated with the assessment, a large number of analyses had to be conducted. The procedures TIMSS used for the analyses were developed to produce accurate results for groups of students while limiting the testing burden on individual students. Furthermore, these procedures provided data that could be readily used in secondary analyses. IRT scaling provides estimates of item parameters (e.g., difficulty, discrimination) that define the relationship between the item and the underlying variable measured by the test. Parameters of the IRT model are estimated for each test question, with an overall scale being established as well as scales for each predefined content area specified in the assessment framework. For example, the TIMSS 2003 eighth-grade assessment had five scales describing mathematics content strands, and science had scales for five fields of science.

TIMSS 1995 utilized a one parameter IRT model to produce score scales that summarized the achievement results. The TIMSS 1995 data were rescaled using a three-parameter IRT model to match the procedures used to scale the 1999 and 2003 TIMSS data. The three-parameter model was preferred to the one-parameter model because it can more accurately account for the differences among items in their ability to discriminate between students of high and low ability. After careful study of the rescaling process, the International Study Center concluded that the fit between the original TIMSS data and the rescaled TIMSS data met acceptable standards. However, as a result of rescaling, the average achievement scores of some countries changed from those initially reported in 1996 and 1997 (Peak 1996; NCES 1997). The rescaled TIMSS scores are included in this report.

Plausible Values

During the scaling phase, plausible values were used to characterize scale scores for students participating in the assessment. To keep student burden to a minimum, TIMSS administered a limited number of assessment items to each student—too few to produce accurate content-related scale scores for each student. To account for this, for each student, TIMSS generated five possible content-related scale scores that represented selections from the distribution of content-related scale scores of students with similar backgrounds who answered the assessment items the same way. The plausible-values technology is one way to ensure that the estimates of the average performance of student populations and the estimates of variability in those estimates are more accurate than those determined through traditional procedures, which estimate a single score for each student.

During the construction of plausible values, careful quality control steps ensured that the subpopulation estimates based on these plausible values were accurate. Plausible values were constructed separately for each national sample. TIMSS uses the plausible-values methodology to represent what the true performance of an individual might have been, had it been observed. This is done by using a small number of random draws from an empirically derived distribution of score values based on the student's observed responses to assessment items and on background variables. Each random draw from the distribution is considered a representative value from the distribution of potential scale scores for all students in the sample who have similar characteristics and identical patterns of item responses. The draws from the distribution are different from one another to quantify the degree of precision (the width of the spread) in the underlying distribution of possible scale scores that could have caused the observed performances. The TIMSS plausible values function like point estimates of scale scores for many purposes, but they are unlike true point estimates in several respects. They differ from one another for any particular student, and the amount of difference quantifies the spread in the underlying distribution of possible scale scores for that student. Because of the plausible-values approach, secondary researchers can use the TIMSS data to carry out a wide range of analyses.

Data Limitations

As with any study, there are limitations to TIMSS 2003 that researchers should take into consideration. Estimates produced using data from TIMSS 2003 are subject to two types of error, nonsampling and sampling errors. Nonsampling errors can be due to errors made in collecting and processing data. Sampling errors can occur because the data were collected from a sample rather than a complete census of the populations.

Nonsampling Errors

Nonsampling error is a term used to describe variations in the estimates that may be caused by population coverage limitations, nonresponse bias, and measurement error, as well as data collection, processing, and reporting procedures. The sources of nonsampling errors are typically problems like unit and item nonresponse, the difference in respondents' interpretations of the meaning of the questions, response differences related to the particular time the survey was conducted, and mistakes in data preparation.

Missing Data

There are four kinds of missing data: nonresponse, missing or invalid, not applicable, and not reached. *Nonresponse* data occurs when a respondent was expected to answer an item but no response was given. Responses that are *missing or invalid* occur in multiple-choice items where an invalid response is given. The code is not used for opened-ended questions. An item is *not applicable* when it is not possible for the respondent to answer the question. Finally, items that are *not reached* are consecutive missing values starting from the end of each test session. All four kinds of missing data are coded differently in the TIMSS 2003 database.

Missing background data are not included in the analyses for this report and are not imputed. In general, item response rates for variables discussed in this report were over the NCES standard of 85 percent to report without notation (table A6).

In general, it is difficult to identify and estimate either the amount of nonsampling error or the bias caused by this error. In TIMSS 2003, efforts were made to prevent such errors from occurring and to compensate for them when possible. For example, the design phase entailed a field test that evaluated items as well as the implementation procedures for the survey. It should also be recognized that most background information was obtained from students' self-reports, which are subject to respondent bias. One potential source of respondent bias in this survey was social desirability bias, for example, if students reported that they enjoyed mathematics.

Sampling Errors

Sampling errors occur when the discrepancy between a population characteristic and the sample estimate arises because not all members of the reference population are sampled for the survey. The size of the sample relative to the population and the variability of the population characteristics both influence the magnitude of sampling error. The particular sample of students in fourth and eighth grade from the 2002-03 school year was just one of many possible samples that could have been selected. Therefore, estimates produced from the TIMSS sample may differ from estimates that would have been produced had another student sample been drawn. This type of variability is called sampling error because it arises from using a sample of students in fourth or eighth grade, rather than all students in the grade in that year.

The standard error is a measure of the variability due to sampling when estimating a statistic. The approach used for calculating sampling variances in TIMSS was the Jackknife Repeated Replication (JRR). Standard errors can be used as a measure for the precision expected from a particular sample. Standard errors for all of the estimates are included in appendix C. The standard errors can be used to produce confidence intervals. There is a 95 percent chance that the true average lies within the range of 1.96 times the standard errors above or below the estimated score. For example, the average mathematics score for the U.S. eighth-grade students was 504 in 2003, and this statistic had a standard error of 3.3. Therefore, it can be stated with 95 percent confidence that the actual

Table A6. Weighted response rates for unimputed variables for TIMSS grade 4 and 8: 2003

Variable	Variable ID	Source of information	Grade 4 U.S. response rate	Grade 4 Range of response rates in other countries	Grade 8 U.S. response rate	Grade 8 Range of response rates in other countries
Sex	ITSEX	Classroom Tracking Form	100	94 – 100	100	92 – 100
Race/ethnicity	STRACE	Student Questionnaire	98	—	98	—
Free or reduced-priced lunch[1]	FRLUNCH	School Questionnaire	85	—	82	—

—Not available.

[1]The response rate is calculated for public schools only.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

average of U.S. eighth-grade students in 2003 was between 498 and 511 (1.96 x 3.3 = 6.5; confidence interval = 504 +/- 6.5).

Description of Background Variables

The international version of the TIMSS 2003 student, teacher and school questionnaires are available at http://timss.bc.edu. The U.S. versions of these questionnaires are available at http://nces.ed.gov/timss.

Race/Ethnicity

Students' race/ethnicity was obtained through student responses to a two-part question. Students were asked first whether they were Hispanic or Latino, and then asked whether they were members of the following racial groups: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, or White. Multiple responses to the race classification question were allowed. Results are shown separately for Asians, Blacks, Hispanics, and Whites. Students identifying themselves as Hispanic and also other races were included in the Hispanic group.

Poverty Level in Public Schools (Percentage of Students Eligible for Free or Reduced-price Lunch)

The poverty level in public schools was obtained from principal responses to the school questionnaire. The question asked what percentage of students at the school was eligible to receive free or reduced-price lunch through the National School Lunch Program around the first of October, 2002. The answers were grouped into five categories: less than 10 percent; 10 to 24.9 percent; 25 to 49.9 percent; 50 to 74.9 percent; and 75 percent or more. Analysis was limited to public schools only.

Confidentiality and Disclosure Limitations

The TIMSS 2003 data are hierarchical and include school data and student data from the participating schools. Confidentiality analyses for the United States were designed to provide reasonable assurance that public use data files issued by the IEA would not allow identification of individual U.S. schools or students when compared against public data collections. Disclosure limitation included the identification and masking of potential disclosure-risk TIMSS schools and adding an additional measure of uncertainty of school, teacher, and student identification through random swapping of data elements within the student, teacher, and school files.

Statistical Procedures

Tests of Significance

Comparisons made in the text of this report have been tested for statistical significance. For example, in the commonly made comparison of country averages against the average of the United States, tests of statistical significance were used to establish whether or not the observed differences from the U.S. average were statistically significant. The estimation of the standard errors that are required in order to undertake the tests of significance is complicated by the complex sample and assessment designs which both generate error variance. Together they mandate a set of statistically complex procedures in order to estimate the correct standard errors. As a consequence, the estimated standard errors contain a sampling variance component estimated by Jackknife Repeated Replication (JRR); and, where the assessments are concerned, an additional imputation variance component arising from the assessment design. Details on the procedures used can be found in the WesVar 4.0 User's Guide (Westat 2000).

In almost all instances, the tests for significance used were standard *t* tests. These fell into two categories according to the nature of the comparison being made: comparisons of independent and non-independent samples. Before describing the *t* tests used, some background on the two types of comparisons is provided below:

The variance of a difference is equal to the sum of the variances of the two initial variables minus two times the covariance between the two initial variables. A sampling distribution has the same characteristics as any distribution, except that units consist of sample estimates and not observations. Therefore,

$$\sigma^2_{(\hat{\mu}_X - \hat{\mu}_Y)} = \sigma^2_{(\hat{\mu}_X)} + \sigma^2_{(\hat{\mu}_Y)} - 2\text{cov}(\hat{\mu}_X, \hat{\mu}_Y)$$

The sampling variance of a difference is equal to the sum of the two initial sampling variances minus two times the covariance between the two sampling distributions on the estimates.

If one wants to determine whether the girls' performance differs from the boys' performance, for example, then as for all statistical analyses, a null hypothesis has to be tested. In this particular example, it consists of computing the difference between the boys' performance mean and the girls' performance mean (or the inverse). The null hypothesis is:

$$H_0 : \hat{\mu}_{(boys)} - \hat{\mu}_{(girls)} = 0$$

To test this null hypothesis, the standard error on this difference is computed and then compared to the observed difference. The respective standard errors on the mean estimate for boys and girls ($\sigma_{(\hat{\mu}_{boys})}$, $\sigma_{(\hat{\mu}_{girls})}$) can be easily computed.

The expected value of the covariance will be equal to 0 if the two sampled groups are independent. If the two groups are not independent, as is the case with girls and boys attending the same schools within a country, or comparing a country mean with the international mean which includes that particular country, then the expected value of the covariance might differ from 0.

In TIMSS, country samples are independent. Therefore, for any comparison between two countries, the expected value of the covariance will be equal to 0, and thus the standard error on the estimate is:

$$\sigma_{(\hat{\theta}_i - \hat{\theta}_j)} = \sqrt{\sigma^2_{(\hat{\theta}_i)} + \sigma^2_{(\hat{\theta}_j)}}$$

with θ being any statistic.

Within a particular country, any sub-samples will be considered as independent only if the categorical variable used to define the sub-samples was used as an explicit stratification variable.

If sampled groups are not independent, the estimation of the covariance between, for instance, $\hat{\mu}_{(boys)}$ and $\hat{\mu}_{(girls)}$ would require the selection of several samples and then the analysis of the variation of $\hat{\mu}_{(boys)}$ in conjunction with $\hat{\mu}_{(girls)}$. Such a procedure is of course unrealistic. Therefore, as for any computation of a standard error in TIMSS, replication methods using the supplied replicate weights are used to estimate the standard error on a difference. Use of the replicate weights implicitly incorporates the covariance between the two estimates into the estimate of the standard error on the difference.

Thus, in simple comparisons of independent averages such as the U.S. average with other country averages, the following formula was used to compute the *t* statistic:

$$t = \frac{(est_1 - est_2)}{\sqrt{(se_1)^2 + (se_2)^2}}$$

Est$_1$ and est$_2$ are the estimates being compared (e.g., average of country A and the U.S. average) and se$_1$ and se$_2$ are the corresponding standard errors of these averages.

The second type of comparison used in this report occurred when comparing differences of non-subset, non-independent groups, such as when comparing the average scores of males versus females within the United States. In such comparisons, the following formula was used to compute the *t* statistic:

$$t = \frac{(est_{grp1} - est_{grp2})}{se(est_{grp1} - est_{grp2})}$$

Est$_{grp1}$ and est$_{grp2}$ are the non-independent group estimates being compared. Se(est$_{grp1}$ - est$_{grp2}$) is the standard error of the difference calculated using Jackknife Repeated Replication (JRR), which accounts for any covariance between the estimates for the two non-independent groups.

Effect size

Tests of statistical significance are, in part, influenced by sample sizes. To provide the reader with an increased understanding of the importance of the significant difference between student populations in the United States, effect sizes are included in the report. Effect sizes use standard deviations, rather than standard errors, and are therefore not influenced by the size of the student population samples. Following Cohen (1988) and Rosnow and Rosenthal (1996), effect size is calculated by finding the difference between the means of two groups and dividing that result by the pooled standard deviation of the two groups:

$$d = \frac{est_{grp1} - est_{grp2}}{sd_{pooled}}$$

Est_{grp1} and est_{grp2} are the student group estimates being compared. Sd_{pooled} is the pooled standard deviation of the groups being compared. The formula for the pooled standard deviation is as follows (Rosnow and Rosenthal 1996):

$$sd_{pooled} = \sqrt{\frac{sd_1^2 + sd_2^2}{2}}$$

Sd_1 and sd_2 are the standard deviations of the groups being compared. In social sciences, an effect size of .2 is considered small, one of .5 is of medium importance, and one of .8 or larger is considered large (Cohen 1988).

Country participation

Table A7 shows the countries that participated in TIMSS 2003 at fourth and eighth grades. The countries are grouped by continent. In addition, countries that are members of the Organization for Economic Cooperation and Development (OECD) are indicated with a check mark.

Table A7. Countries that participated in TIMSS grade 4 and 8 by continent and OECD membership: 2003

Grade 4

Continent and country	OECD member
Africa	
Morocco	
Tunisia	
Asia	
Armenia	
Chinese Taipei	
Hong Kong SAR[1]	
Iran, Islamic Republic of	
Japan	✓
Philippines	
Singapore	
Europe	
Belgium-Flemish	✓
Cyprus	
England	✓
Hungary	✓
Italy	✓
Latvia	
Lithuania	
Moldova, Republic of	
Netherlands	✓
Norway	✓
Russian Federation	
Scotland	✓
Slovenia	
The Americas	
United States	✓
Australia/Oceania	
Australia	✓
New Zealand	✓

Grade 8

Continent and country	OECD member
Africa	
Morocco	
Egypt	
Ghana	
Tunisia	
South Africa	
Asia	
Armenia	
Bahrain	
Botswana	
Bulgaria	
Chinese Taipei	
Hong Kong SAR[1]	
Indonesia	
Iran, Islamic Republic of	
Israel	
Japan	✓
Jordan	
Korea, Republic of	✓
Lebanon	
Malaysia	
Palestinian National Authority	
Philippines	
Saudi Arabia	
Singapore	
Europe	
Belgium-Flemish	✓
Cyprus	
Estonia	
Hungary	✓
Italy	✓
Latvia	
Lithuania	
Macedonia, Republic of	
Moldova, Republic of	
Netherlands	✓
Norway	✓
Romania	
Russian Federation	
Scotland	✓
Serbia	
Slovak Republic	✓
Slovenia	
Sweden	✓
The Americas	
Chile	
United States	✓
Australia/Oceania	
Australia	✓
New Zealand	✓

[1]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
NOTE: The Organization for Economic Cooperation and Development (OECD) is an intergovernmental organization of 30 industrialized countries that serves as a forum for member countries to cooperate in research and policy development on social and economic topics of common interest.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study, 2003.

Appendix B: Example Items and 2003 Country Results

Exhibit B1: Fourth-grade example item for number: 2003

There are 600 balls in a box, and $\frac{1}{3}$ of the balls are red.

How many red balls are in the box?

Answer: ___200___ red balls

Country	Percent full credit
International average	49
Lithuania[1]	85
Singapore	84
Latvia	83
Belgium-Flemish	82
Russian Federation	78
Moldova, Republic of	68
Cyprus	64
Hong Kong SAR[2,3]	64
Armenia	63
Netherlands[3]	63
Hungary	62
Japan	56
Chinese Tapei	55
Italy	43
England[3]	41
Scotland[3]	40
United States[3]	**38**
New Zealand	34
Slovenia	32
Australia[3]	30
Tunisia	24
Norway	19
Philippines	14
Iran, Islamic Republic of	9
Morocco	7

[1]National desired population does not cover all international desired population.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]Met international guidelines for participation rates only after replacement schools were included.
NOTE: Countries are sorted by 2003 average percent correct. Countries were required to sample students in the upper of the two grades that contained the most number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Exhibit B2. Fourth-grade example item for patterns, equations and relationships: 2003

□ represents the number of magazines that Lina reads each week.
Which of these represents the total number of magazines that Lina reads in 6 weeks?

Ⓐ 6 + □

● 6 × □

Ⓒ □ + 6

Ⓓ (□ + □) × 6

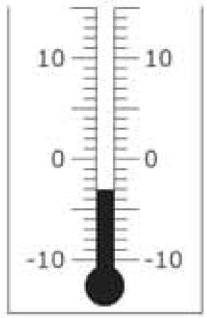
M012046

Country	Percent full credit
International average	58
Singapore	86
Chinese Tapei	81
Hong Kong SAR[1,2]	76
Netherlands[2]	72
United States[2]	**72**
Belgium-Flemish	67
Japan	67
Russian Federation	67
England[2]	66
Latvia	66
Cyprus	65
Moldova, Republic of	64
Lithuania[3]	62
Hungary	61
Scotland[2]	60
Slovenia	60
Australia[2]	56
New Zealand	54
Italy	50
Armenia	46
Philippines	38
Norway	37
Iran, Islamic Republic of	34
Morocco	29
Tunisia	20

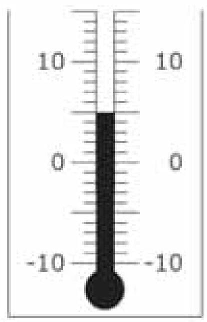
[1]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[2]Met international guidelines for participation rates only after replacement schools were included.
[3]National desired population does not cover all international desired population.
NOTE: Countries are sorted by 2003 average percent correct. Countries were required to sample students in the upper of the two grades that contained the most number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Exhibit B3: Fourth-grade example item for measurement: 2003

When Tracy left for school, the temperature was minus 3 degrees.



At recess, the temperature was 5 degrees.



How many degrees did the temperature rise?

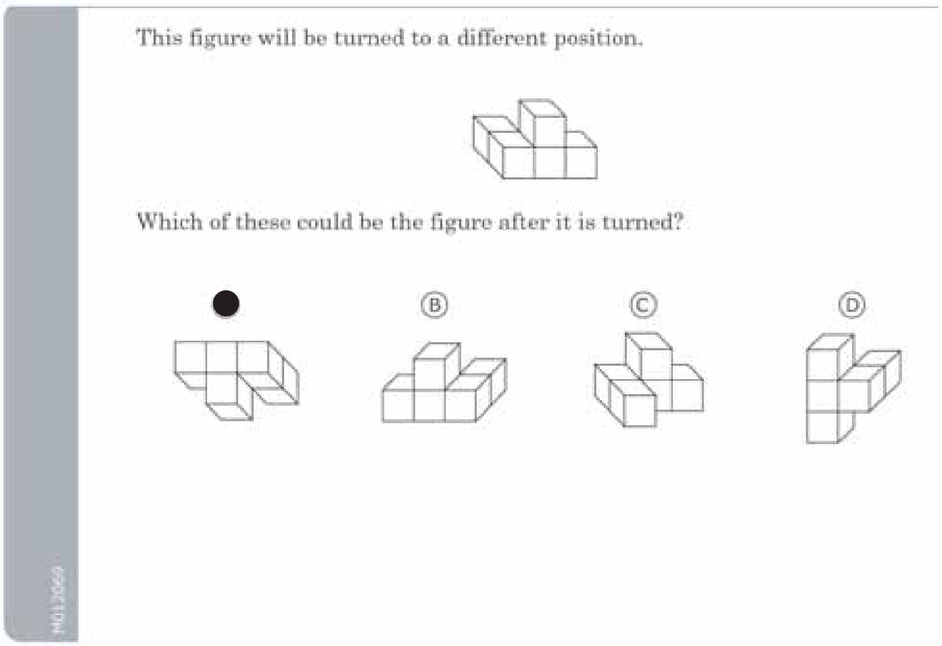
- Ⓐ 2 degrees
- Ⓑ 3 degrees
- Ⓒ 5 degrees
- ● 8 degrees

M011013

Country	Percent full credit
International average	58
Netherlands[1]	78
Belgium-Flemish	76
England[1]	76
Hungary	74
Japan	71
Latvia	70
Italy	66
Norway	65
Lithuania[2]	63
Russian Federation	63
Chinese Tapei	60
Scotland[1]	59
Hong Kong SAR[1,3]	58
Cyprus	55
Moldova, Republic of	55
Slovenia	54
Singapore	53
United States[1]	**52**
Armenia	51
New Zealand	49
Australia[1]	46
Iran, Islamic Republic of	46
Philippines	44
Morocco	29
Tunisia	27

[1]Met international guidelines for participation rates only after replacement schools were included.
[2]National desired population does not cover all international desired population.
[3]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
NOTE: Countries are sorted by 2003 average percent correct. Countries were required to sample students in the upper of the two grades that contained the most number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

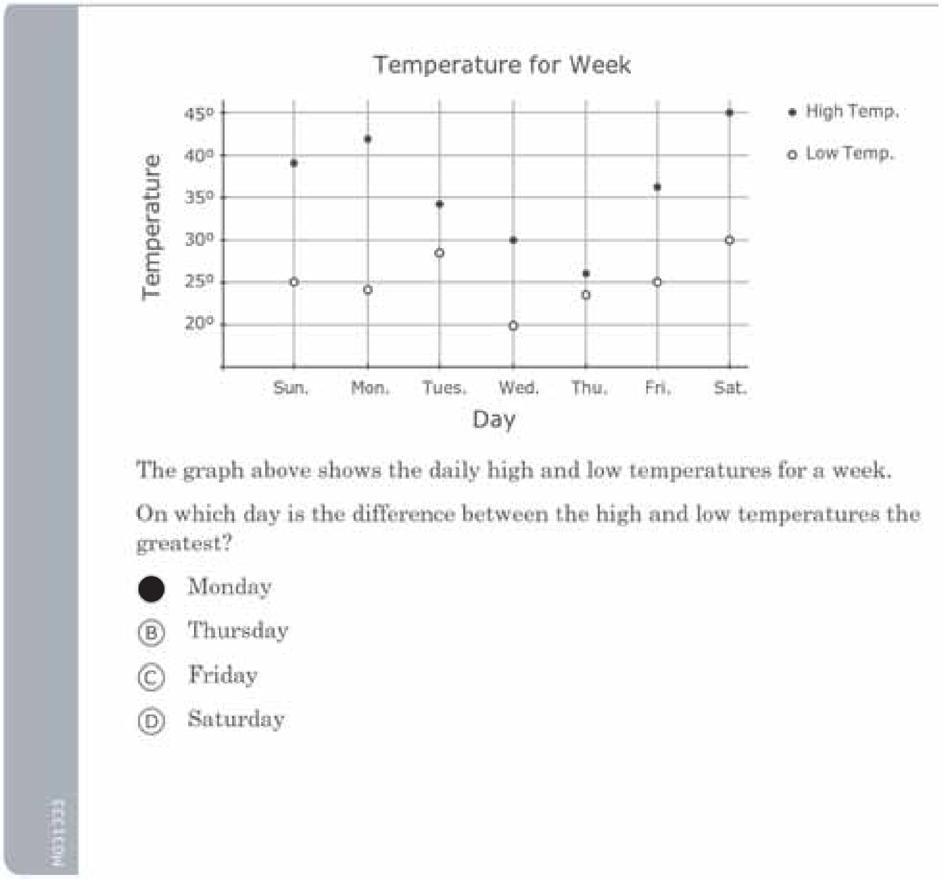
Exhibit B4: Fourth-grade example item for geometry: 2003



Country	Percent full credit
International average	43
Norway	60
Latvia	59
Chinese Tapei	58
Singapore	54
Belgium-Flemish	52
Slovenia	51
Hungary	50
Japan	50
Italy	49
Scotland[1]	49
England[1]	46
New Zealand	45
Hong Kong SAR[1,2]	43
Australia[1]	42
Russian Federation	41
Netherlands[1]	40
Moldova, Republic of	39
United States[1]	**39**
Tunisia	35
Armenia	34
Lithuania[3]	32
Cyprus	31
Iran, Islamic Republic of	26
Philippines	23
Morocco	20

[1]Met international guidelines for participation rates only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]National desired population does not cover all international desired population.
NOTE: Countries are sorted by 2003 average percent correct. Countries were required to sample students in the upper of the two grades that contained the most number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

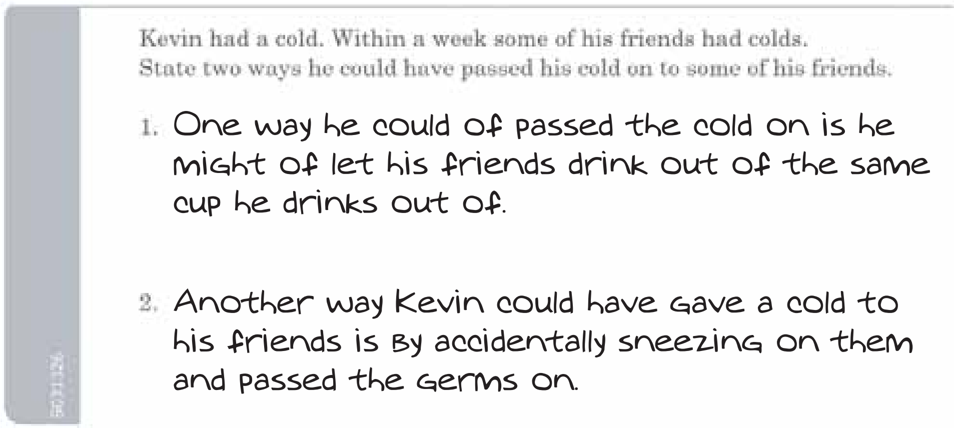
Exhibit B5: Fourth-grade example item for data: 2003



Country	Percent full credit
International average	42
Japan	73
Hong Kong SAR[1,2]	69
Belgium-Flemish	68
Chinese Tapei	57
Lithuania[3]	56
Netherlands[2]	56
England[2]	54
Latvia	48
Singapore	47
Russian Federation	44
Hungary	41
Cyprus	40
Moldova, Republic of	39
Scotland[2]	39
New Zealand	38
Slovenia	38
United States[2]	**38**
Italy	37
Australia[2]	34
Norway	32
Philippines	30
Morocco	25
Armenia	22
Iran, Islamic Republic of	16
Tunisia	13

[1]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[2]Met international guidelines for participation rates only after replacement schools were included.
[3]National desired population does not cover all international desired population.
NOTE: Countries are sorted by 2003 average percent correct. Countries were required to sample students in the upper of the two grades that contained the most number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Exhibit B6: Fourth-grade example item for life science: 2003



Kevin had a cold. Within a week some of his friends had colds.
State two ways he could have passed his cold on to some of his friends.

1. One way he could of passed the cold on is he might of let his friends drink out of the same cup he drinks out of.

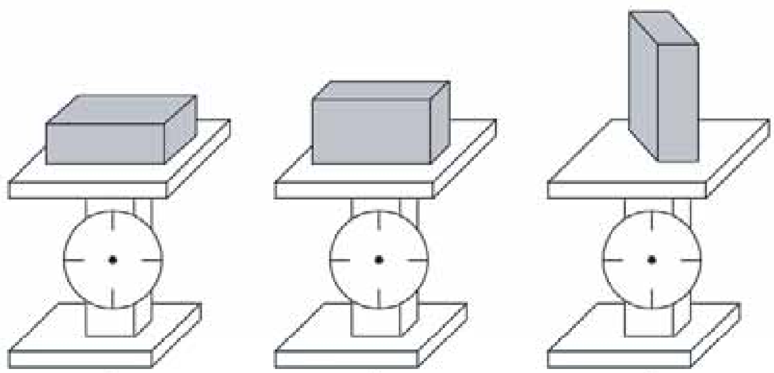
2. Another way Kevin could have gave a cold to his friends is by accidentally sneezing on them and passed the germs on.

Country	Percent full credit
International average	29
Netherlands[1]	45
Singapore	45
Japan	43
Belgium-Flemish	40
Italy	39
Latvia	37
Chinese Tapei	36
Hong Kong SAR[1,2]	35
Cyprus	34
Russian Federation	33
Slovenia	32
Hungary	31
Norway	31
Australia[1]	28
England[1]	28
Lithuania[3]	28
United States[1]	**27**
Iran, Islamic Republic of	24
New Zealand	24
Scotland[1]	24
Tunisia	20
Moldova, Republic of	16
Armenia	9
Morocco	7
Philippines	5

[1]Met international guidelines for participation rates only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]National desired population does not cover all international desired population.
NOTE: Countries are sorted by 2003 average percent correct. Countries were required to sample students in the upper of the two grades that contained the most number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Exhibit B7: Fourth-grade example item for physical science, forces and motion: 2003

The same brick is put on a scale in three different ways.



What will the scale show?

Ⓐ 1 will show the greatest weight.

Ⓑ 2 will show the greatest weight.

Ⓒ 3 will show the greatest weight.

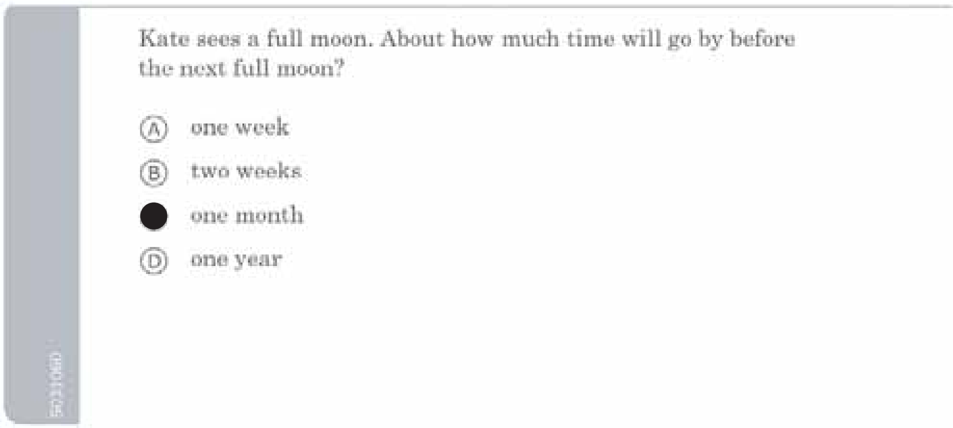
● All will show the same weight.

S011009

Country	Percent full credit
International average	72
Lithuania[1]	88
Moldova, Republic of	87
Russian Federation	86
Chinese Tapei	85
Slovenia	85
Latvia	84
Hungary	79
Singapore	79
Italy	78
England[2]	76
Armenia	74
Australia[2]	74
Netherlands[2]	74
Belgium-Flemish	73
United States[2]	**73**
Iran, Islamic Republic of	72
Hong Kong SAR[2,3]	69
Scotland[1]	68
Japan	66
New Zealand	66
Cyprus	63
Morocco	54
Norway	54
Philippines	52
Tunisia	45

[1]National desired population does not cover all international desired population.
[2]Met international guidelines for participation rates only after replacement schools were included.
[3]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
NOTE: Countries are sorted by 2003 average percent correct. Countries were required to sample students in the upper of the two grades that contained the most number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Exhibit B8: Fourth-grade example item for earth science, earth in the solar system and universe: 2003



Kate sees a full moon. About how much time will go by before the next full moon?

(A) one week

(B) two weeks

● one month

(D) one year

Country	Percent full credit
International average	37
Chinese Tapei	62
Latvia	47
Moldova, Republic of	46
New Zealand	45
Slovenia	45
United States[1]	**43**
Norway	40
Australia[1]	39
England[1]	39
Japan	38
Russian Federation	38
Hong Kong SAR[1,2]	37
Netherlands[1]	37
Scotland[1]	36
Singapore	36
Belgium-Flemish	34
Iran, Islamic Republic of	34
Italy	34
Philippines	33
Lithuania[3]	32
Armenia	30
Cyprus	27
Tunisia	27
Hungary	26
Morocco	25

[1]Met international guidelines for participation rates only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]National desired population does not cover all international desired population.
NOTE: Countries are sorted by 2003 average percent correct. Countries were required to sample students in the upper of the two grades that contained the most number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Exhibit B9. Eighth-grade example item for number: 2003

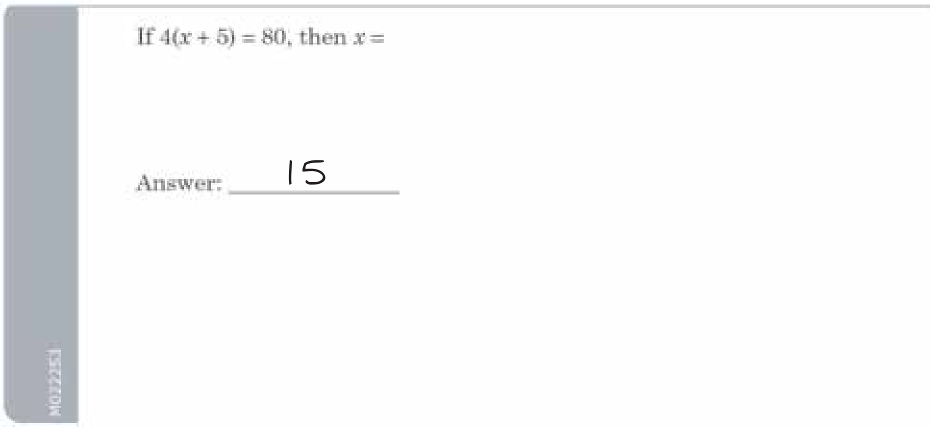
If n is a negative integer, which of these is the largest number?

- (A) $3 + n$
- (B) $3 \times n$
- (●) $3 - n$
- (D) $3 \div n$

Country	Percent full credit
International average	40
Korea, Republic of	79
Chinese Taipei	74
Russian Federation	69
Japan	68
Hong Kong SAR[1,2]	67
Singapore	58
Hungary	53
Estonia	51
Latvia	49
Belgium-Flemish	48
(Israel)	48
(United States)	**48**
Armenia	47
Serbia[3]	46
Slovak Republic	46
Netherlands[1]	44
Bulgaria	42
Lebanon	41
Romania	41
Lithuania[3]	40
Malaysia	40
Moldova, Republic of	39
Slovenia	38
Egypt	37
Australia	36
Cyprus	34
Iran, Islamic Republic of	32
Italy	32
(Macedonia, Republic of)	32
Philippines	32
New Zealand	31
Jordan	29
Palestinian National Authority	29
Scotland[1]	28
Sweden	28
Indonesia[3]	27
South Africa	26
(Morocco)	25
Norway	25
Saudi Arabia	25
Bahrain	22
Botswana	22
Tunisia	22
Chile	18
Ghana	#

#Rounds to zero.
[1]Met international guidelines for participation rates only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]National desired population does not cover all of the international desired population.
NOTE: Countries are sorted by 2003 average percent correct. Parentheses indicate countries that did not meet international sampling or other guidelines. See appendix A for more information. The international average reported here may differ from that reported in Mullis et al. (2004) due to the deletion of England. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA). Trends in International Mathematics and Science Study (TIMSS), 2003.

Exhibit B10. Eighth-grade example item for algebra, equation and formulas: 2003


If $4(x + 5) = 80$, then $x =$

Answer: 15

M022253

Country	Percent full credit
International average	45
Hong Kong SAR[1,2]	90
Korea, Republic of	82
Singapore	82
Chinese Taipei	80
Japan	80
Estonia	72
Hungary	70
Russian Federation	66
Slovak Republic	65
Belgium-Flemish	64
Latvia	64
Slovenia	64
Armenia	61
Romania	61
Serbia[3]	61
Bulgaria	59
(Israel)	57
(United States)	**57**
Cyprus	54
Moldova, Republic of	53
Lithuania[3]	51
Australia	50
Malaysia	46
Netherlands[1]	44
New Zealand	44
Italy	37
(Macedonia, Republic of)	37
Scotland[1]	37
Lebanon	31
Sweden	28
Tunisia	26
Indonesia[3]	25
Jordan	25
Egypt	23
Philippines	23
Bahrain	19
Iran, Islamic Republic of	18
Palestinian National Authority	17
(Morocco)	16
Norway	11
Chile	9
Saudi Arabia	6
South Africa	6
Botswana	5
Ghana	#

#Rounds to zero.
[1]Met international guidelines for participation rates only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]National desired population does not cover all of the international desired population.
NOTE: Countries are sorted by 2003 average percent correct. Parentheses indicate countries that did not meet international sampling or other guidelines. See appendix A for more information. The international average reported here may differ from that reported in Mullis et al. (2004) due to the deletion of England. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA). Trends in International Mathematics and Science Study (TIMSS), 2003.

Exhibit B11. Eighth-grade example item for measurement, attributes and units: 2003

Which of these is the LEAST amount of time?

Ⓐ 1 day

● 20 hours

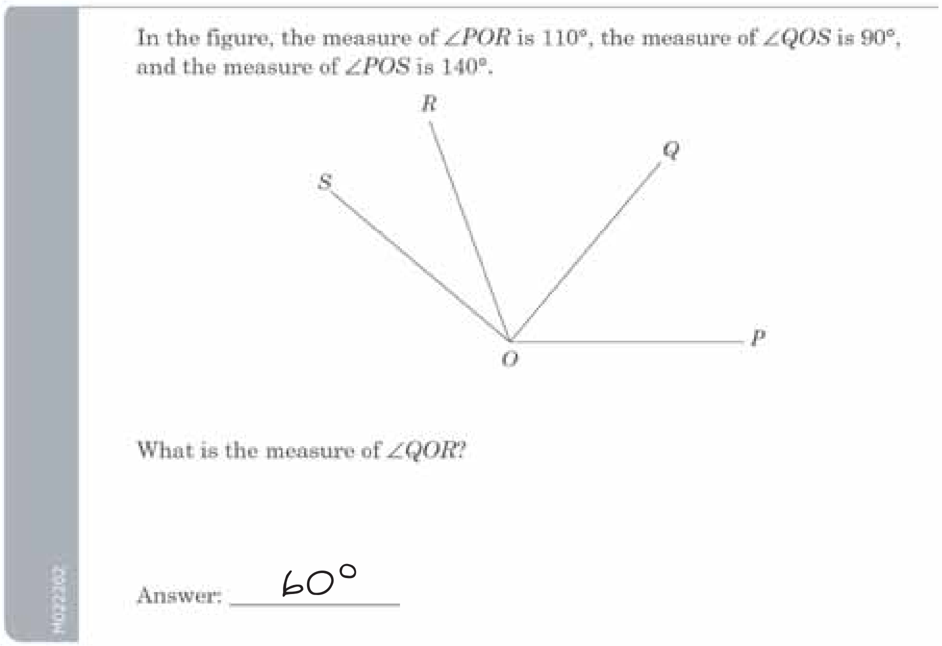
Ⓒ 1800 minutes

Ⓓ 90 000 seconds

Country	Percent full credit
International average	44
Chinese Taipei	66
Hungary	63
Korea, Republic of	63
Singapore	60
Belgium-Flemish	59
Hong Kong SAR[1,2]	54
Japan	54
Slovenia	54
Netherlands[1]	52
Slovak Republic	52
Latvia	51
Armenia	50
Serbia[3]	49
Estonia	48
(Macedonia, Republic of)	48
Russian Federation	48
Malaysia	47
(United States)	**47**
Bulgaria	45
Italy	45
Moldova, Republic of	45
Sweden	44
Romania	43
Lithuania[3]	42
Australia	41
(Israel)	41
Tunisia	41
Lebanon	40
Cyprus	39
Norway	39
Jordan	38
Scotland[1]	38
Palestinian National Authority	37
Egypt	36
New Zealand	36
Chile	35
Iran, Islamic Republic of	35
Philippines	35
Saudi Arabia	35
Bahrain	32
(Morocco)	32
South Africa	32
Ghana	27
Botswana	26
Indonesia[3]	26

[1]Met international guidelines for participation rates only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]National desired population does not cover all of the international desired population.
NOTE: Countries are sorted by 2003 average percent correct. Parentheses indicate countries that did not meet international sampling or other guidelines. See appendix A for more information. The international average reported here may differ from that reported in Mullis et al. (2004) due to the deletion of England. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA). Trends in International Mathematics and Science Study (TIMSS), 2003.

Exhibit B12. Eighth-grade example item for geometry, lines and angles: 2003

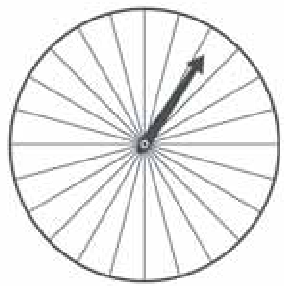


Country	Percent full credit
International average	28
Korea, Republic of	64
Japan	60
Singapore	58
Hong Kong SAR[1,2]	57
Chinese Taipei	49
Hungary	44
Norway	41
Russian Federation	40
Armenia	39
Latvia	37
Belgium-Flemish	36
Estonia	36
Slovak Republic	36
Serbia[3]	35
Bulgaria	34
Romania	34
(Israel)	32
Malaysia	32
Moldova, Republic of	32
Netherlands[1]	28
New Zealand	28
Lithuania[3]	27
Australia	26
Lebanon	26
(Macedonia, Republic of)	26
Italy	25
Slovenia	25
(United States)	**22**
Cyprus	21
Sweden	20
Tunisia	19
Scotland[1]	17
Bahrain	16
Indonesia[3]	16
Palestinian National Authority	16
Egypt	15
Jordan	14
Iran, Islamic Republic of	11
(Morocco)	11
Philippines	11
Chile	10
Botswana	9
Saudi Arabia	6
South Africa	4
Ghana	#

#Rounds to zero.
[1]Met international guidelines for participation rates only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]National desired population does not cover all of the international desired population.
NOTE: Countries are sorted by 2003 average percent correct. Parentheses indicate countries that did not meet international sampling or other guidelines. See appendix A for more information. The international average reported here may differ from that reported in Mullis et al. (2004) due to the deletion of England. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA). Trends in International Mathematics and Science Study (TIMSS), 2003.

Exhibit B13. Eighth-grade example item for data, uncertainty and probability: 2003

The figure below shows a spinner with 24 sectors. When someone spins the arrow, it is equally likely to stop on any sector.



$\frac{1}{8}$ of the sectors are blue, $\frac{1}{24}$ are purple, $\frac{1}{2}$ are orange, and $\frac{1}{3}$ are red. If a person spins the arrow, on which color sector is the spinner LEAST likely to stop?

Ⓐ blue

● purple

Ⓒ orange

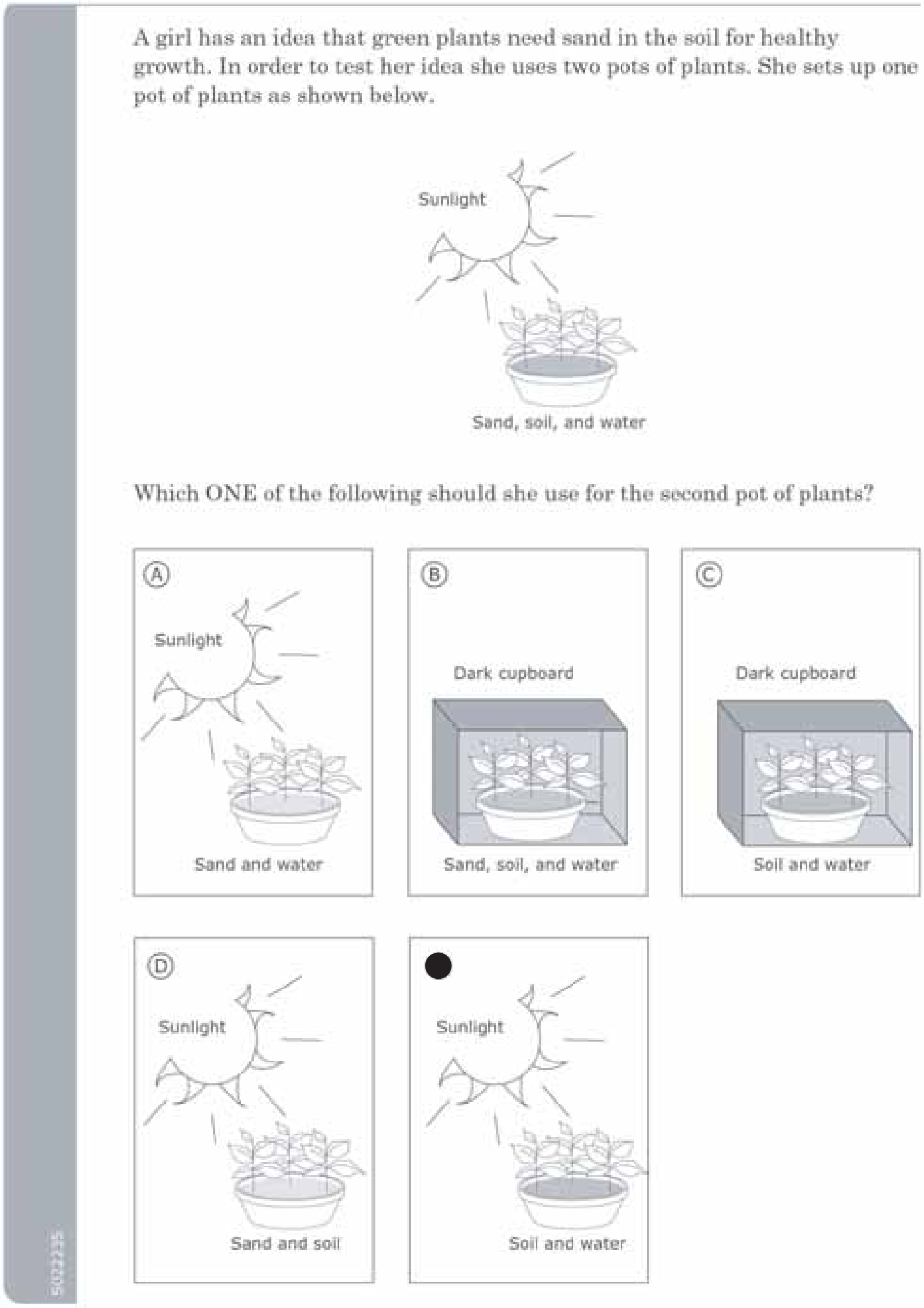
Ⓓ red

M022252

Country	Percent full credit
International average	60
Hong Kong SAR[1,2]	87
Chinese Taipei	85
Netherlands[1]	85
Japan	82
Belgium-Flemish	81
Sweden	81
Korea, Republic of	79
Singapore	79
Australia	78
(United States)	**78**
Hungary	76
Scotland[1]	76
(Israel)	74
Slovenia	74
Estonia	73
Norway	73
Latvia	71
New Zealand	71
Cyprus	69
Slovak Republic	69
Lithuania[3]	67
Serbia[3]	66
Malaysia	65
Bulgaria	60
Russian Federation	60
Italy	58
Romania	57
(Macedonia, Republic of)	54
Armenia	47
Jordan	46
Moldova, Republic of	46
Egypt	43
Iran, Islamic Republic of	43
Philippines	43
Lebanon	42
Palestinian National Authority	41
Bahrain	40
(Morocco)	39
Chile	38
Indonesia[3]	37
Botswana	35
Ghana	34
Saudi Arabia	34
South Africa	34
Tunisia	31

[1]Met international guidelines for participation rates only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]National desired population does not cover all of the international desired population.
NOTE: Countries are sorted by 2003 average percent correct. Parentheses indicate countries that did not meet international sampling or other guidelines. See appendix A for more information. The international average reported here may differ from that reported in Mullis et al. (2004) due to the deletion of England. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA). Trends in International Mathematics and Science Study (TIMSS), 2003.

Exhibit B14. Eighth-grade example item for life science, development, and life cycle of organisms: 2003



Country	Percent full credit
International average	58
Sweden	81
Hungary	77
Hong Kong SAR[1,2]	76
Singapore	76
Japan	74
Armenia	73
Chinese Taipei	72
Estonia	72
Norway	72
(United States)	**70**
Moldova, Republic of	68
Romania	68
Australia	67
Scotland[1]	66
Bulgaria	65
Jordan	65
Russian Federation	65
Chile	64
Italy	64
(Israel)	63
New Zealand	62
Saudi Arabia	62
Serbia[3]	62
Bahrain	60
Korea, Republic of	60
Netherlands[1]	60
Palestinian National Authority	58
Lithuania[3]	57
Slovak Republic	57
Slovenia	57
Cyprus	56
Egypt	55
Malaysia	55
(Morocco)	47
Philippines	45
Botswana	44
Lebanon	42
Tunisia	41
Indonesia[3]	39
Latvia	39
Belgium-Flemish	36
South Africa	34
Ghana	30
Iran, Islamic Republic of	14
(Macedonia, Republic of)	#

#Rounds to zero.
[1]Met international guidelines for participation rates only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]National desired population does not cover all of the international desired population.
NOTE: Countries are sorted by 2003 average percent correct. Parentheses indicate countries that did not meet international sampling or other guidelines. See appendix A for more information. The international average reported here may differ from that reported in Martin et al. (2004) due to the deletion of England. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA). Trends in International Mathematics and Science Study (TIMSS), 2003.

Exhibit B15. Eighth-grade example item for chemistry and chemical change: 2003

Three identical candles are placed in the three jars shown above and lit at the same time. Jars Y and Z are then sealed with lids, and Jar X is left open.

Which candle flame will go out first (X, Y, or Z)? ______________

Explain your answer.

There is less oxygen in Z so the candle will use it up faster and go out.

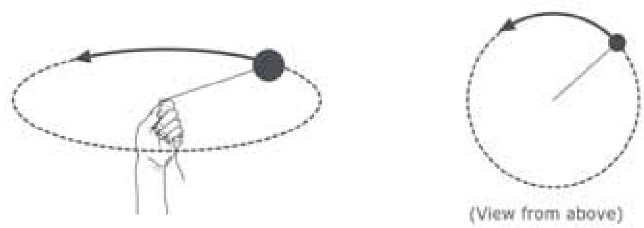
S022191

Country	Percent full credit
International average	46
Netherlands[1]	82
Singapore	78
Sweden	78
Estonia	77
Lithuania[2]	75
Hungary	72
Norway	72
Belgium-Flemish	71
Japan	69
Russian Federation	69
Italy	64
Hong Kong SAR[1,3]	62
Slovenia	62
Chinese Taipei	60
(Israel)	58
Australia	57
Latvia	57
Slovak Republic	55
Scotland[1]	54
New Zealand	53
Korea, Republic of	52
Serbia[3]	48
(United States)	**48**
Lebanon	44
Malaysia	44
Bulgaria	43
(Macedonia, Republic of)	43
Cyprus	42
Romania	41
Tunisia	41
Jordan	37
Egypt	34
Chile	32
Armenia	29
Moldova, Republic of	29
Bahrain	27
Palestinian National Authority	26
(Morocco)	25
Saudi Arabia	23
Iran, Islamic Republic of	19
Indonesia[2]	12
South Africa	9
Philippines	5
Botswana	3
Ghana	1

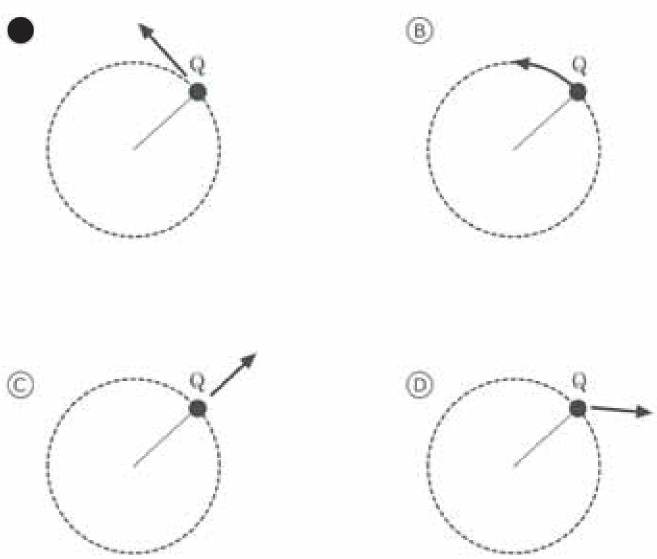
[1]Met international guidelines for participation rates only after replacement schools were included.
[2]National desired population does not cover all of the international desired population.
[3]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
NOTE: Countries are sorted by 2003 average percent correct. Parentheses indicate countries that did not meet international sampling or other guidelines. See appendix A for more information. The international average reported here may differ from that reported in Martin et al. (2004) due to the deletion of England. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA). Trends in International Mathematics and Science Study (TIMSS), 2003.

Exhibit B16. Eighth-grade example item for physics, forces and motion: 2003

The diagram on the left shows a ball on the end of a string being whirled in a circle. The diagram on the right shows the whirling ball as viewed from above.



After several whirls, the string is released when the ball is at Q. Which of these diagrams shows the direction in which the ball will fly the instant the string is released?



S022040

Country	Percent full credit
International average	59
Korea, Republic of	87
Netherlands[1]	82
Estonia	80
Singapore	79
Australia	77
Hungary	77
Japan	77
New Zealand	77
Scotland[1]	77
Belgium-Flemish	76
(United States)	**76**
Lithuania[2]	75
Malaysia	75
Sweden	75
Russian Federation	74
Norway	72
Slovak Republic	72
Latvia	71
Slovenia	70
Hong Kong SAR[1,3]	69
Chinese Taipei	68
Italy	62
Bulgaria	61
Serbia[3]	60
Cyprus	59
Armenia	58
Chile	58
(Israel)	58
Romania	58
(Macedonia, Republic of)	54
Moldova, Republic of	52
Iran, Islamic Republic of	48
Indonesia[2]	47
Jordan	47
Bahrain	44
Philippines	42
Saudi Arabia	38
Palestinian National Authority	36
(Morocco)	33
Tunisia	31
Botswana	30
Egypt	30
Lebanon	30
Ghana	22
South Africa	22

[1]Met international guidelines for participation rates only after replacement schools were included.
[2]National desired population does not cover all of the international desired population.
[3]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
NOTE: Countries are sorted by 2003 average percent correct. Parentheses indicate countries that did not meet international sampling or other guidelines. See appendix A for more information. The international average reported here may differ from that reported in Martin et al. (2004) due to the deletion of England. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA). Trends in International Mathematics and Science Study (TIMSS), 2003.

Exhibit B17. Eighth-grade example item for earth science, earth in the solar system and universe: 2003

The planet Jupiter is bigger than Earth's moon but it appears to be smaller when viewed from Earth. Why is this?

Jupiter is much farther away from Earth than the moon is.

S022283

Country	Percent full credit
International average	62
Netherlands[1]	88
Australia	86
New Zealand	86
(United States)	**85**
Estonia	84
Russian Federation	84
Korea, Republic of	83
Scotland[1]	83
Belgium-Flemish	82
Norway	82
Singapore	81
Sweden	81
Hungary	77
Hong Kong SAR[1,2]	76
Italy	76
Japan	75
Latvia	74
Moldova, Republic of	74
Slovenia	73
Lithuania[3]	71
Armenia	69
Malaysia	69
Chinese Taipei	66
(Israel)	65
Indonesia[3]	64
Slovak Republic	64
Bulgaria	62
Jordan	61
Cyprus	58
Palestinian National Authority	58
Iran, Islamic Republic of	56
Chile	55
Serbia[3]	55
Bahrain	53
Tunisia	50
(Macedonia, Republic of)	45
Romania	45
Egypt	40
Philippines	38
(Morocco)	33
Lebanon	32
Saudi Arabia	31
Botswana	17
South Africa	13
Ghana	8

[1]Met international guidelines for participation rates only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]National desired population does not cover all of the international desired population.
NOTE: Countries are sorted by 2003 average percent correct. Parentheses indicate countries that did not meet international sampling or other guidelines. See appendix A for more information. The international average reported here may differ from that reported in Martin et al. (2004) due to the deletion of England. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA). Trends in International Mathematics and Science Study (TIMSS), 2003.

Exhibit B18. Eighth-grade example item for environmental science, changes in environment: 2003

The burning of fossil fuels has increased the carbon dioxide content of the atmosphere. What is a possible effect that the increased amount of carbon dioxide is likely to have on our planet?

- ● A warmer climate
- Ⓑ A cooler climate
- Ⓒ Lower relative humidity
- Ⓓ More ozone in the atmosphere

S012017

Country	Percent full credit
International average	44
Singapore	83
Japan	80
Hong Kong SAR[1,2]	72
Netherlands[1]	71
Malaysia	68
Sweden	66
Korea, Republic of	65
Australia	64
Chinese Taipei	62
Norway	62
Scotland[1]	62
Estonia	58
Hungary	56
New Zealand	56
(United States)	**56**
Indonesia[3]	52
(Israel)	51
Italy	49
Latvia	47
Moldova, Republic of	46
Belgium-Flemish	45
Iran, Islamic Republic of	45
Russian Federation	44
Slovenia	44
Bulgaria	43
Slovak Republic	43
Cyprus	42
Chile	40
Romania	40
Armenia	35
(Macedonia, Republic of)	35
Lithuania[3]	34
(Morocco)	32
Philippines	32
Serbia[3]	30
Botswana	27
South Africa	23
Ghana	21
Jordan	21
Lebanon	21
Bahrain	18
Egypt	17
Tunisia	17
Palestinian National Authority	15
Saudi Arabia	12

[1]Met international guidelines for participation rates only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]National desired population does not cover all of the international desired population.
NOTE: Countries are sorted by 2003 average percent correct. Parentheses indicate countries that did not meet international sampling or other guidelines. See appendix A for more information. The international average reported here may differ from that reported in Martin et al. (2004) due to the deletion of England. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA). Trends in International Mathematics and Science Study (TIMSS), 2003.

Appendix C: Detailed Tables

Table C1. Average mathematics and science scale scores of fourth-grade students, by country: 2003

Country	Mathematics		Science	
	Scale score	s.e.	Scale score	s.e.
International average	495	0.8	489	0.9
Armenia	456	3.5	437	4.3
Australia[1]	499	3.9	521	4.2
Belgium-Flemish	551	1.8	518	1.8
Chinese Taipei	564	1.8	551	1.7
Cyprus	510	2.4	480	2.4
England[1]	531	3.7	540	3.6
Hong Kong SAR[1,2]	575	3.2	542	3.1
Hungary	529	3.1	530	3.0
Iran, Islamic Republic of	389	4.2	414	4.1
Italy	503	3.7	516	3.8
Japan	565	1.6	543	1.5
Latvia	536	2.8	532	2.5
Lithuania	534	2.8	512	2.6
Moldova	504	4.9	496	4.6
Morocco	347	5.1	304	6.7
Netherlands[1]	540	2.1	525	2.0
New Zealand	493	2.2	520	2.5
Norway	451	2.3	466	2.6
Philippines	358	7.9	332	9.4
Russian Federation	532	4.7	526	5.2
Scotland[1]	490	3.3	502	2.9
Singapore	594	5.6	565	5.5
Slovenia	479	2.6	490	2.5
Tunisia	339	4.7	314	5.7
United States[1]	518	2.4	536	2.5

[1]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]National desired population does not cover all of the international desired population.
NOTE: Countries were required to sample students in the upper of the two grades that contained the most number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details. s.e. means standard error.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Table C2. Average mathematics and science scale scores of eighth-grade students, by country: 2003

Country	Mathematics		Science	
	Scale score	s.e.	Scale score	s.e.
International average	466	0.5	473	0.5
Armenia	478	3.0	461	3.5
Australia	505	4.6	527	3.8
Bahrain	401	1.7	438	1.8
Belgium-Flemish	537	2.8	516	2.5
Botswana	366	2.6	365	2.8
Bulgaria	476	4.3	479	5.2
Chile	387	3.3	413	2.9
Chinese Taipei	585	4.6	571	3.5
Cyprus	459	1.7	441	2.0
Egypt	406	3.5	421	3.9
Estonia	531	3.0	552	2.5
Ghana	276	4.7	255	5.9
Hong Kong SAR[1,2]	586	3.3	556	3.0
Hungary	529	3.2	543	2.8
Indonesia[3]	411	4.8	420	4.1
Iran, Islamic Republic of	411	2.4	453	2.3
(Israel)	496	3.4	488	3.1
Italy	484	3.2	491	3.1
Japan	570	2.1	552	1.7
Jordan	424	4.1	475	3.8
Korea, Republic of	589	2.2	558	1.6
Latvia	505	3.8	513	2.9
Lebanon	433	3.1	393	4.3
Lithuania[3]	502	2.5	519	2.1
(Macedonia, Republic of)	435	3.5	449	3.6
Malaysia	508	4.1	510	3.7
Moldova, Republic of	460	4.0	472	3.4
(Morocco)	387	2.5	396	2.5
Netherlands[1]	536	3.8	536	3.1
New Zealand	494	5.3	520	5.0
Norway	461	2.5	494	2.2
Palestinian National Authority	390	3.1	435	3.2
Philippines	378	5.2	377	5.8
Romania	475	4.8	470	4.9
Russian Federation	508	3.7	514	3.7
Saudi Arabia	332	4.6	398	4.0
Scotland[1]	498	3.7	512	3.4
Serbia[3]	477	2.6	468	2.5
Singapore	605	3.6	578	4.3
Slovak Republic	508	3.3	517	3.2
Slovenia	493	2.2	520	1.8
South Africa	264	5.5	244	6.7
Sweden	499	2.6	524	2.7
Tunisia	410	2.2	404	2.1
(United States)	504	3.3	527	3.1

[1]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]National desired population does not cover all of the international desired population.
NOTE: Parentheses indicate countries that did not meet international sampling or other guidelines in 2003. See appendix A for details regarding 2003 data. The estimates for the international average reported here may differ from that reported in Martin et al. (2004) and Mullis et al. (2004) due to the deletion of England. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details. s.e. means standard error.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Table C3. Average mathematics scale scores of fourth-grade students, by country: 1995 and 2003

Country	1995		2003	
	Scale score	s.e.	Scale score	s.e.
(Australia)[1]	495	3.4	499	3.9
Cyprus	475	3.2	510	2.4
England[1]	484	3.3	531	3.7
Hong Kong SAR[1,2]	557	4.0	575	3.2
(Hungary)	521	3.6	529	3.1
Iran, Islamic Republic of	387	5.0	389	4.2
Japan	567	1.9	565	1.6
(Latvia–LSS)[3]	499	4.6	533	3.1
(Netherlands)[1]	549	3.0	540	2.1
New Zealand[4]	469	4.4	496	2.1
Norway	476	3.0	451	2.3
Scotland[1]	493	4.2	490	3.3
Singapore	590	4.5	594	5.6
(Slovenia)	462	3.1	479	2.6
United States[1]	518	2.9	518	2.4

[1]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]Designated LSS because only Latvian-speaking schools were included in 1995. For this analysis, only Latvian-speaking schools are included in the 2003 average.
[4]In 1995, Maori-speaking students did not participate. Estimates in this table are computed for students taught in English only, which represents between 98-99 percent of the student population in both years.
NOTE: Parentheses indicate countries that did not meet international sampling or other guidelines in 1995. All countries met international sampling and other guidelines in 2003, except as noted. See NCES (1997) for details regarding 1995 data. Countries were required to sample students in the upper of the two grades that contained the most number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details. s.e. means standard error.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2003.

Table C4. Average mathematics scale scores of eighth-grade students, by country: 1995, 1999, and 2003

Country	1995		1999		2003	
	Scale score	s.e.	Scale score	s.e.	Scale score	s.e.
(Australia)[1]	509	3.7	—	—	505	4.6
Belgium-Flemish	550	5.9	558	3.3	537	2.8
(Bulgaria)	527	5.8	511	5.8	476	4.3
Chile	—	—	392	4.4	387	3.3
Chinese Taipei	—	—	585	4.0	585	4.6
Cyprus	468	2.2	476	1.8	459	1.7
Hong Kong SAR[2,3]	569	6.1	582	4.3	586	3.3
Hungary	527	3.2	532	3.7	529	3.2
Indonesia[4]	—	—	403	4.9	411	4.8
Iran, Islamic Republic of	418	3.9	422	3.4	411	2.4
(Israel)[5]	—	—	466	3.9	496	3.4
Italy[5]	—	—	479	3.8	484	3.2
Japan	581	1.6	579	1.7	570	2.1
Jordan	—	—	428	3.6	424	4.1
Korea, Republic of	581	2.0	587	2.0	589	2.2
(Latvia-LSS)[6]	488	3.6	505	3.4	505	3.8
Lithuania[4]	472	4.1	482	4.3	502	2.5
(Macedonia, Republic of)	—	—	447	4.2	435	3.5
Malaysia	—	—	519	4.4	508	4.1
Moldova, Republic of	—	—	469	3.9	460	4.0
(Netherlands)[2]	529	6.1	540	7.1	536	3.8
New Zealand	501	4.7	491	5.2	494	5.3
Norway	498	2.2	—	—	461	2.5
Philippines	—	—	345	6.0	378	5.2
(Romania)	474	4.6	472	5.8	475	4.8
Russian Federation	524	5.3	526	5.9	508	3.7
(Scotland)[2]	493	5.7	—	—	498	3.7
Singapore	609	4.0	604	6.3	605	3.6
Slovak Republic	534	3.1	534	4.0	508	3.3
(Slovenia)[1]	494	2.9	—	—	493	2.2
South Africa[7]	—	—	275	6.8	264	5.5
Sweden	540	4.3	—	—	499	2.6
Tunisia	—	—	448	2.4	410	2.2
(United States)	492	4.7	502	4.0	504	3.3

—Not available.

[1]Because of national-level changes in the starting age/date for school, 1999 data for Australia and Slovenia cannot be compared to 2003.

[2]Met international guidelines for participation rates in 2003 only after replacement schools were included.

[3]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

[4]National desired population does not cover all of the international desired population in all years for Lithuania, and in 2003 for Indonesia.

[5]Because of changes in the population tested, 1995 data for Israel and Italy are not shown.

[6]Designated LSS because only Latvian-speaking schools were included in 1995 and 1999. For this analysis, only Latvian-speaking schools are included in the 2003 average.

[7]Because within classroom sampling was not accounted for, 1995 data are not shown for South Africa.

NOTE: Parentheses indicate countries that did not meet international sampling or other guidelines in 1995, 1999, or 2003. See appendix A for details regarding 2003 data. See Gonzales et al. (2000) for details regarding 1995 and 1999 data. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details. s.e. means standard error.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995, 1999, and 2003.

Table C5. Average mathematics scale scores of eighth-grade students, by country: 1995 and 2003

Country	1995		2003	
	Scale score	s.e.	Scale score	s.e.
(Australia)	509	3.7	505	4.6
Belgium-Flemish	550	5.9	537	2.8
(Bulgaria)	527	5.8	476	4.3
Cyprus	468	2.2	459	1.7
Hong Kong SAR[1,2]	569	6.1	586	3.3
Hungary	527	3.2	529	3.2
Iran, Islamic Republic of	418	3.9	411	2.4
Japan	581	1.6	570	2.1
Korea, Republic of	581	2.0	589	2.2
(Latvia-LSS)[3]	488	3.6	505	3.8
(Lithuania)[4]	472	4.1	502	2.5
(Netherlands)[1]	529	6.1	536	3.8
New Zealand	501	4.7	494	5.3
Norway	498	2.2	461	2.5
(Romania)	474	4.6	475	4.8
Russian Federation	524	5.3	508	3.7
Scotland[1]	493	5.7	498	3.7
Singapore	609	4.0	605	3.6
Slovak Republic	534	3.1	508	3.3
(Slovenia)	494	2.9	493	2.2
Sweden	540	4.3	499	2.6
(United States)	492	4.7	504	3.3

[1]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]Designated LSS because only Latvian-speaking schools were included in 1995. For this analysis, only Latvian-speaking schools are included in the 2003 average.
[4]National desired population does not cover all of the international desired population in all years for Lithuania.
NOTE: Parentheses indicate countries that did not meet international sampling or other guidelines in 1995 or 2003. See appendix A for details regarding 2003 data. See Gonzales et al. (2000) for details regarding 1995 data. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details. s.e. means standard error.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2003.

Table C6. Percent correct of eighth-grade students in five mathematics content areas, by country: 1999 and 2003

Country	All mathematics trend items				Mathematics content area: Number				Mathematics content area: Algebra			
	1999		2003		1999		2003		1999		2003	
	Percent	s.e.	Percent	s.e.	Percent	s.e.	Percent	s.e.	Percent	s.e.	Percent	s.e.
Belgium-Flemish	64	0.8	60	0.7	64	1.0	61	0.8	56	1.0	52	0.8
Bulgaria	53	1.5	45	1.0	54	1.5	47	1.0	53	1.6	43	1.1
Chile	29	0.8	29	0.6	32	0.9	31	0.6	24	0.9	23	0.7
Chinese Taipei	70	0.9	69	1.0	73	0.9	70	1.1	68	1.1	66	1.2
Cyprus	46	0.4	43	0.4	49	0.5	46	0.5	40	0.7	38	0.6
Hong Kong SAR[1,2]	71	1.1	70	0.7	71	1.2	69	0.8	69	1.3	68	0.9
Hungary	59	0.8	57	0.9	60	0.9	59	1.0	57	0.9	56	1.0
Indonesia[3]	34	0.8	32	0.8	36	0.8	35	0.9	32	0.9	30	0.8
Iran, Islamic Republic of	35	0.7	32	0.5	39	0.7	36	0.5	31	0.8	29	0.6
(Israel)	43	0.9	50	0.9	44	0.9	52	0.9	42	1.1	48	0.9
Italy	48	0.9	47	0.9	49	0.9	48	0.9	41	0.9	42	1.1
Japan	70	0.5	66	0.6	70	0.6	65	0.7	69	0.7	64	0.7
Jordan	36	0.6	33	0.8	38	0.7	35	0.8	33	0.8	31	0.9
Korea, Republic of	71	0.5	72	0.5	72	0.5	73	0.6	68	0.7	71	0.6
Latvia-LSS[4]	51	0.8	51	1.0	53	0.9	53	1.1	47	0.9	48	1.2
Lithuania[3]	47	1.0	50	0.7	50	1.1	51	0.7	44	1.2	46	0.8
(Macedonia, Republic of)	38	0.8	36	0.7	37	0.9	38	0.8	38	1.0	35	0.9
Malaysia	56	1.2	52	1.1	62	1.2	57	1.1	46	1.0	42	1.0
Moldova, Republic of	44	1.0	43	0.9	46	1.1	47	1.0	41	1.0	40	1.0
Netherlands[1]	58	2.0	60	1.0	58	2.1	60	1.0	51	2.3	51	1.1
New Zealand	47	1.3	48	1.2	47	1.3	47	1.2	43	1.4	43	1.4
Philippines	25	0.7	27	0.8	30	0.8	31	0.8	20	0.9	27	1.0
Romania	46	1.3	45	1.2	46	1.4	46	1.1	44	1.5	44	1.4
Russian Federation	55	1.3	53	1.0	57	1.4	54	1.1	54	1.3	52	1.0
Singapore	76	1.4	74	1.0	80	1.2	78	0.9	69	1.6	69	1.1
Slovak Republic	59	1.1	52	0.9	62	1.2	55	1.0	55	1.3	49	1.0
South Africa	19	0.7	18	0.7	22	0.7	20	0.7	15	0.7	14	0.7
Tunisia	39	0.5	30	0.4	41	0.5	33	0.5	33	0.6	26	0.5
(United States)	50	0.9	51	0.9	54	1.0	54	0.9	47	1.0	50	1.0

See notes at end of table.

Table C6. Percent correct of eighth-grade students in five mathematics content areas, by country: 1999 and 2003 —Continued

Country	Mathematics content area											
	Measurement				Geometry				Data			
	1999		2003		1999		2003		1999		2003	
	Percent	s.e.	Percent	s.e.	Percent	s.e.	Percent	s.e.	Percent	s.e.	Percent	s.e.
Belgium-Flemish	60	0.8	54	0.8	64	1.0	61	0.9	81	0.8	79	0.7
Bulgaria	45	1.5	35	1.2	58	1.6	50	0.9	62	1.6	58	1.1
Chile	19	0.8	21	0.6	32	0.9	30	0.7	45	1.0	44	1.0
Chinese Taipei	64	1.0	61	1.1	72	0.9	71	1.0	80	0.7	79	0.8
Cyprus	40	0.6	34	0.6	47	0.6	45	0.5	61	1.0	61	0.7
Hong Kong SAR[1,2]	66	1.2	66	0.9	72	1.1	73	0.8	78	0.9	76	0.6
Hungary	53	1.0	51	1.0	55	1.1	55	1.0	71	0.9	69	1.0
Indonesia[3]	22	0.8	21	0.8	37	1.0	36	0.8	47	1.1	47	1.1
Iran, Islamic Republic of	22	0.8	20	0.5	39	0.8	36	0.6	49	1.0	46	0.8
(Israel)	32	0.9	39	0.9	44	0.9	51	1.1	59	1.1	65	1.1
Italy	44	1.0	43	1.0	47	1.0	46	1.0	64	1.2	64	0.9
Japan	63	0.7	58	0.7	75	0.6	74	0.6	79	0.5	76	0.5
Jordan	27	0.8	23	0.8	41	0.7	37	0.8	49	0.7	46	1.1
Korea, Rep. of	64	0.6	63	0.7	74	0.6	75	0.6	82	0.4	80	0.4
Latvia-LSS[4]	40	1.1	38	1.0	59	1.0	57	1.2	63	1.0	67	1.4
Lithuania[3]	34	1.2	38	0.8	49	1.3	54	0.8	64	1.2	68	0.8
(Macedonia, Republic of)	29	1.0	27	0.9	42	1.0	39	0.7	48	1.0	49	1.0
Malaysia	51	1.4	45	1.3	53	1.3	51	1.2	68	1.0	67	1.0
Moldova, Republic of	37	1.3	36	1.1	47	1.2	46	1.3	50	1.1	49	1.0
Netherlands[1]	56	2.0	58	1.2	58	1.7	57	1.2	75	2.4	79	1.0
New Zealand	42	1.5	42	1.5	48	1.3	49	1.3	65	1.4	66	1.4
Philippines	15	0.6	18	0.8	25	0.8	25	0.7	39	0.9	40	0.9
Romania	40	1.4	39	1.4	48	1.3	45	1.3	54	1.3	55	1.4
Russian Federation	47	1.6	44	1.2	58	1.5	56	1.1	65	1.3	64	1.2
Singapore	76	1.6	74	1.1	73	1.6	71	1.1	81	1.2	79	0.8
Slovak Republic	53	1.5	44	1.1	61	1.2	53	1.0	71	1.1	64	1.0
South Africa	13	0.6	12	0.7	21	0.8	19	0.8	30	0.9	29	1.1
Tunisia	32	0.7	20	0.5	46	0.6	34	0.6	52	0.7	39	0.6
(United States)	40	1.1	42	1.0	44	1.0	45	0.9	68	0.9	72	0.8

[1]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]National desired population does not cover all of the international desired population in all years for Lithuania, and in 2003 for Indonesia.
[4]Designated LSS because only Latvian-speaking schools were included in 1999. For this analysis, only Latvian-speaking schools are included in the 2003 average.
NOTE: Parentheses indicate countries that did not meet international sampling or other guidelines in 1999 or 2003. See appendix A for details regarding 2003 data. See Gonzales et al. (2000) for details regarding 1999 data. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details. s.e. means standard error.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1999 and 2003.

Table C7. Average mathematics scale scores of fourth-grade students, by sex and country: 1995 and 2003

Country	Boys				Girls			
	1995		2003		1995		2003	
	Scale score	s.e.	Scale score	s.e.	Scale score	s.e.	Scale score	s.e.
(Australia)[1]	496	4.1	500	4.3	493	3.9	497	4.5
Cyprus	479	3.8	514	2.9	471	3.5	505	2.7
England[1]	488	3.7	532	4.5	480	4.3	530	3.9
Hong Kong SAR[1,2]	557	4.4	575	3.4	558	3.9	575	3.4
(Hungary)	524	4.0	530	3.3	519	4.0	527	3.8
Iran, Islamic Republic of	394	8.0	386	5.5	379	6.0	394	6.5
Japan	571	2.4	566	2.1	563	2.0	563	1.8
(Latvia–LSS)[3]	493	5.6	531	3.9	505	5.1	535	3.2
(Netherlands)[1]	556	3.5	543	2.2	543	3.3	537	2.7
New Zealand[4]	465	6.1	496	2.4	474	4.3	495	2.8
Norway	478	3.6	454	2.7	474	4.3	449	2.7
Scotland[1]	493	4.7	496	4.4	493	4.2	485	3.2
Singapore	586	4.7	590	6.2	595	5.5	599	5.5
(Slovenia)	466	3.5	481	3.5	457	3.8	477	3.0
United States[1]	520	3.1	522	2.7	516	3.0	514	2.4

[1]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]Designated LSS because only Latvian-speaking schools were included in 1995. For the purposes of this analysis, only Latvian-speaking schools were included in the 2003 average.
[4]In 1995, Maori-speaking students did not participate. Estimates in this table are computed for students taught in English only, which represents between 98-99 percent of the student population in both years.
NOTE: Parentheses indicate countries that did not meet international sampling or other guidelines in 1995. All countries met international sampling and other guidelines in 2003, except as noted. See NCES (1997) for details regarding 1995 data. Countries were required to sample students in the upper of the two grades that contained the most number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details. s.e. means standard error.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2003.

Table C8. Average mathematics scale scores of U.S. fourth-grade students, by selected characteristics: 1995 and 2003

Selected characteristics	1995		2003	
	Scale score	s.e.	Scale score	s.e.
Race/ethnicity[1]				
White	541	3.5	542	2.2
Black	457	4.4	472	3.4
Hispanic	493	5.7	492	3.6
Asian	‡	‡	551	8.1
Poverty level in public schools (percentage of students eligible for free or reduced-price lunch)				
Less than 10 percent	—	—	567	5.2
10 to 24.9 percent	—	—	543	3.6
25 to 49.9 percent	—	—	533	4.0
50 to 74.9 percent	—	—	500	3.0
75 percent or more	—	—	471	4.3

—Not available.
‡Reporting standards not met.
[1]Other race/ethnicities are included in the U.S. totals shown throughout the report but not shown separately. Racial categories exclude Hispanic origin.
NOTE: The United States met international sampling guidelines for participation rates in 2003 only after replacement schools were included. s.e. means standard error.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2003.

Table C9. Standard deviations of mathematics and science scores of fourth-grade students, by country: 2003

Country	Mathematics		Science	
	Standard deviation	s.e.	Standard deviation	s.e.
Armenia	87	1.9	96	2.2
Australia[1]	81	2.1	82	2.6
Belgium-Flemish	59	1.1	55	1.0
Chinese Taipei	63	1.1	69	1.3
Cyprus	85	1.3	74	1.3
England[1]	87	1.9	83	2.2
Hong Kong SAR[1,2]	63	1.5	60	1.2
Hungary	77	2.0	79	1.8
Iran, Islamic Republic of	86	2.1	97	2.4
Italy	82	2.2	85	1.9
Japan	74	1.0	73	1.2
Latvia	73	1.5	69	1.5
Lithuania[3]	74	1.7	66	1.5
Moldova	87	3.2	85	3.0
Morocco	90	1.9	125	2.7
Netherlands[1]	55	1.5	53	1.1
New Zealand	84	1.8	85	2.0
Norway	80	1.6	84	1.6
Philippines	110	5.9	145	5.7
Russian Federation	78	2.0	82	2.3
Scotland[1]	78	1.8	78	1.9
Singapore	84	3.2	87	3.3
Slovenia	78	1.3	77	1.4
Tunisia	100	2.5	126	2.6
United States[1]	76	1.0	81	1.1

[1]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]National desired population does not cover all of the international desired population.
NOTE: Countries were required to sample students in the upper of the two grades that contained the most number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details. s.e. means standard error.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Table C10. Average mathematics scale scores of eighth-grade students, by sex and country: 1995, 1999, and 2003

Country	Boys					
	1995		1999		2003	
	Scale score	s.e.	Scale score	s.e.	Scale score	s.e.
(Australia)[1]	507	4.7	—	—	511	5.8
Belgium-Flemish	547	8.7	556	8.3	542	3.8
(Bulgaria)	521	6.2	511	6.9	477	4.3
Chile	—	—	397	5.8	394	4.3
Chinese Taipei	—	—	587	5.3	582	5.2
Cyprus	465	3.3	474	2.7	452	2.3
Hong Kong SAR[2,3]	577	7.2	581	5.9	585	4.6
Hungary	527	3.6	535	4.3	533	3.5
Indonesia[4]	—	—	405	5.0	410	5.3
Iran, Islamic Republic of	429	4.7	432	4.8	408	4.2
(Israel)[5]	—	—	474	4.8	500	4.5
Italy[5]	—	—	484	4.3	486	3.9
Japan	585	2.2	582	2.3	571	3.6
Jordan	—	—	425	5.9	411	5.8
Korea, Republic of	588	2.7	590	2.2	592	2.6
(Latvia-LSS)[6]	490	4.2	508	4.4	502	4.4
Lithuania[4]	472	4.6	483	4.8	499	3.0
(Macedonia, Republic of)	—	—	447	4.3	431	3.9
Malaysia	—	—	517	6.0	505	4.5
Moldova, Republic of	—	—	471	4.7	455	4.8
(Netherlands)[2]	534	6.6	542	7.0	540	4.5
New Zealand	505	6.1	487	7.6	493	7.0
Norway	499	2.9	—	—	460	3.0
Philippines	—	—	337	6.5	370	5.8
(Romania)	475	5.3	470	6.2	473	5.0
Russian Federation	523	6.2	526	6.4	507	4.4
(Scotland)[2]	501	7.0	—	—	495	3.8
Singapore	608	4.7	606	7.5	601	4.3
Slovak Republic	536	3.7	536	4.5	508	4.0
(Slovenia)[1]	497	3.5	—	—	491	2.6
South Africa[7]	—	—	283	7.3	264	6.4
Sweden	539	4.7	—	—	499	2.7
Tunisia	—	—	460	2.9	423	2.2
(United States)	495	5.2	505	4.8	507	3.5

See notes at end of table.

Table C10. Average mathematics scale scores of eighth-grade students, by sex and country: 1995, 1999, and 2003—Continued

Country	Girls					
	1995		1999		2003	
	Scale score	s.e.	Scale score	s.e.	Scale score	s.e.
(Australia)[1]	511	4.1	—	—	499	5.8
Belgium-Flemish	553	8.1	560	7.2	532	3.5
(Bulgaria)	532	6.1	510	5.9	476	5.5
Chile	—	—	388	4.3	379	3.5
Chinese Taipei	—	—	583	3.9	589	4.9
Cyprus	471	2.6	479	2.1	467	1.9
Hong Kong SAR[2,3]	559	7.0	583	4.7	587	3.8
Hungary	527	3.6	529	4.0	526	3.7
Indonesia[4]	—	—	401	5.4	411	4.9
Iran, Islamic Republic of	405	6.1	408	4.2	417	4.3
(Israel)[5]	—	—	459	4.2	492	3.3
Italy[5]	—	—	475	4.5	481	3.0
Japan	577	1.9	575	2.4	569	4.0
Jordan	—	—	431	4.7	438	4.6
Korea, Republic of	571	3.0	585	3.1	586	2.7
(Latvia-LSS)[6]	486	4.0	502	3.8	509	4.0
Lithuania[4]	472	4.6	480	4.7	503	2.9
(Macedonia, Republic of)	—	—	446	5.3	439	4.0
Malaysia	—	—	521	4.7	512	4.7
Moldova, Republic of	—	—	468	4.1	465	4.1
(Netherlands)[2]	522	6.6	538	7.6	533	4.1
New Zealand	497	5.3	495	5.5	495	4.8
Norway	498	2.6	—	—	463	2.7
Philippines	—	—	352	6.9	383	5.2
(Romania)	473	4.4	475	6.3	477	5.1
Russian Federation	524	5.0	526	6.0	510	3.5
(Scotland)[2]	486	5.4	—	—	500	4.3
Singapore	610	4.9	603	6.1	611	3.3
Slovak Republic	532	3.1	532	4.2	508	3.4
(Slovenia)[1]	492	2.9	—	—	495	2.6
South Africa[7]	—	—	267	7.5	262	6.2
Sweden	541	4.6	—	—	499	3.0
Tunisia	—	—	436	2.4	399	2.6
(United States)	490	4.7	498	3.9	502	3.4

—Not available.
[1]Because of national-level changes in the starting age/date for school, 1999 data for Australia and Slovenia cannot be compared to 2003.
[2]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[3]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[4]National desired population does not cover all of the international desired population in all years for Lithuania, and in 2003 for Indonesia.
[5]Because of changes in the population tested, 1995 data for Israel and Italy are not shown.
[6]Designated LSS because only Latvian-speaking schools were included in 1995 and 1999. For this analysis, only Latvian-speaking schools are included in the 2003 average.
[7]Because within classroom sampling was not accounted for, 1995 data are not shown for South Africa.
NOTE: Parentheses indicate countries that did not meet international sampling or other guidelines in 1995, 1999, or 2003. See appendix A for details regarding 2003 data. See Gonzales et al. (2000) for details regarding 1995 and 1999 data. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details. s.e. means standard error.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995, 1999, and 2003.

Table C11. Average mathematics scale scores of U.S. eighth-grade students, by selected characteristics: 1995, 1999, and 2003

Selected characteristics	1995		1999		2003	
	Scale score	s.e.	Scale score	s.e.	Scale score	s.e.
Race/ethnicity[1]						
White	516	3.5	525	4.6	525	3.0
Black	419	6.8	444	5.3	448	5.2
Hispanic or Latino	443	3.8	457	6.3	465	5.4
Poverty level in public schools (percentage of students eligible for free or reduced-price lunch)						
Less than 10 percent	—	—	562	13.9	547	7.3
10 to 24.9 percent	—	—	535	3.1	531	7.4
25 to 49.9 percent	—	—	495	7.5	505	5.2
50 to 74.9 percent	—	—	476	6.6	480	5.1
75 percent or more	—	—	448	11.1	444	10.4

—Not available.

[1]Other race/ethnicities are included in the U.S. totals shown throughout the report but not shown separately. Racial categories exclude Hispanic origin.

NOTE: The United States did not meet international sampling guidelines in 2003. See appendix A for more information. s.e. means standard error.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995, 1999, and 2003.

Table C12. Standard deviations of mathematics and science scores of eighth-grade students, by country: 2003

Country	Mathematics		Science	
	Standard deviation	s.e.	Standard deviation	s.e.
Armenia	84	1.4	81	1.7
Australia	82	3.2	75	2.0
Bahrain	76	0.9	74	1.0
Belgium-Flemish	73	2.6	67	2.2
Botswana	72	1.5	86	2.2
Bulgaria	84	2.3	93	3.7
Chile	83	1.9	84	1.5
Chinese Taipei	100	2.2	79	1.7
Cyprus	81	1.3	79	1.3
Egypt	93	1.5	104	1.8
Estonia	69	1.6	65	1.3
Ghana	91	2.3	120	2.2
Hong Kong SAR[1,2]	72	3.2	66	2.9
Hungary	80	2.3	76	1.7
Indonesia[3]	89	2.6	79	2.3
Iran, Islamic Republic of	74	1.4	73	1.2
(Israel)	85	1.8	85	1.6
Italy	77	1.8	78	1.9
Japan	80	1.3	71	1.1
Jordan	89	1.8	89	1.7
Korea, Republic of	84	1.3	70	1.2
Latvia	73	1.4	67	1.2
Lebanon	67	1.6	93	2.3
Lithuania[3]	78	1.3	70	1.2
(Macedonia, Republic of)	88	2.3	92	2.3
Malaysia	74	2.2	66	1.9
Moldova, Republic of	81	1.7	74	1.4
(Morocco)	68	1.0	69	1.2
Netherlands[1]	69	2.8	61	2.5
New Zealand	78	3.6	74	3.1
Norway	71	1.3	70	1.2
Palestinian National Authority	92	1.5	92	1.7
Philippines	87	2.6	102	2.4
Romania	90	1.7	91	1.9
Russian Federation	77	1.4	75	1.7
Saudi Arabia	78	2.6	72	1.5
Scotland[1]	75	2.3	76	1.5
Serbia[3]	89	1.4	84	1.2
Singapore	80	2.4	92	3.1
Slovak Republic	82	1.7	76	1.3
Slovenia	71	1.5	67	1.7
South Africa	107	5.1	132	5.5
Sweden	71	1.7	74	1.5
Tunisia	60	1.3	60	1.0
(United States)	80	1.8	81	1.6

[1]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]National desired population does not cover all of the international desired population.
NOTE: Parentheses indicate countries that did not meet international sampling or other guidelines in 2003. See appendix A for details regarding 2003 data. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details. s.e. means standard error.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Table C13. Average science scale scores of fourth-grade students, by country: 1995 and 2003

Country	1995		2003	
	Scale score	s.e.	Scale score	s.e.
(Australia)[1]	521	3.8	521	4.2
Cyprus	450	3.2	480	2.4
England[1]	528	3.1	540	3.6
Hong Kong SAR[1,2]	508	3.3	542	3.1
(Hungary)	508	3.4	530	3.0
Iran, Islamic Republic of	380	4.6	414	4.1
Japan	553	1.8	543	1.5
(Latvia–LSS)[3]	486	4.9	530	2.8
(Netherlands)[1]	530	3.2	525	2.0
New Zealand[4]	505	5.3	523	2.3
Norway	504	3.7	466	2.6
Scotland[1]	514	4.5	502	2.9
Singapore	523	4.8	565	5.5
(Slovenia)	464	3.1	490	2.5
United States[1]	542	3.3	536	2.5

[1]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]Designated LSS because only Latvian-speaking schools were included in 1995. For the purposes of this analysis, only Latvian-speaking schools were included in the 2003 average.
[4]In 1995, Maori-speaking students did not participate. Estimates in this table are computed for students taught in English only, which represents between 98-99 percent of the student population in both years.
NOTE: Parentheses indicate countries that did not meet international sampling or other guidelines in 1995. All countries met international sampling and other guidelines in 2003, except as noted. See NCES (1997) for details regarding 1995 data. Countries were required to sample students in the upper of the two grades that contained the most number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details. s.e. means standard error.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2003.

Table C14. Average science scale scores of eighth-grade students, by country: 1995, 1999, and 2003

Country	1995		1999		2003	
	Scale score	s.e.	Scale score	s.e.	Scale score	s.e.
(Australia)[1]	514	3.9	—	—	527	3.8
Belgium-Flemish	533	6.4	535	3.1	516	2.5
(Bulgaria)	545	5.2	518	5.4	479	5.2
Chile	—	—	420	3.7	413	2.9
Chinese Taipei	—	—	569	4.4	571	3.5
Cyprus	452	2.1	460	2.4	441	2.0
Hong Kong SAR[2,3]	510	5.8	530	3.7	556	3.0
Hungary	537	3.1	552	3.7	543	2.8
Indonesia[4]	—	—	435	4.5	420	4.1
Iran, Islamic Republic of	463	3.6	448	3.8	453	2.3
(Israel)[5]	—	—	468	4.9	488	3.1
Italy[5]	—	—	493	3.9	491	3.1
Japan	554	1.8	550	2.2	552	1.7
Jordan	—	—	450	3.8	475	3.8
Korea, Republic of	546	2.0	549	2.6	558	1.6
(Latvia-LSS)[6]	476	3.3	503	4.8	513	2.9
Lithuania[4]	464	4.0	488	4.1	519	2.1
(Macedonia, Republic of)	—	—	458	5.2	449	3.6
Malaysia	—	—	492	4.4	510	3.7
Moldova, Republic of	—	—	459	4.0	472	3.4
(Netherlands)[2]	541	6.0	545	6.9	536	3.1
New Zealand	511	4.9	510	4.9	520	5.0
Norway	514	2.4	—	—	494	2.2
Philippines	—	—	345	7.5	377	5.8
(Romania)	471	5.1	472	5.8	470	4.9
Russian Federation	523	4.5	529	6.4	514	3.7
(Scotland)[2]	501	5.6	—	—	512	3.4
Singapore	580	5.5	568	8.0	578	4.3
Slovak Republic	532	3.3	535	3.3	517	3.2
(Slovenia)[1]	514	2.7	—	—	520	1.8
South Africa[7]	—	—	243	7.8	244	6.7
Sweden	553	4.4	—	—	524	2.7
Tunisia	—	—	430	3.4	404	2.1
(United States)	513	5.6	515	4.6	527	3.1

—Not available.
[1]Because of national-level changes in the starting age/date for school, 1999 data for Australia and Slovenia cannot be compared to 2003.
[2]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[3]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[4]National desired population does not cover all of the international desired population in all years for Lithuania, and in 2003 for Indonesia.
[5]Because of changes in the population tested, 1995 data for Israel and Italy are not shown.
[6]Designated LSS because only Latvian-speaking schools were included in 1995 and 1999. For this analysis, only Latvian-speaking schools are included in the 2003 average.
[7]Because within classroom sampling was not accounted for, 1995 data are not shown for South Africa.
NOTE: Parentheses indicate countries that did not meet international sampling or other guidelines in 1995, 1999, or 2003. See appendix A for details regarding 2003 data. See Gonzales et al. (2000) for details regarding 1995 and 1999 data. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details. s.e. means standard error.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995, 1999, and 2003.

Table C15. Average science scale scores of eighth-grade students, by country: 1995 and 2003

Country	1995		2003	
	Scale score	s.e.	Scale score	s.e.
(Australia)	514	3.9	527	3.8
Belgium-Flemish	533	6.4	516	2.5
(Bulgaria)	545	5.2	479	5.2
Cyprus	452	2.1	441	2.0
Hong Kong SAR[1,2]	510	5.8	556	3.0
Hungary	537	3.1	543	2.8
Iran, Islamic Republic of	463	3.6	453	2.3
Japan	554	1.8	552	1.7
Korea, Republic of	546	2.0	558	1.6
(Latvia-LSS)[3]	476	3.3	513	2.9
(Lithuania)[4]	464	4.0	519	2.1
(Netherlands)[1]	541	6.0	536	3.1
New Zealand	511	4.9	520	5.0
Norway	514	2.4	494	2.2
(Romania)	471	5.1	470	4.9
Russian Federation	523	4.5	514	3.7
(Scotland)[1]	501	5.6	512	3.4
Singapore	580	5.5	578	4.3
Slovak Republic	532	3.3	517	3.2
(Slovenia)	514	2.7	520	1.8
Sweden	553	4.4	524	2.7
(United States)	513	5.6	527	3.1

[1]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]Designated LSS because only Latvian-speaking schools were included in 1995. For this analysis, only Latvian-speaking schools are included in the 2003 average.
[4]National desired population does not cover all of the international desired population in all years for Lithuania.
NOTE: Parentheses indicate countries that did not meet international sampling or other guidelines in 1995 or 2003. See appendix A for details regarding 2003 data. See Gonzales et al. (2000) for details regarding 1995 data. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details. s.e. means standard error.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2003.

Table C16. Percent correct of eighth-grade students in five science content areas, by country: 1999 and 2003

Country	All science trend items				Science content areas: Life science				Science content areas: Chemistry			
	1999		2003		1999		2003		1999		2003	
	Percent	s.e.	Percent	s.e.	Percent	s.e.	Percent	s.e.	Percent	s.e.	Percent	s.e.
Belgium-Flemish	60	0.5	56	0.5	64	0.5	61	0.6	51	1.0	49	0.5
Bulgaria	57	1.1	50	1.1	58	1.3	50	1.2	62	1.1	53	1.2
Chile	38	0.7	40	0.5	41	0.8	43	0.6	38	0.7	41	0.7
Chinese Taipei	67	0.6	66	0.7	64	0.6	62	0.6	72	0.8	71	0.9
Cyprus	46	0.3	42	0.4	49	0.6	41	0.5	47	0.7	42	0.5
Hong Kong SAR[1,2]	59	0.7	61	0.7	59	0.8	61	0.6	56	0.7	57	0.7
Hungary	63	0.7	62	0.5	61	0.8	61	0.7	67	0.8	66	0.7
Indonesia[3]	40	0.6	39	0.6	38	0.7	38	0.6	32	0.6	31	0.4
Iran, Islamic Republic of	44	0.7	44	0.5	40	0.7	39	0.6	48	0.7	46	0.6
(Israel)	49	0.8	53	0.6	50	0.9	56	0.7	51	0.9	56	0.8
Italy	53	0.7	53	0.6	54	0.8	55	0.8	53	1.0	52	0.8
Japan	63	0.4	61	0.5	63	0.5	61	0.5	61	0.6	59	0.6
Jordan	47	0.6	48	0.7	46	0.7	50	0.9	52	0.8	51	0.8
Korea, Republic of	64	0.4	63	0.4	62	0.5	64	0.5	61	0.5	54	0.5
Latvia-LSS[4]	53	0.6	54	0.7	50	0.8	53	0.8	53	0.8	54	1.0
Lithuania[3]	50	0.8	58	0.6	48	0.9	57	0.7	53	0.9	60	0.7
(Macedonia, Republic of)	46	0.7	45	0.7	47	0.8	45	0.8	52	1.1	52	0.9
Malaysia	52	0.8	53	0.8	51	1.0	49	1.0	49	0.7	52	0.9
Moldova, Republic of	47	0.8	48	0.7	48	0.9	46	1.0	46	1.0	50	0.8
Netherlands[1]	61	1.4	61	0.7	63	1.5	66	0.8	53	1.2	53	0.8
New Zealand	54	1.0	56	1.0	56	1.1	59	1.0	50	1.1	50	1.2
Philippines	33	0.9	35	0.8	34	1.0	38	1.0	34	0.8	31	0.7
Romania	48	0.9	48	1.0	48	1.1	50	1.1	52	1.2	49	1.1
Russian Federation	57	1.3	56	0.6	54	1.5	55	0.5	64	1.5	61	1.0
Singapore	67	1.4	67	0.9	66	1.5	65	0.9	65	1.6	70	1.1
Slovak Republic	58	0.7	56	0.7	59	0.8	57	0.8	61	0.8	57	0.9
South Africa	24	0.7	23	0.7	24	0.9	23	0.7	29	0.6	27	0.6
Tunisia	41	0.4	35	0.5	39	0.5	34	0.6	45	0.5	40	0.4
(United States)	57	0.7	58	0.6	61	0.9	63	0.7	55	0.9	55	0.7

See notes at end of table.

Table C16. Percent correct of eighth-grade students in five science content areas, by country: 1999 and 2003 —Continued

Country	Science content areas											
	Physics				Earth science				Environmental science			
	1999		2003		1999		2003		1999		2003	
	Percent	s.e.	Percent	s.e.	Percent	s.e.	Percent	s.e.	Percent	s.e.	Percent	s.e.
Belgium-Flemish	64	0.8	61	0.6	59	1.0	56	0.7	54	0.7	49	0.8
Bulgaria	52	1.4	48	1.1	63	1.2	57	1.3	50	1.3	43	1.3
Chile	37	0.7	40	0.5	38	0.7	41	0.6	37	0.8	33	0.6
Chinese Taipei	64	0.7	62	0.8	71	0.7	69	0.8	69	0.8	70	0.9
Cyprus	47	0.5	46	0.6	46	0.6	43	0.6	42	0.7	35	0.6
Hong Kong SAR[1,2]	62	0.8	61	0.7	65	0.9	64	0.8	55	1.0	62	1.0
Hungary	63	0.8	62	0.7	70	0.9	66	0.7	53	1.0	52	1.0
Indonesia[3]	43	0.7	42	0.7	45	0.9	43	0.8	46	0.9	40	0.8
Iran, Islamic Republic of	42	0.7	41	0.6	53	0.9	54	0.8	40	0.8	42	0.7
(Israel)	48	0.9	53	0.8	50	1.1	54	0.7	42	1.0	42	0.9
Italy	50	0.8	49	0.7	58	1.0	61	0.9	49	0.9	47	0.9
Japan	68	0.4	65	0.5	66	0.6	62	0.6	50	0.7	54	0.9
Jordan	42	0.6	42	0.8	52	0.7	53	0.8	44	0.8	44	1.0
Korea, Republic of	67	0.4	68	0.5	67	0.7	67	0.6	58	0.7	58	0.8
Latvia-LSS[4]	57	0.8	57	0.9	51	1.0	54	1.0	48	1.0	49	1.2
Lithuania[3]	55	0.9	61	0.6	49	1.0	59	0.8	38	1.0	46	0.8
(Macedonia, Republic of)	45	0.9	45	0.7	45	1.1	47	0.9	35	0.9	34	1.0
Malaysia	53	0.8	55	0.8	56	1.0	56	1.0	50	1.0	51	1.1
Moldova, Republic of	47	0.9	49	0.9	52	1.0	53	0.9	38	1.2	38	1.1
Netherlands[1]	64	1.5	65	0.8	61	1.5	62	0.9	59	2.0	58	1.3
New Zealand	57	1.0	60	1.0	53	1.0	53	1.1	54	1.1	52	1.4
Philippines	33	0.8	35	0.8	35	1.0	36	1.0	26	1.1	33	1.3
Romania	47	1.0	47	0.9	52	1.1	51	1.2	42	1.2	44	1.2
Russian Federation	58	1.1	56	0.7	60	1.4	61	0.7	46	1.5	45	1.0
Singapore	69	1.3	68	0.7	63	1.5	65	0.8	73	1.8	68	1.1
Slovak Republic	59	0.9	56	0.7	57	1.0	60	0.9	53	0.9	50	1.0
South Africa	24	0.7	23	0.8	23	0.6	24	0.7	20	0.9	19	1.0
Tunisia	39	0.5	33	0.6	44	0.7	38	0.7	38	0.5	30	0.7
(United States)	54	0.7	57	0.6	58	0.8	60	0.7	54	0.7	55	0.9

[1]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]National desired population does not cover all of the international desired population in all years for Lithuania, and in 2003 for Indonesia.
[4]Designated LSS because only Latvian-speaking schools were included in 1999. For this analysis, only Latvian-speaking schools are included in the 2003 average.
NOTE: Parentheses indicate countries that did not meet international sampling or other guidelines in 1999 or 2003. See appendix A for details regarding 2003 data. See Gonzales et al. (2000) for details regarding 1999 data. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details. s.e. means standard error.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1999 and 2003.

Table C17. Average science scale scores of fourth-grade students, by sex and country: 1995 and 2003

Country	Boys				Girls			
	1995		2003		1995		2003	
	Scale score	s.e.	Scale score	s.e.	Scale score	s.e.	Scale score	s.e.
(Australia)[1]	524	4.8	519	5.5	519	3.7	522	3.8
Cyprus	455	3.9	484	2.9	445	3.1	477	2.5
England[1]	530	4.0	538	4.6	525	3.5	542	3.3
Hong Kong SAR[1,2]	515	3.9	541	3.2	501	3.4	544	3.3
(Hungary)	515	3.9	533	3.2	501	3.8	527	3.7
Iran, Islamic Republic of	383	7.3	406	4.7	377	5.5	426	7.0
Japan	559	2.2	545	2.0	547	2.0	542	1.8
(Latvia–LSS)[3]	485	5.5	526	3.7	488	5.7	534	3.0
(Netherlands)[1]	544	3.9	529	2.2	518	3.3	521	2.2
New Zealand[4]	499	7.0	521	2.3	511	4.8	526	3.2
Norway	509	4.9	466	2.9	497	3.6	467	3.2
Scotland[1]	517	5.3	508	4.0	512	4.5	496	3.1
Singapore	526	5.3	565	6.4	521	5.8	565	5.4
(Slovenia)	470	4.1	490	3.2	458	3.3	491	3.0
United States[1]	548	3.3	538	2.8	536	3.6	533	2.5

[1]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]Designated LSS because only Latvian-speaking schools were included in 1995. For the purposes of this analysis, only Latvian-speaking schools were included in the 2003 average.
[4]In 1995, Maori-speaking students did not participate. Estimates in this table are computed for students taught in English only, which represents between 98-99 percent of the student population in both years.
NOTE: Parentheses indicate countries that did not meet international sampling or other guidelines in 1995 regarding 1995 data. All countries met international sampling and other guidelines in 2003, except as noted. See NCES (1997) for details regarding 1995 data. Countries were required to sample students in the upper of the two grades that contained the most number of 9-year-olds. In the United States and most countries, this corresponds to grade 4. See table A1 in appendix A for details. s.e. means standard error.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2003.

Table C18. Average science scale scores of U.S. fourth-grade students, by selected characteristics: 1995 and 2003

Selected characteristics	1995		2003	
	Scale score	s.e.	Scale score	s.e.
Race/ethnicity[1]				
White	572	3.0	565	2.1
Black	462	5.1	487	3.3
Hispanic	503	5.3	498	3.6
Asian	‡	‡	544	6.7
Poverty level in public schools (percentage of students eligible for free or reduced-price lunch)				
Less than 10 percent	—	—	579	4.9
10 to 24.9 percent	—	—	567	4.0
25 to 49.9 percent	—	—	551	4.0
50 to 74.9 percent	—	—	519	4.2
75 percent or more	—	—	480	4.3

—Not available.
‡Reporting standards not met.
[1]Other race/ethnicities are included in the U.S. totals shown throughout the report but not shown separately. Racial categories exclude Hispanic origin.
NOTE: The United States met international sampling guidelines for participation rates in 2003 only after replacement schools were included. s.e. means standard error.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2003.

Table C19. Average science scale scores of eighth-grade students, by sex and country: 1995, 1999, and 2003

Country	Boys					
	1995		1999		2003	
	Scale score	s.e.	Scale score	s.e.	Scale score	s.e.
(Australia)[1]	520	5.3	—	—	537	4.6
Belgium-Flemish	542	9.0	544	7.2	528	3.4
(Bulgaria)	543	5.7	525	6.5	487	5.2
Chile	—	—	432	5.1	427	3.6
Chinese Taipei	—	—	578	5.7	572	3.8
Cyprus	451	2.4	465	3.0	440	2.8
Hong Kong SAR[2,3]	525	6.3	537	5.1	561	3.8
Hungary	549	3.5	565	4.5	556	3.0
Indonesia[4]	—	—	444	4.8	426	4.6
Iran, Islamic Republic of	475	4.6	461	4.4	453	3.7
(Israel)[5]	—	—	476	5.5	498	4.1
Italy[5]	—	—	503	5.6	496	3.8
Japan	564	2.2	556	3.6	557	2.7
Jordan	—	—	442	5.9	462	5.6
Korea, Republic of	559	2.8	559	3.2	564	1.9
(Latvia-LSS)[6]	490	4.3	510	4.8	515	3.3
(Lithuania)[4]	477	4.5	499	5.0	522	2.4
(Macedonia, Republic of)	—	—	458	5.4	445	4.2
Malaysia	—	—	498	5.8	515	4.0
Moldova, Republic of	—	—	465	5.4	468	3.7
(Netherlands)[2]	554	7.4	554	7.3	543	3.8
New Zealand	524	6.1	513	7.0	525	6.7
Norway	523	3.5	—	—	498	3.0
Philippines	—	—	339	8.9	374	6.4
(Romania)	478	5.6	475	6.5	474	4.9
Russian Federation	530	5.1	540	6.2	519	4.2
Scotland[2]	515	6.7	—	—	517	3.5
Singapore	587	7.0	578	9.7	579	5.0
Slovak Republic	545	3.3	546	4.5	525	3.4
(Slovenia)[1]	524	3.4	—	—	524	2.3
South Africa[5]	—	—	253	7.7	244	7.7
Sweden	559	4.9	—	—	528	2.7
Tunisia	—	—	442	4.3	416	2.6
(United States)	520	6.1	524	5.5	536	3.4

See notes at end of table.

Table C19. Average science scale scores of eighth-grade students, by sex and country: 1995, 1999, and 2003—Continued

Country	Girls					
	1995		1999		2003	
	Scale score	s.e.	Scale score	s.e.	Scale score	s.e.
(Australia)[1]	508	3.9	—	—	517	4.6
Belgium-Flemish	524	8.7	526	4.6	505	3.0
(Bulgaria)	548	6.1	511	5.8	470	6.3
Chile	—	—	409	4.3	398	3.2
Chinese Taipei	—	—	561	3.9	571	3.8
Cyprus	454	2.9	455	3.1	443	2.3
Hong Kong SAR[2,3]	492	6.5	522	4.4	552	3.4
Hungary	525	3.7	540	4.0	530	3.4
Indonesia[4]	—	—	427	6.5	415	3.9
Iran, Islamic Republic of	448	5.7	430	5.7	454	3.9
(Israel)[5]	—	—	461	6.0	479	3.2
Italy[5]	—	—	484	4.1	486	2.7
Japan	544	1.9	543	2.8	548	3.0
Jordan	—	—	460	5.0	489	4.5
Korea, Republic of	530	2.5	538	4.0	552	2.1
(Latvia-LSS)[6]	464	3.8	495	5.6	511	3.2
(Lithuania)[4]	452	4.3	478	4.4	516	2.7
(Macedonia, Republic of)	—	—	458	6.0	454	3.7
Malaysia	—	—	488	5.5	505	4.3
Moldova, Republic of	—	—	454	4.4	477	3.5
(Netherlands)[2]	528	5.7	536	7.1	528	3.3
New Zealand	497	5.6	506	5.4	515	4.8
Norway	506	2.5	—	—	490	2.2
Philippines	—	—	351	8.2	380	5.9
(Romania)	464	5.4	468	6.4	465	5.5
Russian Federation	516	4.5	519	7.1	508	3.7
Scotland[2]	487	5.2	—	—	506	4.0
Singapore	574	6.7	557	7.9	576	4.0
Slovak Republic	520	4.1	525	3.4	508	3.8
(Slovenia)[1]	505	2.8	—	—	517	2.4
South Africa[5]	—	—	234	9.2	242	7.2
Sweden	546	4.8	—	—	521	3.2
Tunisia	—	—	417	3.3	392	2.3
(United States)	505	5.4	505	4.6	519	3.2

—Not available.
[1]Because of national-level changes in the starting age/date for school, 1999 data for Australia and Slovenia cannot be compared to 2003.
[2]Met international guidelines for participation rates in 2003 only after replacement schools were included.
[3]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[4]National desired population does not cover all of the international desired population in all years for Lithuania, and in 2003 for Indonesia.
[5]Because of changes in the population tested, 1995 data for Israel and Italy are not shown.
[6]Designated LSS because only Latvian-speaking schools were included in 1995 and 1999. For this analysis, only Latvian-speaking schools are included in the 2003 average.
[7]Because within classroom sampling was not accounted for, 1995 data are not shown for South Africa.
NOTE: Parentheses indicate countries that did not meet international sampling or other guidelines in 1995, 1999 or 2003. See appendix A for details regarding 2003 data. See Gonzales et al. (2000) for details regarding 1995 and 1999 data. Countries were required to sample students in the upper of the two grades that contained the most number of 13-year-olds. In the United States and most countries, this corresponds to grade 8. See table A1 in appendix A for details. s.e. means standard error.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995, 1999, and 2003.

Table C20. Average science scale scores of U.S. eighth-grade students, by selected characteristics: 1995, 1999, and 2003

Selected characteristics	1995		1999		2003	
	Scale score	s.e.	Scale score	s.e.	Scale score	s.e.
Race/ethnicity[1]						
White, not Hispanic or Latino	544	3.3	547	4.0	552	2.6
Black, not Hispanic or Latino	422	8.3	438	5.7	463	5.1
Hispanic or Latino	446	5.0	462	7.4	482	5.3
Poverty level in public schools (percentage of students eligible for free or reduced-price lunch)						
Less than 10 percent	—	—	579	12.0	571	6.6
10 to 24.9 percent	—	—	559	4.6	554	6.8
25 to 49.9 percent	—	—	513	8.8	529	5.1
50 to 74.9 percent	—	—	484	7.4	504	5.3
75 percent or more	—	—	439	10.0	461	10.2

—Not available.

[1]Other race/ethnicities are included in the U.S. totals shown throughout the report but not shown separately. Racial categories exclude Hispanic origin.

NOTE: The United States did not meet international sampling guidelines in 2003. See appendix A for more information. s.e. means standard error.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995, 1999, and 2003.

Table C21. Standard deviation of mathematics and science scores of U.S. fourth-grade and eighth-grade students, by selected characteristics: 2003

Selected characteristics	Fourth grade		Eighth grade	
	Mathematics	Science	Mathematics	Science
Sex				
Boys	75	80	80	80
Girls	72	75	75	76
Race/ethnicity[1]				
White	67	68	70	66
Black	63	67	69	69
Hispanic	69	73	73	75
Asian	71	71	81	79
Poverty level in public schools (percentage of students eligible for free or reduced-price lunch)				
Less than 10 percent	58	60	64	61
10 to 24.9 percent	66	66	76	72
25 to 49.9 percent	67	68	67	69
50 to 74.9 percent	66	72	66	69
75 percent or more	66	71	67	70

[1]Other races/ethnicities are included in the U.S. totals shown throughout the report but not shown separately. Racial categories exclude Hispanic origin.

NOTE: The United States met international sampling guidelines for participation rates in 2003 only after replacement schools were included at grade 4. The United States did not meet international sampling guidelines for participation rates in 2003 at grade 8. See appendix A for more information. s.e. means standard error.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Appendix D: Comparisons Between TIMSS, NAEP, and PISA

The analyses presented in this report examine the performance of U.S. fourth- and eighth-grade students in comparison to their counterparts in other countries. The TIMSS data are best understood in relation to data from other large assessments of similar subjects, such as the National Assessment of Educational Progress (NAEP) or the Program for International Student Assessment (PISA). Some of the TIMSS results for the United States mirror similar findings in the 2003 NAEP mathematics assessment (Braswell, Daane, and Grigg 2003). For example, as in TIMSS eighth grade, the national mathematics average of eighth-graders in NAEP increased from 1996, the most comparable dates between NAEP and TIMSS. However, some of the TIMSS results, particularly at fourth grade, do not mirror the findings in NAEP. Both TIMSS and NAEP are curriculum-based studies, while PISA, an international assessment of the reading, mathematics, and science literacy skills and abilities of 15-year-olds in the 30-member countries of the Organization for Economic Cooperation and Development, is less so. PISA 2003 results indicate that U.S. 15-year-olds performed relatively poorly in mathematical literacy in comparison to their peers in the other OECD-member nations (Lemke et al. 2004). In 2003, 15-year-olds in the United States scored below the international average in mathematical literacy and below their peers in 20 of the 28 other OECD-member countries.

Consistent with the principle of assessing curriculum- and school-based mathematics learning, NAEP and TIMSS focus on the performance of students in the same grade (fourth, eighth and, for NAEP, twelfth). This is important from the NAEP perspective because it allows the development of proficiency benchmarks—what students should know by the end of eighth grade—against which to compare what students actually know at the end of eighth grade. In the case of TIMSS, this allows comparisons of countries based on student populations with similar numbers of years of schooling. PISA, on the other hand, measures the demonstrated mathematics and science literacy of students of the same age—15-year-olds. This allows an internationally comparable measure of system yield: the knowledge output of the education system at a point when students are nearing the end of compulsory education.

The scores on NAEP, TIMSS, and PISA are not directly comparable, for both technical and practical reasons. Rather, the information on student achievement collected through these three studies can be understood through comparisons of their conceptual frameworks as well as the assessment items. NCES sponsored two comparative studies of TIMSS, PISA, and NAEP items. The first was a comparison of the conceptual frameworks and assessment items using NAEP as the centerpiece (Neidorf, Binkley, Gattis, and Nohara forthcoming; Neidorf, Binkley, and Stephens forthcoming). The second was a comparison of the conceptualization of and implementation of problem-solving assessments items in PISA and TIMSS (Dossey, O'Sullivan, and McCrone forthcoming).

Based on the NAEP conceptual framework, a panel of mathematics and science experts compared the mathematics assessment items from TIMSS, PISA, and NAEP on several dimensions: content, grade level, item type, and cognitive processes. The results of this study indicate that mathematics items from TIMSS 2003 and NAEP 2000 and 2003 appear more similar in content than do PISA 2003 and NAEP 2003 (Neidorf, Binkley, Gattis, and Nohara forthcoming; Neidorf, Binkley, and Stephens forthcoming). Examination of the mathematics frameworks and items showed that a major difference is that both TIMSS 2003 and NAEP 2003 mathematics have a relatively high percentage (33 and 26 percent, respectively) of items focused on the content area Number compared to PISA, which has the highest percentage of items (40 percent) focused on the content area Data, the content area of least focus in TIMSS and NAEP (Neidorf, Binkley, Gattis, and Nohara forthcoming). Grade-level analysis suggests that an eighth-grade TIMSS mathematics item or a PISA item designed for 15-year-olds could also be an eighth-grade NAEP item—in other words, that almost all the items seemed to fit within the age/grade descriptions for each assessment. Examination of the science frameworks and items showed that while NAEP 2000 and TIMSS 2003 are generally similar in terms of their broad content areas in science, there is some difference in relative emphasis (Neidorf, Binkley, and Stephens forthcoming). For example, NAEP currently has a greater emphasis than TIMSS on Earth Science at both the fourth and eighth grades than does TIMSS. TIMSS has a greater emphasis than NAEP on Life Science in the fourth grade and on Physical Science in the eighth grade. TIMSS also includes Environmental Science as an explicit part of its framework whereas NAEP does not. Over 80 percent of the science items from TIMSS and NAEP map to the other's framework at the corresponding grade level. The study also found that NAEP science items require more conceptual understanding than TIMSS science items, whereas TIMSS gives relatively more emphasis to items requiring factual knowledge than does NAEP. For more detailed information on the comparative item study, see Neidorf, Binkley, Gattis, and Nohara (forthcoming); and Neidorf, Binkley, and Stephens (forthcoming).

In a separate study (Dossey, O'Sullivan, and McCrone forthcoming), PISA and TIMSS mathematics and science items were examined for their connection to problem-solving skills and abilities. While PISA 2003 provided students with a separate assessment focused on problem-solving, TIMSS 2003 incorporated problem-solving and inquiry (PSI) tasks into the regular assessment booklets. In addition to items that were specifically designed to tap into problem-solving skills and abilities, the remaining items were also examined for the range of problem-solving skills embedded in them. A review of all the assessment items in PISA 2003 and TIMSS 2003 showed that 38 percent of eighth-grade TIMSS 2003 mathematics items and 48 percent of PISA 2003 mathematical literacy items measured some aspect of problem-solving; similarly, 26 percent of eighth-grade TIMSS 2003 science items and 49 percent of PISA science literacy items measured problem-solving skills (Dossey, O'Sullivan, and McCrone forthcoming). More items in PISA were found to require students to critically evaluate information than in TIMSS, both in mathematics and science. A similar percentage of problem-solving items in TIMSS science and PISA science measured scientific inquiry skills (33 percent). Eighty percent of TIMSS science items required students to know science information and knowledge compared to 35 percent of PISA science items. And, PISA items were more likely to involve a reading passage than TIMSS items. NAEP and TIMSS were similar in the predominance of multiple-choice items; PISA was more likely to employ extended-response items. For more detailed information on the comparative item study, see Dossey, O'Sullivan, and McCrone (forthcoming).

In sum, among the three studies, TIMSS and NAEP appear to have the most in common, with a focus on material that is more likely to be taught through the school curriculum than PISA, which is more situation- and phenomena-based. The content in TIMSS and NAEP mathematics and science overlap substantially. Nonetheless, NAEP was found to have a greater emphasis on Earth Science and TIMSS has a greater emphasis on Physical Science in the eighth grade. TIMSS also includes Environmental Science as an explicit part of its framework whereas NAEP does not. TIMSS and PISA appear to have less in common than TIMSS and NAEP. TIMSS and PISA differ in a number of respects, including a greater focus on factual knowledge in mathematics and science in TIMSS than in PISA, and a greater focus on problem solving and the critical evaluation of information in PISA than in TIMSS. Moreover, PISA has a greater focus on data analysis, statistics and probability in mathematics than either TIMSS or NAEP.

The detailed examinations of the conceptual underpinnings and assessment items in TIMSS, PISA, and NAEP described above offer, among other possibilities, at least one way to understand the most recent results in mathematics and science from these studies. Assuming that TIMSS and NAEP mathematics and science have more in common than do either TIMSS and PISA or NAEP and PISA, it seems reasonable to have expected recent improvements in the average performance of eighth-graders on NAEP mathematics to be found in the TIMSS data as well. However, the TIMSS results at fourth grade do not mirror the most recent NAEP results. Assuming that PISA places more emphasis on items that require a greater focus on problem solving, the critical evaluation of information, as well as a greater focus on data analysis, statistics and probability in mathematics than either TIMSS or NAEP, it also seems reasonable to have expected the PISA results in mathematics to differ from results in either TIMSS or NAEP.

A more thorough and detailed examination of the results from all three studies—TIMSS, PISA, and NAEP—may reveal other differences and similarities between them. Moreover, such analyses may provide insights into the actual reasons that U.S. students perform differently in seemingly similar subject areas on national and international assessments. Finally, the results from the comparisons among TIMSS, NAEP, and PISA frameworks and items, carried out in anticipation of the release of TIMSS and PISA 2003 data, will likely change in the future whenever any of the guiding frameworks for these three assessments are updated.

Appendix E: TIMSS Online Resources and Publications

Online Resources

The NCES website (http://nces.ed.gov/timss) provides background information on the TIMSS surveys, copies of NCES publications that relate to TIMSS, and information for educators about how to use TIMSS in the classroom.

NCES Publications

The following publications are intended to serve as examples of some of the numerous reports that have been produced in relation to the Trends in International Mathematics and Science Study (TIMSS) by NCES. All of the publications listed here are available at http://nces.ed.gov/timss.

TIMSS 1999 Summary and Achievement Reports

Gonzales, P., Calsyn, C., Jocelyn, L., Mak, K., Kastberg, D., Arafeh, S., Williams, T., and Tsen, W. (2000). *Pursuing Excellence: Comparisons of International Eighth-Grade Mathematics and Science Achievement From a U.S. Perspective, 1995 and 1999* (NCES 2001–028). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Gonzales, P., Calsyn, C., Jocelyn, L., Mak, D., Kastberg, D., Arafeh, S., Williams, T., and Tsen, W. (2000). *Highlights From TIMSS-R* (NCES 2001–027). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

TIMSS 1995 Summary and Achievement Reports

National Center for Education Statistics, U.S. Department of Education. (1997). *Pursuing Excellence: A Study of U.S. Fourth-Grade Mathematics and Science Achievement in International Context* (NCES 97–255). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Peak, L. (1996). *Pursuing Excellence: A Study of U.S. Eighth-Grade Mathematics and Science Teaching, Learning, Curriculum, and Achievement in International Context* (NCES 97–198). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Takahira, S., Gonzales, P., Frase, M., and Salganik, L.H. (1998). *Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context* (NCES 98–049). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

TIMSS Videotape Classroom Study Reports

Hiebert, J., Gallimore, R., Garnier, H., Givvin Bogard, K., Hollingsworth, H., Jacobs, J., Miu-Ying Chui, A., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., Gonzales, P., and Stigler, J. (2003). *Teaching Mathematics in Seven Countries: Results From the TIMSS 1999 Video Study* (NCES 2003–013 Revised). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

National Center for Education Statistics, U.S. Department of Education. (2000). *Highlights From the TIMSS Videotape Classroom Study* (NCES 2000–094). Washington, DC: U.S. Government Printing Office.

Roth, K.J., Druker, S.L., Garnier, H., Lemmens, M., Chen, C., Kawanaka, T., Rasmussen, D., Trubacova, S., Warvi, D., Gonzales, P., Stigler, J., and Gallimore, R. (forthcoming). *Teaching Science in Five Countries: Results From the TIMSS 1999 Video Study.* U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Stigler, J.W., Gonzales, P., Kawanaka, T., Knoll, S., and Serrano, A. (1999). *The TIMSS Videotape Classroom Study: Methods and Findings From an Exploratory Research Project on Eighth-Grade Mathematics Instruction in Germany, Japan, and the United States* (NCES 1999–074). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

TIMSS Data Products

National Center for Education Statistics, U.S. Department of Education (2003). *Third International Mathematics and Science Study (TIMSS) 1999 U.S. National Restricted-Use Data and User's* Guide (NCES 2003–075). Washington, DC: U.S. Government Printing Office.

IEA Publications

The following publications are intended to serve as examples of some of the numerous reports that have been produced in relation to TIMSS by the IEA. All of the publications listed here are available at http://timss.bc.edu.

TIMSS 2003 Summary and Achievement Reports

Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., and Chrostowski, S.J. (2004). *TIMSS 2003 International Science Report: Findings From IEA's Trends in International Mathematics and Science Study at the Eighth and Fourth Grades.* Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., and Chrostowski, S.J. (2004). *TIMSS 2003 International Mathematics Report: Findings From IEA's Trends in International Mathematics and Science Study at the Eighth and Fourth Grades.* Chestnut Hill, MA: Boston College.

TIMSS 1999 Summary and Achievement Reports

Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., Gregory, K.D., Smith, T.A., Chrostowski, S.J., Garden, R.A., and O'Connor, K.M. (2000). *TIMSS 1999 International Science Report: Findings From IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade.* Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Gregory, K.D., Garden, R.A., O'Connor, K.M., Chrostowski, S.J., and Smith, T.A. (2000). *TIMSS 1999 International Mathematics Report: Findings From IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade.* Chestnut Hill, MA: Boston College.

TIMSS 1995 Summary and Achievement Reports

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

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

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

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

Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., and Smith, T.A. (1998). *Mathematics and Science Achievement in the Final Year of Secondary School: IEA's Third International Mathematics and Science Study.* Chestnut Hill, MA: Boston College.

TIMSS Technical Reports and Frameworks

Martin, M.O., Gregory, K.D., and Stemler, S.E. (2000). *TIMSS 1999 Technical Report.* Chestnut Hill, MA: Boston College.

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

Martin, M.O., and Kelly, D.L. (Eds.). (1998). *Third International Mathematics and Science Study Technical Report, Volume II: Implementation and Analysis, Primary and Middle School Years.* Chestnut Hill, MA: Boston College.

Martin, M.O., Mullis, I.V.S. and Chrostowski, S.J. (2004). *TIMSS 2003 Technical Report: Findings From IEA's Trends in International Mathematics and Science Study at the Eighth and Fourth Grades.* Chestnut Hill, MA: Boston College.

Martin, M.O., and Kelly, D.L. (Eds.). (1999). *Third International Mathematics and Science Study Technical Report, Volume III: Implementation and Analysis, Final Year of Secondary School.* Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzalez, E.J., Chrostowski, S.J., and O'Connor, K.M. (2003). *TIMSS Assessment Frameworks and Specifications 2003: 2nd Edition.* Chestnut Hill, MA: Boston College.

TIMSS Data Products

Gonzalez, E.J., and Miles, J.A. (Eds.). (2001). *TIMSS 1999 User Guide for the International Database: IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade.* Chestnut Hill, MA: Boston College.