Highlights From TIMSS 2007:

Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context

December 2008

Patrick Gonzales
Project Officer
National Center for Education Statistics

Trevor Williams
Leslie Jocelyn
Stephen Roey
David Kastberg
Summer Brenwald
Westat

NCES 2009-001
U.S. DEPARTMENT OF EDUCATION

NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Sue Betka
Acting Director

National Center for Education Statistics
Stuart Kerachsky
Acting 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 countries. 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. Unless specifically noted, all information contained herein is in the public domain.

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 Special Education Research

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

December 2008
The NCES World Wide Web Home Page address is http://nces.ed.gov.
The NCES World Wide Web Electronic Catalog is http://nces.ed.gov/pubsearch.

Suggested Citation
Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., and Brenwald, S. (2008). *Highlights From TIMSS 2007: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context* (NCES 2009–001). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

For ordering information on this report, write to

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

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

Content Contact
Patrick Gonzales
(415) 920-9229
patrick.gonzales@ed.gov

Executive Summary

The 2007 Trends in International Mathematics and Science Study (TIMSS) is the fourth administration since 1995 of this international comparison. Developed and implemented at the international level by the International Association for the Evaluation of Educational Achievement (IEA)—an international organization of national research institutions and governmental research agencies—TIMSS is used to measure over time the mathematics and science knowledge and skills of fourth- and eighth-graders. TIMSS is designed to align broadly with mathematics and science curricula in the participating countries.

This report focuses on the performance of U.S. students relative to that of their peers in other countries in 2007, and on changes in mathematics and science achievement since 1995.[1] Thirty-six countries or educational jurisdictions participated at grade four in 2007, while 48 participated at grade eight.[2] This report also describes additional details about the achievement of U.S. student subpopulations. All differences described in this report are statistically significant at the .05 level. No statistical adjustments to account for multiple comparisons were used.

Key findings from the report include the following:

- In 2007, the average mathematics scores of both U.S. fourth-graders (529) and eighth-graders (508) were higher than the TIMSS scale average (500 at both grades).[3] The average U.S. fourth-grade mathematics score was higher than those of students in 23 of the 35 other countries, lower than those in 8 countries (all located in Asia or Europe), and not measurably different from those in the remaining 4 countries.[4] At eighth grade, the average U.S. mathematics score was higher than those of students in 37 of the 47 other countries, lower than those in 5 countries (all of them located in Asia), and not measurably different from those in the other 5 countries.
- Compared to 1995, the average mathematics scores for both U.S. fourth- and eighth-grade students were higher in 2007. At fourth grade, the U.S. average score in 2007 was 529, 11 points higher than the 1995 average of 518. At eighth grade, the U.S. average mathematics score in 2007 was 508, 16 points higher than the 1995 average of 492.
- In 2007, 10 percent of U.S. fourth-graders and 6 percent of U.S. eighth-graders scored at or above the advanced international benchmark in mathematics.[5] At grade four, seven countries had higher percentages of students performing at or above the advanced international mathematics benchmark than the United States: Singapore, Hong Kong SAR, Chinese Taipei, Japan, Kazakhstan, England, and the Russian Federation. Fourth-graders in these seven countries were also found to outperform U.S. fourth-graders, on average, on the overall mathematics scale. At grade eight, a slightly different set of seven countries had higher percentages of students performing at or above the advanced mathematics benchmark than the United States: Chinese Taipei, Korea, Singapore, Hong Kong SAR, Japan, Hungary, and the Russian Federation. These seven countries include the five countries that had higher average overall mathematics scores than the United States, as well as Hungary and the Russian Federation.
- In 2007, the average science scores of both U.S. fourth-graders (539) and eighth-graders (520) were higher than the TIMSS scale average (500 at both grades). The average U.S. fourth-grade science score was higher than those of students in 25 of the 35 other countries, lower than those in 4 countries (all of them in Asia), and not measurably different from those in the remaining 6 countries. At eighth grade, the average U.S. science score was higher than the average scores of students in 35 of the 47 other countries, lower than those in 9 countries (all located in Asia or Europe), and not measurably different from those in the other 3 countries.

[1]At grade four, a total of 257 schools and 10,350 students participated in the United States in 2007. At grade eight, 239 schools and 9,723 students participated. The overall weighted school response rate in the United States was 70 percent at grade four before the use of substitute schools. The final weighted student response rate at grade four was 95 percent. At grade eight, the overall weighted school response rate before the use of substitute schools was 68 percent. The final weighted student response rate at grade eight was 93 percent.
[2]The total number of countries reported here differs from the total number reported in the international TIMSS reports (Mullis et al. 2008; Martin et al. 2008). In addition to the 36 countries at grade four and 48 countries at grade eight, 8 other educational jurisdictions, or "benchmarking" entities, participated: the states of Massachusetts and Minnesota; the Canadian provinces of Alberta, British Columbia, Ontario, and Quebec; Dubai, United Arab Emirates; and the Basque region of Spain.
[3]TIMSS provides two overall scales—mathematics and science—as well as several content and cognitive domain subscales for each of the overall scales. The scores are reported on a scale from 0 to 1,000, with the TIMSS scale average set at 500 and standard deviation set at 100.
[4]TIMSS is open to countries and subnational entities, or educational jurisdictions, which are part of larger countries. For example, Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China. For convenience, this report uses the term "country" or "nation" to refer to all participating entities.
[5]TIMSS reports on four benchmarks to describe student performance in mathematics and science. Each benchmark is associated with a score on the achievement scale and a description of the knowledge and skills demonstrated by students at that level of achievement. The advanced international benchmark indicates that students scored 625 or higher. More information on the benchmarks can be found in the main body of the report and appendix A.

- The average science scores for both U.S. fourth- and eighth-grade students in 2007 were not measurably different from those in 1995. The U.S. fourth-grade average science score in 2007 was 539 and in 1995 was 542. The U.S. eighth-grade average science score in 2007 was 520 and in 1995 was 513.

- In 2007, 15 percent of U.S. fourth-graders and 10 percent of U.S. eighth-graders scored at or above the advanced international benchmark in science. At grade four, two countries had higher percentages of students performing at or above the advanced international science benchmark than the United States: Singapore and Chinese Taipei. Fourth-graders in these two countries were also found to outperform U.S. fourth-graders, on average, on the overall science scale. At grade eight, six countries had higher percentages of students performing at or above the advanced science benchmark than the United States: Singapore, Chinese Taipei, Japan, England, Korea, and Hungary. These six countries also had higher average overall eighth-grade science scores than the United States.

Acknowledgments

The authors wish to thank all those who assisted with TIMSS 2007, from its design to the reporting of findings. Most importantly, the authors wish to thank the many principals, teachers, and students who participated in the study.

Page intentionally left blank

Contents

List of Tables

List of Figures

LIST OF FIGURES

List of Exhibits

Introduction

TIMSS in brief

The Trends in International Mathematics and Science Study (TIMSS) 2007 is the fourth time since 1995 that this international comparison of student achievement has been conducted. Developed and implemented at the international level by the International Association for the Evaluation of Educational Achievement (IEA), an international organization of national research institutions and governmental research agencies, TIMSS is used to measure over time the mathematics and science knowledge and skills of fourth- and eighth-graders.

TIMSS is designed to align broadly with mathematics and science curricula in the participating countries. The results, therefore, suggest the degree to which students have learned mathematics and science concepts and skills likely to have been taught in school. TIMSS also collects background information on students, teachers, and schools to allow cross-national comparison of educational contexts that may be related to student achievement. In 2007, there were 58 countries and educational jurisdictions[1] that participated in TIMSS, at the fourth- or eighth-grade level, or both.[2]

This report presents the performance of U.S. students relative to their peers in other countries, and on changes in mathematics and science achievement since 1995. Most of the findings in the report are based on the results presented in two reports published by the IEA and available online at http://www.timss.org:

- *TIMSS 2007 International Mathematics Report: Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades* (Mullis et al. 2008); and
- *TIMSS 2007 International Science Report: Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades* (Martin et al. 2008).

For a number of participating countries, changes in achievement can be documented over the last 12 years, from 1995 to 2007. For other countries, changes can be documented over a shorter period of time. Table 1 and figure 1 show the countries that participated in TIMSS 2007 as well as their participation status in the earlier TIMSS data collections. The TIMSS fourth-grade assessment was implemented in 1995, 2003, and 2007, while the eighth-grade assessment was implemented in 1995, 1999, 2003, and 2007.

This report describes additional details about the achievement of U.S. students that are not available in the international reports, such as trends in the achievement of students of different racial and ethnic and socioeconomic backgrounds.

Design and administration of TIMSS

TIMSS 2007 is sponsored by the IEA and carried out under a contract with the TIMSS & PIRLS[3] International Study Center at Boston College. The National Center for Education Statistics (NCES), in the Institute of Education Sciences at the U.S. Department of Education, is responsible for the implementation of TIMSS in the United States. Data collection in the United States was carried out under contract to Windwalker Corporation and its subcontractors, Westat and Pearson Educational Measurement.

Participating countries administered TIMSS to two national probability samples of students and schools, based on a standardized definition. Countries were required to draw samples of students who were nearing the end of their fourth year or eighth year of formal schooling, beginning with the International Standard Classification of Education (ISCED) Level 1.[4] In most countries, including the United States, these students were in the fourth and eighth grades. Details on the grades assessed in each country are included in appendix A.

In the United States, TIMSS was administered between April and June 2007. The U.S. sample included both public and private schools, randomly selected and weighted to be representative of the nation.[5] In total, 257 schools and 10,350 students participated at grade four, and 239 schools and 9,723 students participated at grade eight. The overall weighted school response rate in the United States was 70

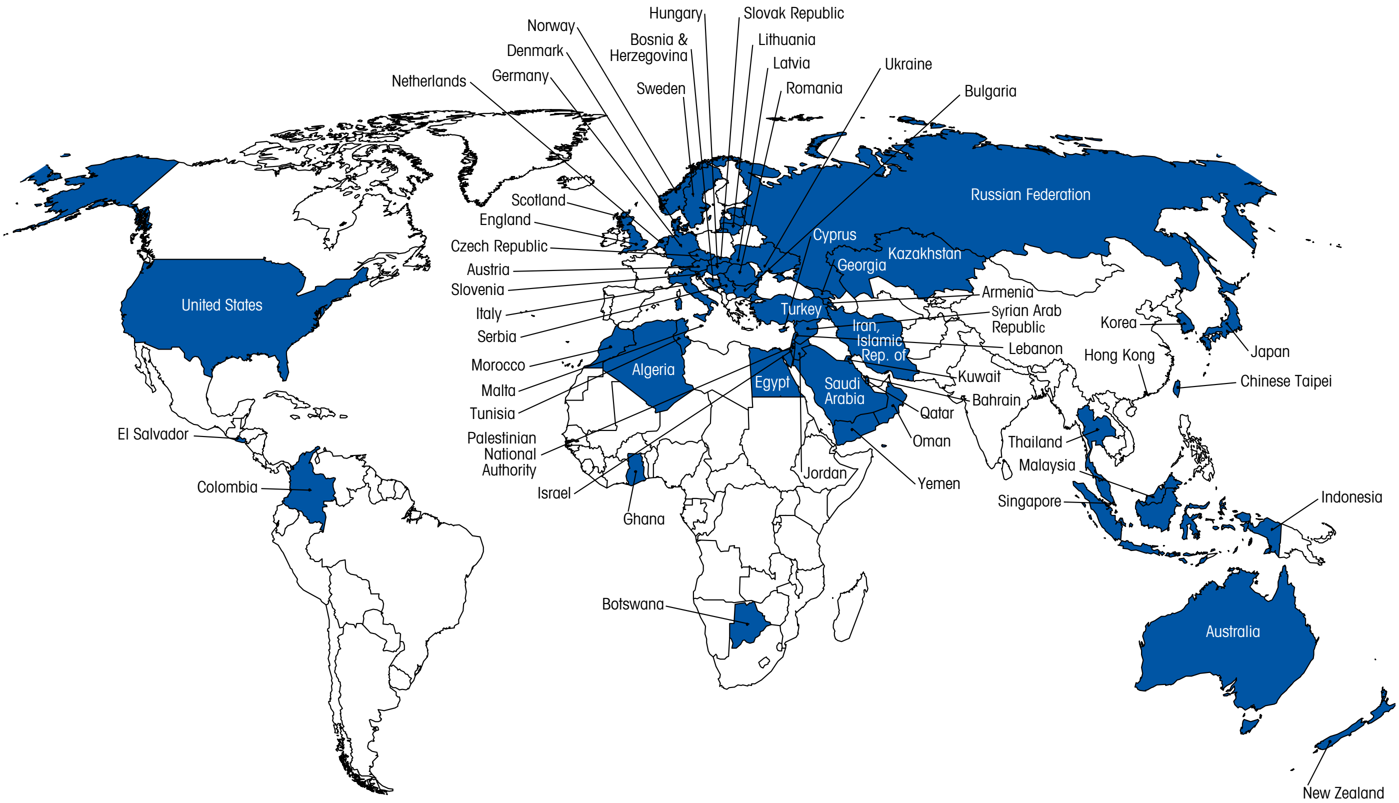
[1]TIMSS is open to countries and subnational entities, or educational jurisdictions, which are part of larger countries. For example, Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China. For convenience, this report uses the term "country" or "nation" to refer to all participating entities.
[2]Data from two nations were judged problematic by the IEA. Morocco failed to meet the required school participation rates in grade eight because of a procedural difficulty with some schools. Also, the quality of the data from Mongolia was not well documented at either grade level. In the international reports, Morocco is included in the fourth-grade tables but is shown "below the line" in the eighth-grade tables to indicate a problem in data quality. Data on Mongolia are reported in an appendix. For the purposes of the present report, statistics relating to Moroccan eighth-graders and to Mongolian students in both grades are not reported.
[3]The international study center takes its name from the two main IEA studies it coordinates; the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS).
[4]The ISCED was developed by the United Nations Educational, Scientific, and Cultural Organization (UNESCO) to assist countries in providing comparable, cross-national data. ISCED Level 1 is termed primary schooling, and in the United States is equivalent to the first through sixth grades (Matheson et al. 1996).
[5]The sample frame data for public schools in the United States was based on the 2006 National Assessment of Educational Progress (NAEP) sampling frame. This was done because recruitment of districts and schools began at the end of the 2005-06 school year to maximize response rates. The 2006 NAEP sampling frame was based on the 2003-04 Common Core of Data (CCD), and the data for private schools were from the 2003-04 Private School Universe Survey (PSS). Any school containing at least one grade four or one grade eight class was included in the school sampling frame.

Table 1. Participation in the TIMSS fourth- and eighth-grade assessments, by grade and country: 1995, 1999, 2003, and 2007

	Grade four			Grade eight			
Country	1995	2003	2007	1995	1999	2003	2007
Total	26	25	36	41	38	46	48
Algeria			✓				✓
Armenia		✓	✓			✓	✓
Australia[1]	✓	✓	✓	✓	✓	✓	✓
Austria	✓		✓	✓			
Bahrain						✓	✓
Belgium (Flemish)		✓		✓	✓	✓	
Belgium (French)				✓			
Bosnia and Herzegovina							✓
Botswana						✓	✓
Bulgaria				✓	✓	✓	✓
Canada	✓			✓	✓		
Chile					✓	✓	
Chinese Taipei		✓	✓		✓	✓	✓
Colombia			✓	✓			✓
Cyprus	✓	✓		✓	✓	✓	✓
Czech Republic	✓		✓	✓	✓		✓
Denmark			✓	✓			
Egypt						✓	✓
El Salvador			✓				✓
England[2]	✓	✓	✓	✓	✓	✓	✓
Estonia						✓	
Finland					✓		
France				✓			
Georgia			✓				✓
Germany			✓	✓			
Ghana						✓	✓
Greece	✓			✓			
Hong Kong SAR[3]	✓	✓	✓	✓	✓	✓	✓
Hungary	✓	✓	✓	✓	✓	✓	✓
Iceland	✓			✓			
Indonesia					✓	✓	✓
Iran, Islamic Rep. of	✓	✓	✓	✓	✓	✓	✓
Ireland	✓			✓			
Israel[4]	✓			✓	✓	✓	✓
Italy[4]		✓	✓		✓	✓	✓
Japan	✓	✓	✓	✓	✓	✓	✓
Jordan					✓	✓	✓
Kazakhstan			✓				
Korea, Rep. of	✓			✓	✓	✓	✓
Kuwait	✓		✓	✓			✓
Latvia[5]	✓	✓	✓	✓	✓	✓	
Lebanon						✓	✓
Lithuania		✓	✓	✓	✓	✓	✓
Macedonia, Rep. of					✓	✓	
Malaysia					✓	✓	✓
Malta							✓
Moldova, Rep. of		✓			✓	✓	
Morocco[4]		✓	✓		✓	✓	
Netherlands	✓	✓	✓	✓	✓	✓	
New Zealand	✓	✓	✓	✓	✓	✓	
Norway	✓	✓	✓	✓		✓	✓
Oman							✓
Palestinian Nat'l Auth.						✓	✓
Philippines		✓			✓	✓	
Portugal	✓			✓			
Qatar			✓				✓
Romania				✓	✓	✓	✓
Russian Federation		✓	✓	✓	✓	✓	✓
Saudi Arabia						✓	✓
Scotland	✓	✓	✓	✓		✓	✓
Serbia						✓	✓
Singapore	✓	✓	✓	✓	✓	✓	✓
Slovak Republic			✓	✓	✓	✓	
Slovenia[1]	✓	✓	✓	✓	✓	✓	✓
South Africa[6]				✓	✓	✓	
Spain				✓			
Sweden			✓	✓		✓	✓
Switzerland				✓			
Syrian Arab Republic							✓
Thailand	✓			✓	✓		✓
Tunisia		✓	✓		✓	✓	✓
Turkey					✓		✓
Ukraine			✓				✓
United States	✓	✓	✓	✓	✓	✓	✓
Yemen			✓				

[1]Because of national-level changes in the starting age/date for school, 1999 data for Australia and Slovenia cannot be compared to 2003 data.
[2]England collected data at grade eight 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.
[3]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[4]Because of changes in the population tested, 1995 data for Israel and Italy, and 1999 data for Morocco are not shown.
[5]Only Latvian-speaking schools were included in 1995 and 1999. For trend analyses, only Latvian-speaking schools are included in the estimates.
[6]Because within-classroom sampling was not accounted for, 1995 data are not shown for South Africa.
NOTE: No fourth-grade assessment was conducted in 1999. 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, eight separate jurisdictions participated in the Trends in International Mathematics and Science Study (TIMSS) 2007: the provinces of Alberta, British Columbia, Ontario, and Quebec in Canada; the Basque region of Spain; Dubai, UAE, and the states of Massachusetts and Minnesota. Information on these eight jurisdictions can be found in the international TIMSS 2007 reports. Morocco participated in TIMSS 2007 at both the fourth and eighth grades, but due to sampling difficulties, its grade eight data are not shown in this report. Mongolia also participated in TIMSS 2007 but could not complete the steps necessary to have its data included in the report. Countries could participate at either grade level. Countries were required to sample students enrolled in the grade corresponding to the fourth and eighth year of schooling, beginning with International Standard Classification of Education (ISCED) level 1, providing that the mean age at the time of testing was at least 9.5 years and 13.5 years, respectively. In the United States and most countries, this corresponds to grade four and grade eight. 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, 2003 and 2007.

Figure 1. Countries that participated in TIMSS 2007



SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

percent at grade four before the use of substitute schools and 89 percent with the inclusion of substitute schools.[6] At grade eight, the overall weighted school response rate before the use of substitute schools was 68 percent and 83 percent with the inclusion of substitute schools. The final weighted student response rate at grade four was 95 percent and at grade eight was 93 percent. Student response rates are based on a combined total of students from both sampled and substitute schools. Detailed information on sampling, administration, response rates, and other technical issues are included in appendix A.

Reporting TIMSS results

Achievement results from TIMSS are reported on a scale from 0 to 1,000, with a TIMSS scale average of 500 and standard deviation of 100. Even though the countries participating in TIMSS have changed across the four assessments between 1995 and 2007, comparisons between the 2007 results and prior results are still possible because the achievement scores in each of the TIMSS assessments are placed on a scale which is not dependent on the list of participating countries in any particular year. A brief description of the assessment equating and scaling is presented in appendix A to this volume. A more detailed presentation can be found in the *TIMSS 2007 Technical Report* (Olson, Martin, and Mullis 2008).

In addition to numerical scale results, TIMSS also includes international benchmarks. The TIMSS international benchmarks provide a way to interpret the scale scores and to understand how students' proficiency in mathematics and science varies along the TIMSS scale. The TIMSS benchmarks describe four levels of student achievement in each subject, based on the kinds of skills and knowledge students at each score cutpoint would need to successfully answer the mathematics and science items. In general, the score cutpoints for the TIMSS benchmarks were set based on the distribution of students

[6]NCES standards advise that substitute schools should not be included in the calculation of response rates (standard 1-3-8; National Center for Education Statistics 2002). Response rates calculated "before replacement" are consistent with this standard. Response rates calculated "after replacement" include substitute schools and hence are not consistent with NCES standards. Both kinds of response rates are reported here in the interests of comparability with the TIMSS international reports which report response rates before and after replacement.

along the TIMSS scale. More information on the development of the benchmarks and the procedures used to set the score cutpoints can be found in the *TIMSS 2007 Technical Report* (Olson, Martin, and Mullis 2008).

All differences described in this report are statistically significant at the .05 level. No statistical adjustments to account for multiple comparisons were used. Differences 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 "not measurably different" or "not statistically significant." In this latter case, failure to find a difference as statistically significant does not necessarily mean that there was no difference. It simply means that, given the precision of the estimates, there is a larger than five percent chance that the difference was zero. In addition, because the results of tests of statistical significance are, in part, influenced by sample sizes, statistically significant results may not identify those findings that have policy or practical importance. For this reason, this report includes effect sizes to provide the reader with a sense of the magnitude of statistically significant differences. Further information about effect sizes and about the tests conducted to determine statistical significance can be found in appendix A. Supplemental tables providing all estimates and standard errors discussed in this report are available online at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001.

All data presented in this report are used to describe relationships between variables. These data are not intended, nor can they be used, to imply causality. Student performance can be affected by a complex mix of educational and other factors that are not examined here.

Nonresponse bias in the U.S. TIMSS samples

NCES standards require a nonresponse bias analysis if school-level response rates fall below 85 percent, as they did for both the fourth- and eighth-grade school samples in TIMSS 2007.[7] As a consequence, a nonresponse bias analysis was undertaken, similar to that used for TIMSS 2003 (Ferraro and Van De Kerckhove 2006).

These analyses examined whether the participation status of schools (participant/non-participant) was related to seven school characteristics: the region of the country in which the school was located (Northeast, Southeast, Central, West); the type of community served by the school (central city, urban fringe/large town, rural/small town); whether the school was public or private; percentage of students eligible for free or reduced-price lunch; number of students enrolled in fourth or eighth grade; total number of students; and percentage of students from minority backgrounds. Details are provided in appendix A.[8]

The findings indicate some potential for bias in the data arising from regional and community-type differences in participation, along with the fact that schools with higher percentages of minority students were less likely to participate. Specifically, grade 4 schools in the central region were more likely to participate than schools in the other regions, and schools in rural/small towns were more likely to participate than schools in central cities. However with the inclusion of substitute schools there were no measurable differences by region and differences by community type were substantially reduced. At grade 8, after substitution, the results of the analyses indicated that schools in central cities were still more likely to participate than schools in urban/fringe/large towns. At both grades, schools with higher percentages of minority students were less likely to participate, but the measurable differences were small after substitution especially at grade 8. Since TIMSS is conducted under a set of standard rules designed to facilitate international comparisons, the U.S. nonresponse bias analysis results were not used to adjust the U.S. data for this source of bias. While this may be possible at some later date, at present the variables identified above remain as potential sources of bias in the published estimates.

Further information

To assist the reader in understanding how TIMSS relates to the National Assessment of Educational Progress (NAEP), the primary source of national- and state-level data on U.S. students' mathematics and science achievement, NCES compared the form and content of the TIMSS and NAEP mathematics and science assessments. A summary of the results of this comparison is included in appendix C. Appendix D includes a list of TIMSS publications and resources published by NCES and the IEA. Standard errors for the estimates discussed in the report are available online at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001. Detailed information on TIMSS can also be found on the NCES website (http://nces.ed.gov/timss) and the international TIMSS website (http://www.timss.org).

[7]Standard 2-2-2 found in National Center for Education Statistics 2002.
[8]The full text of the nonresponse bias analysis conducted for TIMSS 2007 will be included in a technical report released with the U.S. national dataset. See appendix A for a description of the analyses undertaken and additional details on the findings.

Mathematics Performance in the United States and Internationally

The TIMSS mathematics assessment

The TIMSS mathematics assessment is designed along two dimensions: the mathematical topics or content that students are expected to learn and the cognitive skills students are expected to have developed. The topical or content domains (as they are called in TIMSS) covered at grade four are *number, geometric shapes and measures*, and *data display* (table 2). At grade eight, the content domains are *number*, *algebra*, *geometry*, and *data and chance*. The cognitive domains in each grade are *knowing*, *applying*, and *reasoning*. Example items from the TIMSS mathematics assessment are included in appendix B (see items B1 through B7).

The proportion of items devoted to a domain, and, therefore, the contribution of the domain to the overall mathematics scale score differs somewhat across grades. For example, in 2007 at grade four, 52 percent of the TIMSS mathematics assessment focused on the *number* domain, while the analogous percentage at grade eight was 29 percent. The proportion of items devoted to each cognitive domain was similar across grades.

Also, within a content or cognitive domain, the makeup of items, in terms of difficulty and form of knowledge and skills addressed, differs across grade levels to reflect the nature, difficulty, and emphasis of the subject matter encountered in school at each grade. *TIMSS 2007 Assessment Frameworks* (Mullis et al. 2005) provides a more detailed description of the content and cognitive domains assessed in TIMSS. The development and validation of the cognitive domains is detailed in *IEA's TIMSS 2003 International Report on Achievement in the Mathematics Cognitive Domains: Findings From a Developmental* Project (Mullis, Martin, and Foy 2005).

TIMSS provides an overall mathematics scale score as well as content and cognitive domain scores at each grade level. The TIMSS mathematics scale is from 0 to 1,000 and the international mean score is set at 500, with a standard deviation of 100. The scaling of data is conducted separately for each grade and each content domain. Thus, a score of 500 on the grade four scale is not equivalent to a score of 500 on the grade eight scale The scaling of data is conducted separately for each grade and each content domain. While the scales were created to each have a mean of 500 and a standard deviation of 100, the subject matter and the level of difficulty of items necessarily differ between the assessments at both grades. Therefore, direct comparisons between scores across grades should not be made. See appendix A for more details.

Table 2. Percentage of fourth- and eighth-grade TIMSS mathematics assessment devoted to content and cognitive domains: 2007

Grade four		Grade eight	
Content domains	**Percent of assessment**	**Content domains**	**Percent of assessment**
Number	52	Number	29
Geometric shapes and measures	34	Algebra	30
Data display	15	Geometry	22
		Data and chance	19
Cognitive domains	**Percent of assessment**	**Cognitive domains**	**Percent of assessment**
Knowing	39	Knowing	38
Applying	39	Applying	41
Reasoning	22	Reasoning	21

NOTE: The content and cognitive domains are the foundation of the Trends in International Mathematics and Science Study (TIMSS) assessment. The content domains define the specific mathematics subject matter covered by the assessment, and the cognitive domains define the sets of behaviors expected of students as they engage with the mathematics content. Each mathematics content domain has several topic areas. Each topic area is presented as a list of objectives covered in a majority of participating countries, at either grade four or grade eight. However, the cognitive domains of mathematics are defined by the same three sets of expected behaviors—knowing, applying, and reasoning. 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), 2007.

Scores within a subject and grade are comparable over time. The TIMSS scale was established originally to have a mean of 500 based on the average of all of the countries that participated in TIMSS 1995 at the fourth and eighth grades. Successive TIMSS assessments since then (TIMSS 1999, 2003, and 2007) have scaled the achievement data so that scores are equivalent from assessment to assessment. That is, a score of 500 in eighth-grade mathematics in 2007 is equivalent to a score of 500 in eighth-grade mathematics in 2003, in 1999, and in 1995. The same is true for the fourth-grade scale: a score of 500 in fourth-grade mathematics in 2007 is equivalent to a score of 500 in fourth-grade mathematics in 2003 and 1995. More information on how the TIMSS scale was created can be found in appendix A.

Average scores in 2007

The average mathematics scores for both U.S. fourth- and eighth-graders were higher than the TIMSS scale average (table 3). In 2007, the average score of U.S. fourth-graders was 529 and the average score of U.S. eighth-graders was 508, compared with the TIMSS scale average of 500 at each grade level.

At grade four, the average U.S. mathematics score was higher than those in 23 of the 35 other countries, lower than those in 8 countries (all 8 were in Asia or Europe), and not measurably different from the average scores in the remaining 4 countries.

At grade eight, the average U.S. mathematics score was higher than those in 37 of the 47 other countries, lower than those in 5 countries (all of them located in Asia), and not measurably different from the average scores in the other 5 countries.

Table 3. Average mathematics scores of fourth- and eighth-grade students, by country: 2007

Grade four: Country	Average score
TIMSS scale average	500
Hong Kong SAR[1]	607
Singapore	599
Chinese Taipei	576
Japan	568
Kazakhstan[2]	549
Russian Federation	544
England	541
Latvia[2]	537
Netherlands[3]	535
Lithuania[2]	530
United States[4,5]	**529**
Germany	525
Denmark[4]	523
Australia	516
Hungary	510
Italy	507
Austria	505
Sweden	503
Slovenia	502
Armenia	500
Slovak Republic	496
Scotland[4]	494
New Zealand	492
Czech Republic	486
Norway	473
Ukraine	469
Georgia[2]	438
Iran, Islamic Rep. of	402
Algeria	378
Colombia	355
Morocco	341
El Salvador	330
Tunisia	327
Kuwait[6]	316
Qatar	296
Yemen	224

Grade eight: Country	Average score
TIMSS scale average	500
Chinese Taipei	598
Korea, Rep. of	597
Singapore	593
Hong Kong SAR[1,4]	572
Japan	570
Hungary	517
England[4]	513
Russian Federation	512
United States[4,5]	**508**
Lithuania[2]	506
Czech Republic	504
Slovenia	501
Armenia	499
Australia	496
Sweden	491
Malta	488
Scotland[4]	487
Serbia[2,5]	486
Italy	480
Malaysia	474
Norway	469
Cyprus	465
Bulgaria	464
Israel[7]	463
Ukraine	462
Romania	461
Bosnia and Herzegovina	456
Lebanon	449
Thailand	441
Turkey	432
Jordan	427
Tunisia	420
Georgia[2]	410
Iran, Islamic Rep. of	403
Bahrain	398
Indonesia	397
Syrian Arab Republic	395
Egypt	391
Algeria	387
Colombia	380
Oman	372
Palestinian Nat'l Auth.	367
Botswana	364
Kuwait[6]	354
El Salvador	340
Saudi Arabia	329
Ghana	309
Qatar	307

■ Average score is higher than U.S. average score ($p < .05$)
□ Average score is not measurably different from the U.S. average score ($p < .05$)
□ Average score is lower than the U.S. average score ($p < .05$)

[1]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[2]National Target Population does not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS) (see appendix A).
[3]Nearly satisfied guidelines for sample participation rates only after substitute schools were included (see appendix A).
[4]Met guidelines for sample participation rates only after substitute schools were included (see appendix A).
[5]National Defined Population covers 90 percent to 95 percent of National Target Population (see appendix A).
[6]Kuwait tested the same cohort of students as other countries, but later in 2007, at the beginning of the next school year.
[7]National Defined Population covers less than 90 percent of National Target Population (but at least 77 percent, see appendix A).

NOTE: Countries are ordered by 2007 average score. 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. The standard errors of the estimates are shown in tables E-1 and E-2 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Trends in scores since 1995

Several countries participated in both the first TIMSS in 1995 and the most recent TIMSS in 2007 and therefore the average scores can be compared over a 12-year period. At grade four, 16 countries, including the United States, participated in both the first and most recent TIMSS administrations. Comparing 2007 mathematics scores with those from 1995, one-half of the countries (8 of 16), including the United States, showed improvement in average scores and one-quarter of the countries (4 of 16) showed declines (table 4). In 2007, the U.S. fourth-grade average mathematics score of 529 was 11 scale score points higher than the 1995 average of 518.

The gain in the U.S. fourth-grade average mathematics score (11 scale score points) was greater than the difference in six countries (the four countries with declines in average scores, as well as two other countries) and less than the gain of four countries (England, Hong Kong SAR, Slovenia, and Latvia). There was no measurable difference between the 11 score point gain in the United States and the gains or declines in score points experienced in the other countries.

At grade eight, 20 countries, including the United States, participated in TIMSS in both 1995 and 2007. About one-quarter of the countries (6 of 20), including the United States, had higher average mathematics scores in 2007 than in 1995 and students in one-half of the countries (10 of 20) showed declines in their average scores. The U.S. eighth-grade average mathematics score of 508 was 16 scale score points higher than the 1995 average of 492.

The gain in the U.S. eighth-grade mathematics score (16 scale score points) was greater than the difference

Table 4. Trends in average mathematics scores of fourth- and eighth-grade students, by country: 1995 to 2007

Grade four

Country	Average score 1995	Average score 2007	Difference[1] 2007–1995
England	484	541	57*
Hong Kong SAR[2]	557	607	50*
Slovenia	462	502	40*
Latvia[3]	499	537	38*
New Zealand	469	492	23*
Australia	495	516	22*
Iran, Islamic Rep. of	387	402	15*
United States[4,5]	**518**	**529**	**11***
Singapore	590	599	9
Scotland[4]	493	494	1
Japan	567	568	1
Norway	476	473	-3
Hungary	521	510	-12*
Netherlands[6]	549	535	-14*
Austria	531	505	-25*
Czech Republic	541	486	-54*

Grade eight

Country	Average score 1995	Average score 2007	Difference[1] 2007–1995
Colombia	332	380	47*
Lithuania[3]	472	506	34*
Korea, Rep. of	581	597	17*
United States[4,5]	**492**	**508**	**16***
England[4]	498	513	16*
Slovenia	494	501	7*
Hong Kong SAR[2,4]	569	572	4
Cyprus	468	465	-2
Scotland[4]	493	487	-6
Hungary	527	517	-10*
Japan	581	570	-11*
Russian Federation	524	512	-12
Romania	474	461	-12*
Australia	509	496	-13*
Iran, Islamic Rep. of	418	403	-15*
Singapore	609	593	-16*
Norway	498	469	-29*
Czech Republic	546	504	-42*
Sweden	540	491	-48*
Bulgaria	527	464	-63*

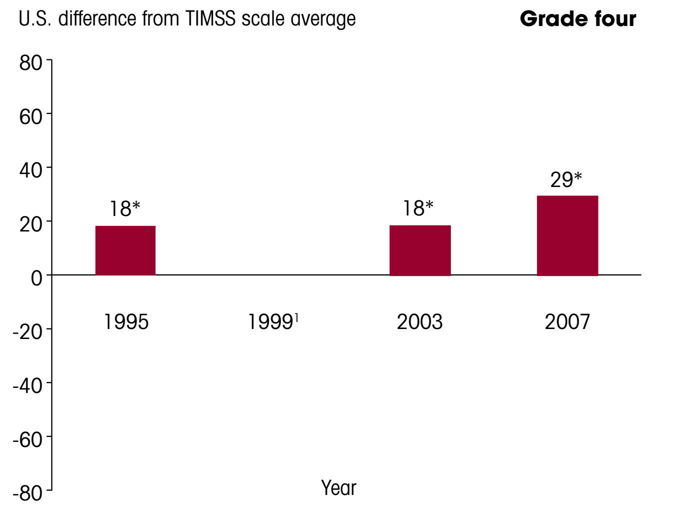
■ Country difference in average scores between 1995 and 2007 is greater than analogous U.S. difference ($p < .05$)
□ Country difference in average scores between 1995 and 2007 is not measurably different from analogous U.S. difference ($p < .05$)
■ Country difference in average scores between 1995 and 2007 is less than analogous U.S. difference ($p < .05$)
*$p < .05$. Within-country difference between 1995 and 2007 average scores is significant.
[1]Difference calculated by subtracting 1995 from 2007 estimate using unrounded numbers.
[2]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[3]In 2007, National Target Population did not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS) (see appendix A).
[4]In 2007, met guidelines for sample participation rates only after substitute schools were included (see appendix A).
[5]In 2007, National Defined Population covered 90 percent to 95 percent of National Target Population (see appendix A).
[6]In 2007, nearly satisfied guidelines for sample participation rates only after substitute schools were included (see appendix A).
NOTE: Countries are ordered based on the difference in 1995 and 2007 average scores. All countries met international sampling and other guidelines in 2007, except as noted. Data are not shown for some countries, because comparable data from previous cycles are not available. 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. Detail may not sum to totals because of rounding. The standard errors of the estimates are shown in tables E-1 and E-2 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995 and 2007.

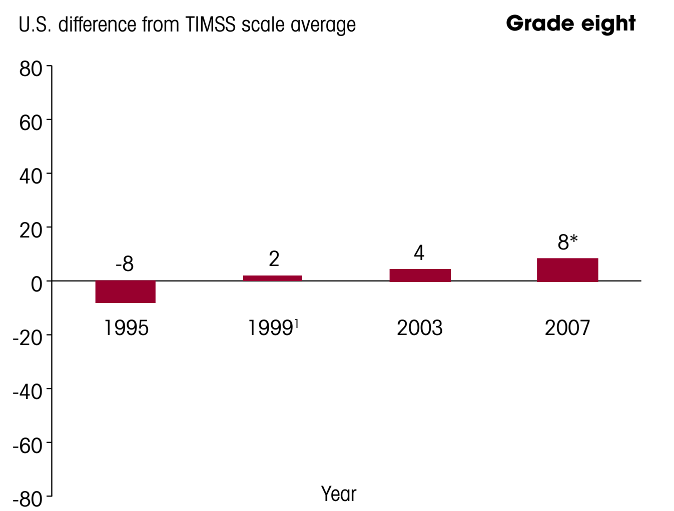
in 13 countries (including the 10 countries with declining scores and 3 others) and less than the gain of 2 countries (Colombia and Lithuania). There was no measurable difference between the 16 score point gain in the United States and the gains or declines in score points experienced in the other countries.

The size of the difference in scores between the U.S. fourth-graders' and TIMSS scale averages was larger in 2007 at 29 scale score points than it was in 1995 at 18 scale score points (figure 2). U.S. fourth-graders' average mathematics scores were higher than the TIMSS scale average in each of the 3 data collection years: 1995, 2003, and 2007.

U.S. eighth-graders' average mathematics scores showed no measurable difference from the TIMSS scale average in 3 of the 4 data collection years between 1995 and 2007. However, the 2007 U.S. score was higher than the U.S. score in 1995, with the U.S. score in 1995 some 8 points below the TIMSS scale average, but 8 points above the average in 2007.

Figure 2. Difference between average mathematics scores of U.S. fourth- and eighth-grade students and the TIMSS scale average: 1995, 1999, 2003, and 2007





**p* < .05. Difference between U.S. average and Trends in International Mathematics and Science Study (TIMSS) scale average is statistically significant.
[1]No fourth-grade assessment was conducted in 1999.
NOTE: In 2007, the United States met guidelines for sample participation rates only after substitute schools were included. The National Defined Population covered 90 percent to 95 percent of National Target Population (see appendix A). Difference calculated by subtracting the TIMSS scale average (500) from the U.S. average mathematics score. The standard errors of the estimates are shown in table E-39 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995, 1999, 2003, and 2007.

Content and cognitive domain scores in 2007

In addition to an overall mathematics score, TIMSS provides scores for content domains and cognitive domains (see table 5 for a description of the cognitive domains). U.S. fourth-graders scored higher than the TIMSS scale average across the mathematics content domains in 2007 (table 6). U.S. fourth-graders' average scores in *number, geometric shapes and measures*, and *data display* were between 22 and 43 scale score points above the TIMSS scale average of 500 in each content domain.

U.S. fourth-graders performed better on average in the *data display* domain than in the *number* and *geometric shapes and measures* domains, at least in terms of comparisons with other countries. That is, there were fewer countries that outperformed the United States in *data display* than in the other two domains. U.S. fourth-graders outperformed their peers in 22 countries in the *number* domain, 20 countries in the *geometric shapes and measures* domain, and 28 countries in the *data display* domain. They were outperformed by their peers in 9 countries in the *number* domain, 10 countries in the *geometric shapes and measures* domain, and 4 countries in the *data display* domain.

In the three cognitive domains, U.S. fourth-graders scored higher than the TIMSS scale average in 2007. U.S. fourth-graders' average scores in the *knowing*, *applying*, and *reasoning* domains were between 23 and 41 scale score points higher than the TIMSS scale average of 500.

In terms of comparisons with other countries, U.S. fourth-graders performed relatively better on average in the *applying* domain than the *knowing* and *reasoning* domains. U.S. fourth-graders outperformed students in 16 to 27 countries across the three cognitive domains and were outperformed by their peers in 5 to 11 countries across the three cognitive domains.

At the eighth-grade level, U.S. students scored higher, on average, than the TIMSS scale average in two of the four mathematics content domains in 2007 (table 7). U.S. eighth-graders' average scores in *number* and *data and chance* were 10 and 31 scale score points above the TIMSS scale score average of 500, respectively. On the other hand, U.S. eighth-graders' average score in the *geometry* domain was lower than the TIMSS scale score average by 20 scale score points. There was no measurable difference between U.S. eighth-graders' average score in *algebra* and the TIMSS scale score average.

U.S. eighth-graders performed relatively better, on average, in the *data and chance* domain than in the *number*, *algebra*,

Table 5. Description of TIMSS mathematics cognitive domains: 2007

Cognitive domain	Description
Knowing	*Knowing* addresses the facts, procedures, and concepts that students need to know to function mathematically. The key skills of this cognitive domain include recalling definitions, terminology, number properties, geometric properties, and notation; recognizing mathematical objects, shapes, numbers, and expressions; recognizing mathematical entities that are mathematically equivalent; computing algorithmic procedures for basic functions with whole numbers, fractions, decimals, and integers; approximating numbers to estimate computations; carrying out routine algebraic procedures; retrieving information from graphs, tables, and charts; reading simple scales; using appropriate units of measure and measuring instruments; estimating measures; classifying or grouping objects, shapes, numbers, and expressions according to common properties; making correct decisions about class membership; and ordering numbers and objects by attributes.
Applying	*Applying* focuses on students' abilities to apply knowledge and conceptual understanding to solve problems or answer questions. The key skills of this cognitive domain include selecting appropriate operations, methods, or strategies for solving problems where there is a known algorithm or method of solution; representing mathematics information and data in diagrams, tables, graphs, and charts; generating equivalent representations for a given mathematical entity or relationship; generating an appropriate mathematical model, such as an equation or diagram for solving a routine problem; following and executing a set of mathematical instructions; drawing figures and shapes given specifications; solving routine problems (i.e., problems similar to those students are likely to have encountered in class); comparing and matching different representations of data (grade eight) and using data from charts, tables, graphs, and maps to solve routine problems.
Reasoning	*Reasoning* goes beyond the cognitive processes involved in solving routine problems to include unfamiliar situations, complex contexts, and multistep problems. The key skills of this cognitive domain include determining and describing relationships between variables or objects in mathematical situations; using proportional reasoning (grade four); decomposing geometric figures to simplify solving a problem; drawing the net of a given unfamiliar solid; visualizing transformations of three-dimensional figures; comparing and matching different representations of the same data (grade four); making valid inferences from given information; generalizing mathematical results to wider applications; combining mathematical procedures to establish results and combining results to produce a further result; making connections between different elements of knowledge and related representations; making linkages between different elements of knowledge and related representations; making linkages between related mathematical ideas; providing a justification for the truth or falsity of a statement by reference to mathematical results or properties; solving problems set in mathematical or real life contexts that students are unlikely to have encountered before; applying mathematical procedures in unfamiliar or complex contexts; and using geometric properties to solve non-routine problems.

NOTE: The descriptions of the cognitive domains are the same for grades four and eight, except where noted.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 6. Average mathematics content and cognitive domain scores of fourth-grade students, by country: 2007

Country	Content domain: Number	Content domain: Geometric shapes and measures	Content domain: Data display	Cognitive domain: Knowing	Cognitive domain: Applying	Cognitive domain: Reasoning
TIMSS scale average	500	500	500	500	500	500
Hong Kong SAR[1]	606	599	585	599	617	589
Singapore	611	570	583	590	620	578
Chinese Taipei	581	556	567	569	584	566
Japan	561	566	578	566	565	563
Kazakhstan[2]	556	542	522	547	559	539
Russian Federation	546	538	530	547	538	540
England	531	548	547	540	544	537
Latvia[2]	536	532	536	540	530	537
Netherlands[3]	535	522	543	540	525	534
Lithuania[2]	533	518	530	539	520	526
United States[4,5]	**524**	**522**	**543**	**524**	**541**	**523**
Germany	521	528	534	531	514	528
Denmark[4]	509	544	529	528	513	524
Australia	496	536	534	523	509	516
Hungary	510	510	504	507	511	509
Italy	505	509	506	501	514	509
Austria	502	509	508	507	505	506
Sweden	490	508	529	508	482	519
Slovenia	485	522	518	504	497	505
Armenia	522	483	458	493	518	489
Slovak Republic	495	499	492	498	492	499
Scotland[4]	481	503	516	500	489	497
New Zealand	478	502	513	495	482	503
Czech Republic	482	494	493	496	473	493
Norway	461	490	487	479	461	489
Ukraine	480	457	462	466	472	474
Georgia[2]	464	415	414	433	450	437
Iran, Islamic Rep. of	398	429	400	405	410	410
Algeria	391	383	361	376	384	387
Colombia	360	361	363	357	360	372
Morocco	353	365	316	346	354	—
El Salvador	317	333	367	339	312	356
Tunisia	352	334	307	329	343	—
Kuwait[6]	321	316	318	305	326	—
Qatar	292	296	326	296	293	—
Yemen	—	—	—	—	—	—

■ Average score is higher than the U.S. average score ($p < .05$)
□ Average score is not measurably different from the U.S. average score ($p < .05$)
▪ Average score is lower than the U.S. average score ($p < .05$)
— Not available. Average achievement could not be accurately estimated.
[1]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[2]National Target Population does not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS) (see appendix A).
[3]Nearly satisfied guidelines for sample participation rates only after substitute schools were included (see appendix A).
[4]Met guidelines for sample participation rates only after substitute schools were included (see appendix A).
[5]National Defined Population covers 90 percent to 95 percent of National Target Population (see appendix A).
[6]Kuwait tested the same cohort of students as other countries, but later in 2007, at the beginning of the next school year.
NOTE: Countries are ordered by 2007 overall mathematics average scale score. 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. The standard errors of the estimates are shown in table E-3 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 7. Average mathematics content and cognitive domain scores of eighth-grade students, by country: 2007

Country	Content domain: Number	Content domain: Algebra	Content domain: Geometry	Content domain: Data and chance	Cognitive domain: Knowing	Cognitive domain: Applying	Cognitive domain: Reasoning
TIMSS scale average	500	500	500	500	500	500	500
Chinese Taipei	577	617	592	566	592	594	591
Korea, Rep. of	583	596	587	580	595	596	579
Singapore	597	579	578	574	593	581	579
Hong Kong SAR[1,2]	567	565	570	549	569	574	557
Japan	551	559	573	573	565	560	568
Hungary	517	503	508	524	513	518	513
England[2]	510	492	510	547	514	503	518
Russian Federation	507	518	510	487	510	521	497
United States[2,3]	**510**	**501**	**480**	**531**	**503**	**514**	**505**
Lithuania[4]	506	483	507	523	511	508	486
Czech Republic	511	484	498	512	504	502	500
Slovenia	502	488	499	511	503	500	496
Armenia	492	532	493	427	493	507	489
Australia	503	471	487	525	500	487	502
Sweden	507	456	472	526	497	478	490
Malta	496	473	495	487	492	490	475
Scotland[2]	489	467	485	517	489	481	495
Serbia[3,4]	478	500	486	458	478	500	474
Italy	478	460	490	491	483	476	483
Malaysia	491	454	477	469	478	477	468
Norway	488	425	459	505	477	458	475
Cyprus	464	468	458	464	465	468	461
Bulgaria	458	476	468	440	458	477	455
Ukraine	460	464	467	458	464	471	445
Romania	457	478	466	429	462	470	449
Israel[5]	469	470	436	465	456	473	462
Bosnia and Herzegovina	451	475	451	437	440	478	452
Lebanon	454	465	462	407	448	464	429
Thailand	444	433	442	453	446	436	456
Turkey	429	440	411	445	425	439	441
Jordan	416	448	436	425	422	432	440
Tunisia	425	423	437	411	423	421	425
Georgia[4]	421	421	409	373	401	427	389
Iran, Islamic Rep. of	395	408	423	415	402	403	427
Bahrain	388	403	412	418	403	395	413
Indonesia	399	405	395	402	398	397	405
Syrian Arab Republic	393	406	417	387	401	393	396
Egypt	393	409	406	384	393	392	396
Algeria	403	349	432	371	412	371	—
Colombia	369	390	371	405	384	364	416
Oman	363	391	387	389	368	372	397
Palestinian Nat'l Auth.	366	382	388	371	371	365	381
Botswana	366	394	325	384	351	376	—
Kuwait[6]	347	354	385	366	361	347	—
El Salvador	355	331	318	362	347	336	—
Saudi Arabia	309	344	359	348	335	308	—
Ghana	310	358	275	321	297	313	—
Qatar	334	312	301	305	305	307	—

■ Average score is higher than the U.S. average score ($p < .05$)
□ Average score is not measurably different from the U.S. average score ($p < .05$)
▪ Average score is lower than the U.S. average score ($p < .05$)
— Not available. Average achievement could not be accurately estimated.
[1]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[2]Met guidelines for sample participation rates only after substitute schools were included (see appendix A).
[3]National Defined Population covers 90 percent to 95 percent of National Target Population (see appendix A).
[4]National Target Population does not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS) (see appendix A).
[5]National Defined Population covers less than 90 percent of National Target Population (but at least 77 percent, see appendix A).
[6]Kuwait tested the same cohort of students as other countries, but later in 2007, at the beginning of the next school year.
NOTE: Countries are ordered by 2007 overall mathematics average scale score. 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. The standard errors of the estimates are shown in table E-4 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

and *geometry* domains and relatively worse, on average, in *geometry* than the other three content domains, at least in terms of comparisons with other countries. U.S. eighth-graders outperformed students in 38 countries in the *data and chance* domain, 35 countries in the *number* domain, 37 countries in the *algebra* domain, and 29 countries in the *geometry* domain. They were outperformed by their peers in 6 countries in the *data and chance* domain, 5 countries in the *number* domain, 7 countries in the *algebra* domain, and 14 countries in the *geometry* domain.

In two of the three cognitive domains, the U.S. eighth-grade average score was higher than the TIMSS scale average in 2007. U.S. eighth-graders' scores in the *applying* and *reasoning* domains were 14 and 5 scale score points above the TIMSS scale score average of 500, respectively. On the other hand, U.S. eighth-graders' average score in the *knowing* domain was not measurably different from the TIMSS scale score average.

Like their fourth-grade counterparts, U.S. eighth-graders performed relatively better in the *applying* domain than in the *knowing* and *reasoning* domains in terms of comparisons with other countries. U.S. eighth-graders outperformed students in 30 to 38 countries across the three cognitive domains. They were outperformed by their peers in 5 to 8 countries across the three cognitive domains.

Performance on the TIMSS international benchmarks

The TIMSS international benchmarks provide a way to understand how students' proficiency in mathematics varies along the TIMSS scale (table 8). TIMSS defines four levels of student achievement: advanced, high, intermediate, and low. The benchmarks can then be used to describe the kinds of skills and knowledge students at each score cutpoint would need to successfully answer the mathematics items included in the assessment. The descriptions of the benchmarks differ between the two grade levels, as the mathematical skills and knowledge needed to respond to the assessment items reflect the nature, difficulty, and emphasis at each grade.

Table 8. Description of TIMSS international mathematics benchmarks, by grade: 2007

Benchmark (score cutpoint)	Grade four
Advanced (625)	*Students can apply their understanding and knowledge in a variety of relatively complex situations and explain their reasoning.* They can apply proportional reasoning in a variety of contexts. They demonstrate a developing understanding of fractions and decimals. They can select appropriate information to solve multistep word problems. They can formulate or select a rule for a relationship. Students can apply geometric knowledge of a range of two- and three-dimensional shapes in a variety of situations. They can organize, interpret, and represent data to solve problems.
High (550)	*Students can apply their knowledge and understanding to solve problems.* Students can solve multistep word problems involving operations with whole numbers. They can use division in a variety of problem situations. They demonstrate understanding of place value and simple fractions. Students can extend patterns to find a later specified term and identify the relationship between ordered pairs. Students show some basic geometric knowledge. They can interpret and use data in tables and graphs to solve problems.
Intermediate (475)	*Students can apply basic mathematical knowledge in straightforward situations.* Students at this level demonstrate an understanding of whole numbers. They can extend simple numeric and geometric patterns. They are familiar with a range of two-dimensional shapes. They can read and interpret different representations of the same data.
Low (400)	*Students have some basic mathematical knowledge.* Students can demonstrate an understanding of adding and subtracting with whole numbers. They demonstrate familiarity with triangles and informal coordinate systems. They can read information from simple bar graphs and tables.
	Grade eight
Advanced (625)	*Students can organize and draw conclusions from information, make generalizations, and solve nonroutine problems.* They can solve a variety of ratio, proportion, and percent problems. They can apply their knowledge of numeric and algebraic concepts and relationships. Students can express generalizations algebraically and model situations. They can apply their knowledge of geometry in complex problem situations. Students can derive and use data from several sources to solve multistep problems.
High (550)	*Students can apply their understanding and knowledge in a variety of relatively complex situations.* They can relate and compute with fractions, decimals, and percents, operate with negative integers, and solve word problems involving proportions. Students can work with algebraic expressions and linear equations. Students use knowledge of geometric properties to solve problems, including area, volume, and angles. They can interpret data in a variety of graphs and table and solve simple problems involving probability.
Intermediate (475)	*Students can apply basic mathematical knowledge in straightforward situations.* They can add and multiply to solve one-step word problems involving whole numbers and decimals. They can work with familiar fractions. They understand simple algebraic relationships. They demonstrate understanding of properties of triangles and basic geometric concepts. They can read and interpret graphs and tables. They recognize basic notions of likelihood.
Low (400)	*Students have some knowledge of whole numbers and decimals, operations, and basic graphs.*

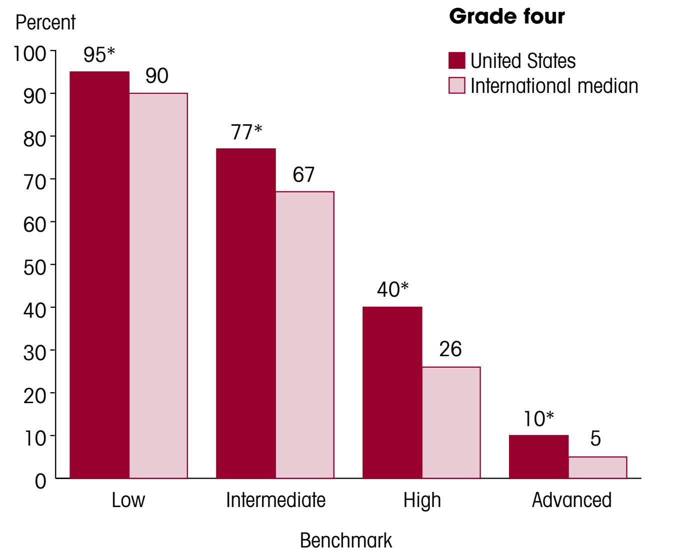
NOTE: Score cutpoints for the international benchmarks are determined through scale anchoring. Scale anchoring involves selecting benchmarks (scale points) on the achievement scales to be described in terms of student performance, and then identifying items that students scoring at the anchor points can answer correctly. The score cutpoints are set at equal intervals along the achievement scales. The score cutpoints were selected to be as close as possible to the standard percentile cutpoints (i.e., 90th, 75th, 50th, and 25th percentiles). More information on the setting of the score cutpoints can be found in appendix A and Martin et al. (2008).

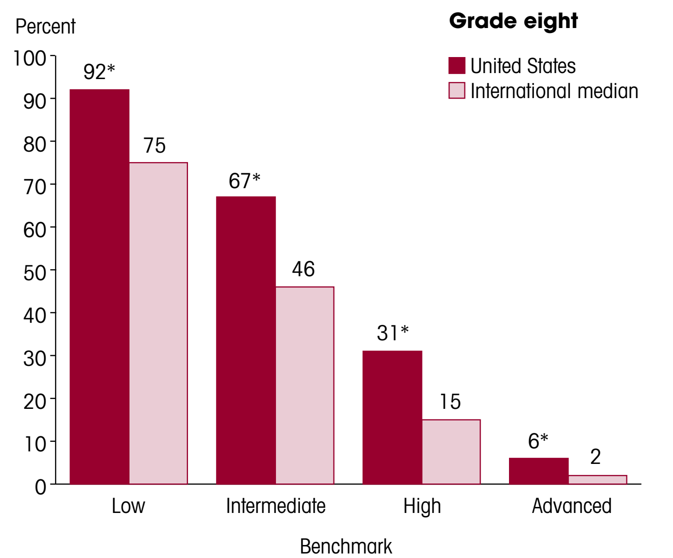
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

In 2007, there were higher percentages of U.S. fourth-graders performing at or above each of the four TIMSS international benchmarks than the international medians[9] of the percentages performing at each level (figure 3). For example, 10 percent of U.S. fourth-graders performed at or above the advanced benchmark (625) compared to the international median of 5 percent. These students demonstrated an ability to apply their understanding and knowledge to a variety of relatively complex mathematical situations (see description in table 8). At the other end of the scale, 95 percent of U.S. fourth-graders performed at or above the low benchmark (400) compared with the international median of 90 percent. These students showed at least some basic mathematical skills by demonstrating an understanding of adding and subtracting with whole numbers, showing familiarity with triangles and informal coordinate systems, and reading information from simple bar graphs and tables.

Similar to their fourth-grade counterparts, there were higher percentages of U.S. eighth-graders performing at or above each of the four TIMSS international benchmarks than the international medians of the percentage performing at each level (figure 3). For example, 6 percent of U.S. eighth-graders performed at or above the advanced benchmark (625) compared to the international median of 2 percent. These students demonstrated an ability to organize information, make generalizations, solve nonroutine problems, and draw and justify conclusions from data (see description in table 8). At the other end of the scale, 92 percent of U.S. eighth-graders performed at or above the low benchmark (400) compared with the international median of 75 percent. These students showed at least a basic mathematical understanding of whole numbers and decimals, could perform simple computations, and complete a basic graph.

Figure 3. Percentage of U.S. fourth- and eighth-grade students who reached each TIMSS international mathematics benchmark compared with the international median percentage: 2007





*$p < .05$. U.S. percentage is significantly different from the Trends in International Mathematics and Science (TIMSS) international median percentage.
NOTE: The United States met guidelines for sample participation rates only after substitute schools were included and the National Defined Population covers 90 percent to 95 percent of National Target Population (see appendix A). The TIMSS international median represents all participating TIMSS jurisdictions, including the United States. The international median represents the percentage at which half of the participating countries have that percentage of students at or above the median and half have that percentage of students below the median. The standard errors for the estimates are shown in table E-5 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

[9]The international median at each benchmark represents the percentage at which half of the participating countries have that percentage of students at or above the median and half have that percentage of students below the median. For example, the low international benchmark median of 90 percent at grade four indicates that half of the countries have 90 percent or more of their students who met the low benchmark, and half have less than 90 percent of their students who met the low benchmark.

At grade four, seven countries had higher percentages of students performing at or above the advanced international mathematics benchmark than the United States (figure 4). Fourth-graders in these seven countries were also found to outperform U.S. fourth-graders, on average, on the overall mathematics scale (see table 3). At grade eight, a slightly different set of seven countries had higher percentages of students performing at or above the advanced mathematics benchmark than the United States (figure 4). These seven countries include the five countries that had higher average overall mathematics scores than the United States (see table 3), as well as Hungary and the Russian Federation.

At grade four in 2007, higher percentages of U.S. students performed at or above the intermediate and low international benchmarks than in 1995 (intermediate: 77 v. 71 percent; low: 95 v. 92 percent; data not shown). There were no measurable differences in the percentage of U.S. fourth-graders performing at or above either the high or advanced international benchmarks between 1995 and 2007 (high: 37 v. 40 percent; advanced: 9 v. 10 percent). At grade eight, higher percentages of U.S. students performed at or above the high, intermediate, and low international benchmarks in 2007 than in 1995 (high: 31 v. 26 percent; intermediate: 67 v. 61 percent; low: 92 v. 86 percent; data not shown). There was no measurable difference in the percentage of U.S. eighth-graders performing at or above the advanced international benchmark in 2007 than in 1995 (6 v. 4 percent).

Performance within the United States

TIMSS not only provides a measure of mathematics performance of the nation as a whole, but also of the performance of student subpopulations. For this report, TIMSS data were analyzed to investigate the performance of students grouped in four ways: higher and lower performing students; males and females; racial and ethnic groups; and public schools serving students with different low-income concentrations.

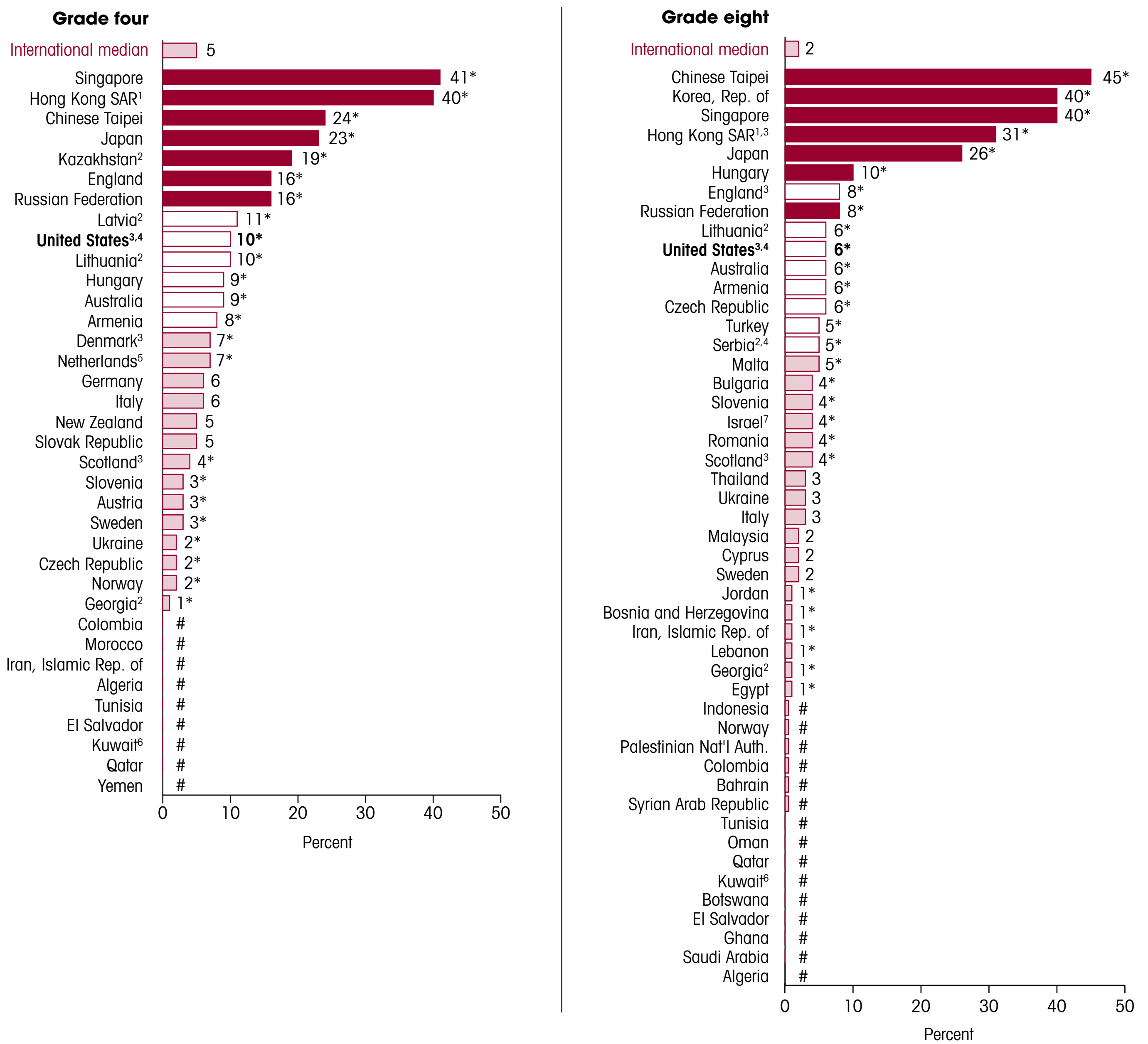
Scores of lower and higher performing students

To examine the mathematics performance of each participating country's higher and lower performing students, cutpoint scores were calculated for students performing at or above the 90th percentile (that is, the top 10 percent of students) and those performing at or below the 10th percentile (the bottom 10 percent of students). The cutpoint scores were calculated for each country, rather than across all countries combined.

In 2007, the highest-performing U.S. fourth-graders (those performing at or above the 90th percentile) scored 625 or higher (table 9). This was higher than the 90th percentile scores for fourth-graders in 23 countries and lower than the 90th percentile score for students in 7 countries. The countries in which the 90th percentile cutpoint score was higher than the cutpoint score for U.S. are the same as those that outperformed the United States as a whole (table 3), with the exception of Latvia where the 90th percentile score of 628 is not significantly different from 625 in the United States. The 90th percentile scores ranged between 371 (Yemen) and 702 (Singapore). The difference in the 90th percentile score between Singapore, the highest performing country, and the United States was 77 score points.

The lowest-performing U.S. fourth-graders (those performing at or below the 10th percentile) scored 430 or lower in 2007 (table 9). This was higher than the 10th percentile score in 23 countries and lower than the 10th percentile score in 6 countries: Singapore, Hong Kong SAR, Japan, Chinese Taipei, Latvia, and the Netherlands. The score at the 10th percentile ranged between 81 (Yemen) and 520 (Hong Kong SAR). The difference in the cutpoint scores between the lowest-performing students in Hong Kong SAR and the United States was 90 score points.

Figure 4. Percentage of fourth- and eighth-grade students who reached the TIMSS advanced international benchmark in mathematics, by country: 2007



■ Percentage is higher than U.S. percentage (*p* < .05)
□ Percentage is not measurably different from U.S. percentage (*p* < .05)
■ Percentage is lower than U.S. percentage (*p* < .05)
**p* < .05. Percentage is significantly different from the international median percentage.
Rounds to zero.
[1]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[2]National Target Population does not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS) (see appendix A).
[3]Met guidelines for sample participation rates only after substitute schools were included (see appendix A).
[4]National Defined Population covers 90 percent to 95 percent of National Target Population (see appendix A).
[5]Nearly satisfied guidelines for sample participation rates only after substitute schools were included (see appendix A).
[6]Kuwait tested the same cohort of students as other countries, but later in 2007, at the beginning of the next school year.
[7]National Defined Population covers less than 90 percent of National Target Population (but at least 77 percent, see appendix A).
NOTE: The TIMSS international median represents all participating TIMSS jurisdictions, including the United States. The international median represents the percentage at which half of the participating countries have that percentage of students at or above the median and half have that percentage of students below the median. 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. The standard errors for the estimates are shown in table E-41 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 9. Mathematics scores of fourth- and eighth-grade students defining 10th and 90th percentiles, by country: 2007

Grade four

Country	90th percentile	10th percentile
International average	576	366
Singapore	702	487
Hong Kong SAR[1]	691	520
Japan	663	471
Chinese Taipei	663	488
Kazakhstan[2]	653	435
England	647	429
Russian Federation	647	436
Latvia[2]	628	444
United States[3,4]	**625**	**430**
Lithuania[2]	624	430
Hungary	620	389
Australia	620	408
Armenia	617	385
Netherlands[5]	612	454
Denmark[3]	611	431
Germany	607	440
Italy	601	406
New Zealand	598	377
Slovak Republic	597	389
Scotland[3]	592	389
Austria	590	416
Slovenia	589	408
Sweden	586	417
Czech Republic	576	392
Ukraine	573	356
Norway	566	372
Georgia[2]	549	322
Iran, Islamic Rep. of	508	290
Algeria	493	261
Colombia	470	238
Tunisia	469	178
Morocco	466	223
El Salvador	448	212
Kuwait[6]	443	184
Qatar	413	179
Yemen	371	81

Grade eight

Country	90th percentile	10th percentile
International average	559	339
Chinese Taipei	721	448
Korea, Rep. of	711	475
Singapore	706	463
Hong Kong SAR[1,3]	681	438
Japan	677	460
Hungary	624	405
England[3]	618	400
Russian Federation	617	402
Lithuania[2]	609	402
United States[3,4]	**607**	**408**
Armenia	601	390
Australia	600	394
Czech Republic	599	408
Malta	597	359
Serbia[2,4]	597	368
Slovenia	594	409
Scotland[3]	590	381
Romania	587	328
Bulgaria	586	324
Israel[7]	584	328
Sweden	582	399
Turkey	581	297
Malaysia	578	372
Cyprus	575	347
Italy	574	381
Ukraine	572	346
Thailand	562	327
Jordan	556	290
Norway	552	382
Bosnia and Herzegovina	552	352
Lebanon	549	354
Georgia[2]	532	280
Egypt	521	258
Iran, Islamic Rep. of	516	295
Indonesia	509	286
Tunisia	508	336
Bahrain	505	289
Syrian Arab Republic	502	290
Palestinian Nat'l Auth.	498	233
Oman	492	245
Colombia	477	281
Algeria	465	311
Botswana	460	264
Kuwait[6]	455	252
El Salvador	433	248
Saudi Arabia	429	231
Ghana	428	192
Qatar	427	186

■ Percentile cutpoint score is higher than U.S. cutpoint score ($p < .05$)
□ Percentile cutpoint score is not measurably different from U.S. cutpoint score ($p < .05$)
▫ Percentile cutpoint score is lower than U.S. cutpoint score ($p < .05$)

[1]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[2]National Target Population does not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS, see appendix A).
[3]Met guidelines for sample participation rates only after substitute schools were included (see appendix A).
[4]National Defined Population covers 90 percent to 95 percent of National Target Population (see appendix A).
[5]Nearly satisfied guidelines for sample participation rates only after substitute schools were included (see appendix A).
[6]Kuwait tested the same cohort of students as other countries, but later in 2007, at the beginning of the next school year.
[7]National Defined Population covers less than 90 percent of National Target Population (but at least 77 percent, see appendix A).
NOTE: Countries are ordered based on the 90th percentile cutpoint for mathematics scores. Cutpoints are calculated based on distribution of student scores within each country. The international average is the average of the cutpoint scores for all reported countries. The standard errors of the estimates are shown in tables E-6 and E-7 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

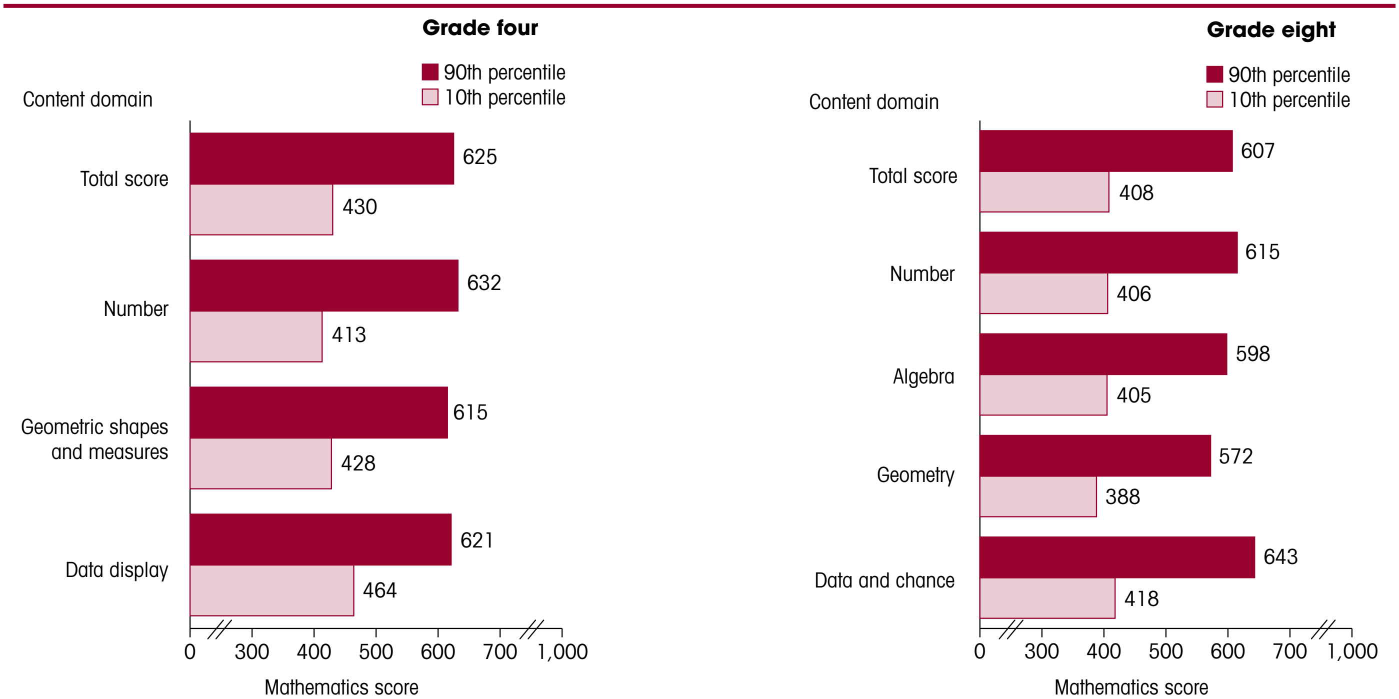
On the three mathematics content domains at grade four, the highest-performing U.S. fourth-graders (90th percentile or higher) scored 632 or higher on the *number* domain, 615 or higher on the *geometric shapes and measures* domain, and 621 or higher on the data display domain (figure 5). The lowest-performing U.S. students (10th percentile or lower) scored 413 or lower on the *number* domain, 428 or lower on the *geometric shapes and measures* domain, and 464 or lower on the data display domain in 2007.

At grade eight, the highest-performing U.S. students (90th percentile or higher) in mathematics scored 607 or higher (table 9). The U.S. 90th percentile score was higher than that of 34 countries and lower than the 90th percentile score in 6 countries: Chinese Taipei, Korea, Singapore, Hong Kong SAR, Japan, and Hungary. The range at the eighth grade in 90th percentile scores was between 427 (Qatar) and 721 (Chinese Taipei). The difference in average scores between the 90th percentile in Chinese Taipei and the United States was 114 score points.

The lowest-performing U.S. eighth-graders (10th percentile or lower) scored 408 or less in 2007 (table 9). The 10th percentile score for U.S. eighth-graders in mathematics was higher than the 10th percentile score in 34 countries and lower than the 10th percentile score in 4 countries: Chinese Taipei, Korea, Singapore, and Japan. The range in 10th percentile scores was between 186 (Qatar) and 475 (Korea). The difference in the cutpoint scores between the lowest-performing students in Korea and the United States was 66 score points.

On the four mathematics content domains at grade eight, the highest-performing U.S. eighth-graders (90th percentile or higher) scored 615 or higher on the *number* domain, 598 or higher on the *algebra* domain, 572 or higher on the *geometry* domain, and 643 or higher on the *data and chance* domain (figure 5). The same general pattern appears to hold among the lowest-performing U.S. students (10th percentile or lower) who scored 406 or lower on the *number* domain, 405 or lower on the *algebra* domain, 388 or lower on the *geometry* domain, and 418 or lower on the *data and chance* domain.

Figure 5. Cutpoints at the 10th and 90th percentile for mathematics content domain scores of U.S. fourth- and eighth-grade students: 2007



NOTE: The United States met guidelines for sample participation rates only after substitute schools were included. The National Defined Population covered 90 percent to 95 percent of National Target Population (see appendix A). Cutpoints are calculated based on distribution of U.S. student scores. The standard errors of the estimates are shown in table E-8 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

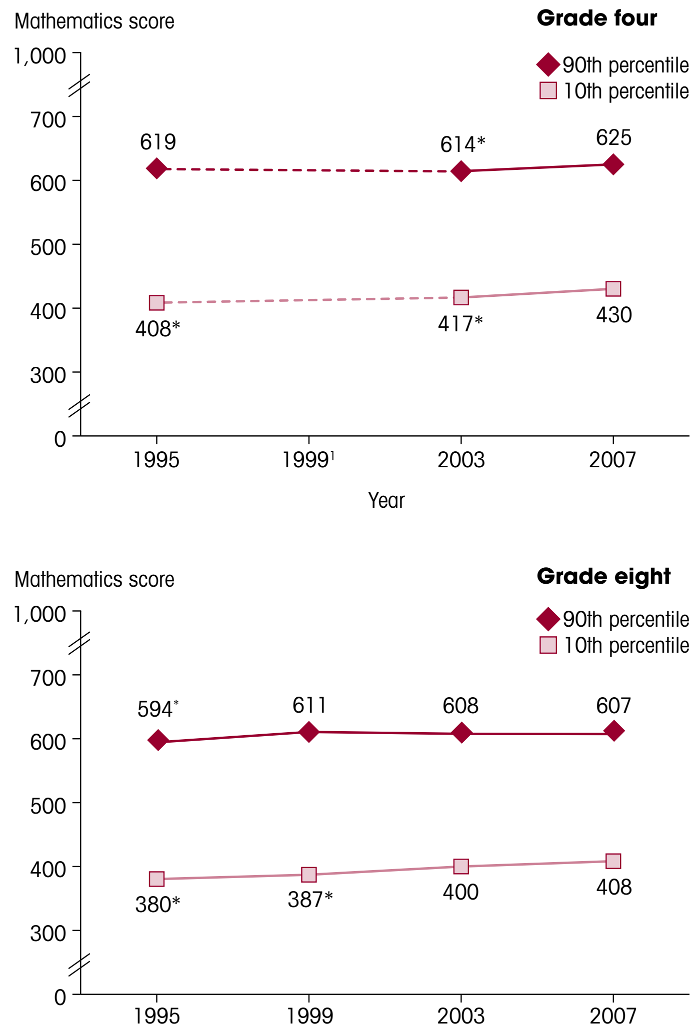
A comparison of 1995, when TIMSS was first administered, and 2007 shows no measurable change in the cutpoint score at the 90th percentile for U.S. fourth graders, the point marking the top 10 percent of students (figure 6). In 2007, the 90th percentile score for U.S. fourth-graders was 625; the 90th percentile score for 1995 was 619. However, a comparison of data from 2003 and 2007 shows there was an increase in the 90th percentile score defining the top-performing students: from 614 to 625. On the other hand, the lowest-performing U.S. fourth graders' showed statistically significant improvement in mathematics: the 10th percentile score increased from 408 in 1995 and 417 in 2003 to 430 in 2007.

At grade eight, both the 90th and 10th percentile scores were higher in 2007 than in 1995 (figure 6). Though the 90th percentile score has been relatively stable over the last three administrations of TIMSS, the 2007 score of 607 was higher than the 1995 score of 594, showing improvement among top students. The 10th percentile score for eighth-graders was higher in 2007 than in 1995 or 1999.

Average scores of male and female students

In 2007, U.S. fourth-grade males outperformed females by 6 score points on average in mathematics (figure 7). In addition to the United States, of the 35 other countries participating at grade four, 20 showed a significant difference in the average mathematics scores of males and females: 12 in favor of males and 8 in favor of females. The difference in average scores between males and females ranged from 37 score points in Kuwait (in favor of females) to 17 score points in Colombia (in favor of males).

Figure 6. Trends in 10th and 90th percentile mathematics scores of U.S. fourth- and eighth-grade students: 1995, 1999, 2003, and 2007



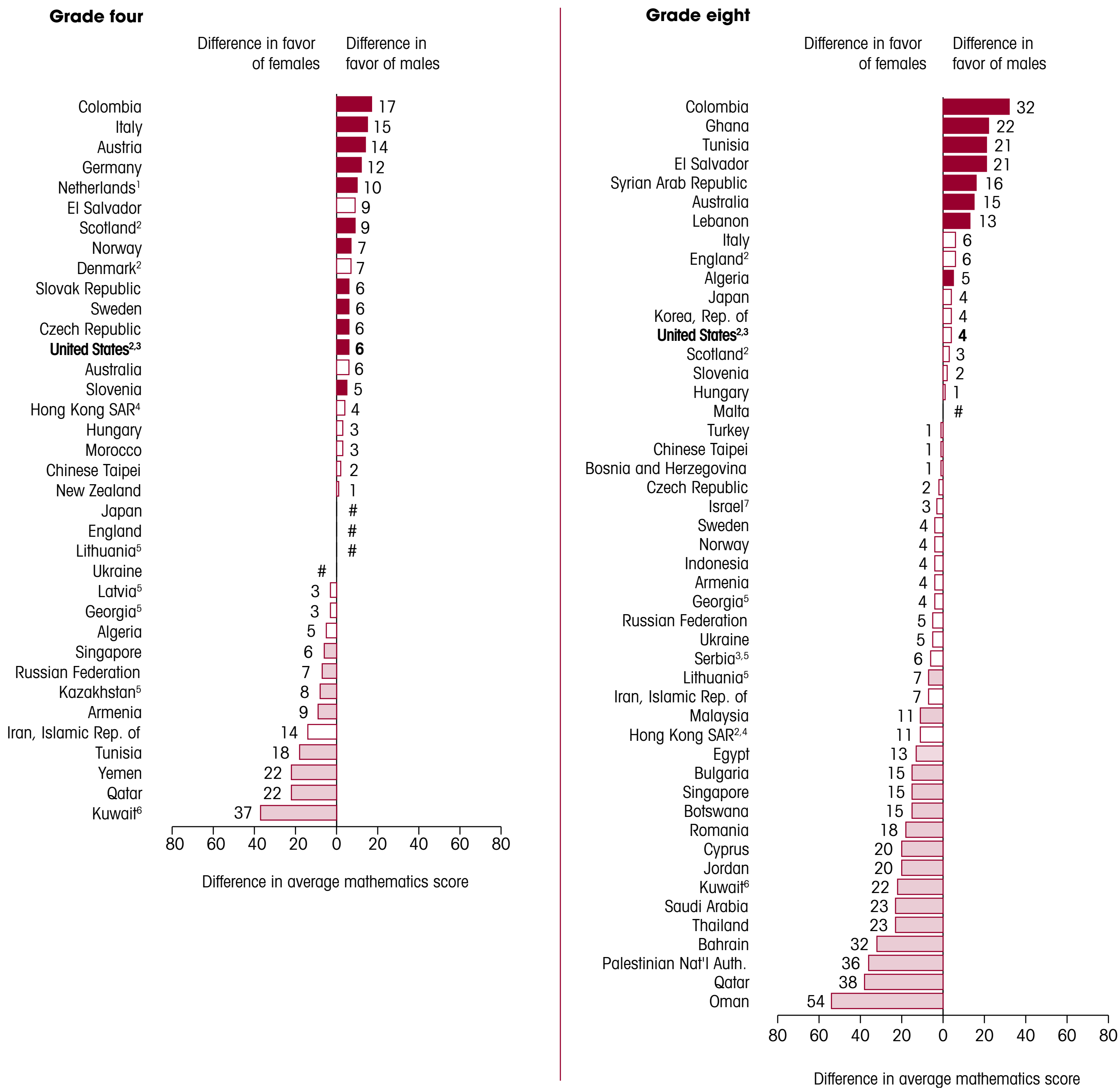
**p* < .05. Percentile cutpoint score is significantly different from 2007 percentile cutpoint score.

[1]No fourth-grade assessment was conducted in 1999.

NOTE: In 2007, the United States met guidelines for sample participation rates only after substitute schools were included. The National Defined Population covered 90 percent to 95 percent of National Target Population (see appendix A). Cutpoints are calculated based on distribution of U.S. student scores. The standard errors of the estimates are shown in table E-9 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995, 1999, 2003, and 2007.

Figure 7. Difference in average mathematics scores of fourth- and eighth-grade students, by sex and country: 2007



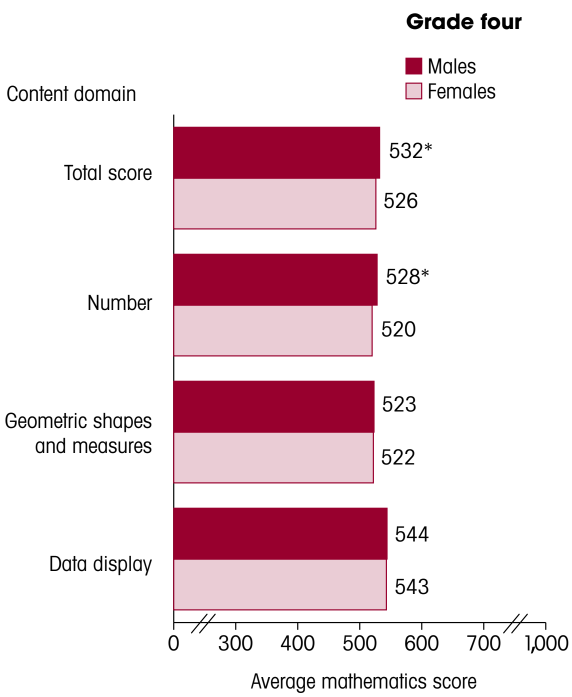
■ Male-female difference in average mathematics scores favors males and is statistically significant ($p < .05$)
□ Male-female difference in average mathematics scores is not measurably different ($p < .05$)
■ Male-female difference in average mathematics scores favors females and is statistically significant ($p < .05$)
Rounds to zero.
[1]Nearly satisfied guidelines for sample participation rates only after substitute schools were included (see appendix A).
[2]Met guidelines for sample participation rates only after substitute schools were included (see appendix A).
[3]National Defined Population covers 90 percent to 95 percent of National Target Population (see appendix A).
[4]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
[5]National Target Population does not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS) (see appendix A).
[6]Kuwait tested the same cohort of students as other countries, but later in 2007, at the beginning of the next school year (see appendix A).
[7]National Defined Population covers less than 90 percent of National Target Population (but at least 77 percent, see appendix A).
NOTE: The standard errors of the estimates are shown in tables E-10 and E-11 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

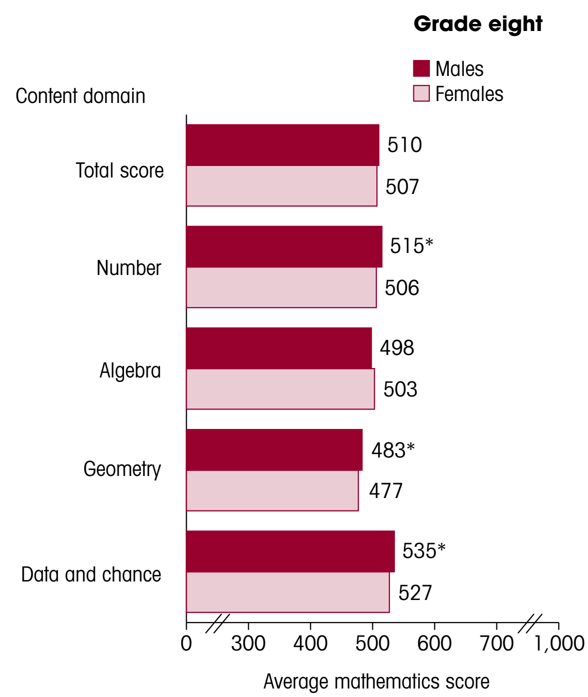
The higher average for U.S. male fourth graders on the total mathematics scale reflects higher average performance on one content area: males outscored females 528 to 520, on average, in *number* (figure 8). There were no measurable sex differences detected in the average scores in either the *geometric shapes and measures* domain or the *data display* domain.

At grade eight, there was no measurable difference in the average mathematics scores of U.S. males and females in 2007 (figure 7). Among the 47 other countries participating in TIMSS at grade eight, 24 showed a difference in the average mathematics scores of males and females: 8 in favor of males and 16 in favor of females. The difference in average scores between males and females ranged from 54 score points in Oman (in favor of females) to 32 score points in Colombia (in favor of males).

Though there was no measurable difference detected in the average mathematics scores of U.S. eighth-grade males and females, U.S. males outperformed U.S. females in three of four mathematics content domains: *number* (515 v. 506), *geometry* (483 v. 477), and *data and chance* (535 v. 527; figure 8).

Figure 8. Average mathematics scores of U.S. fourth- and eighth-grade students, by content domain and sex: 2007



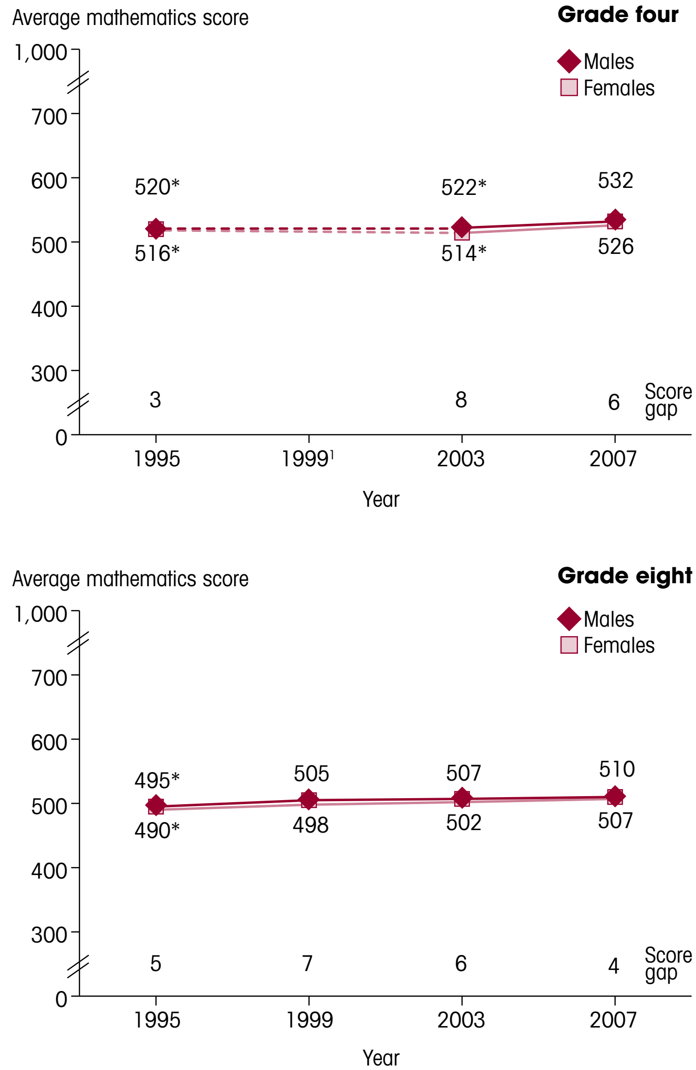


*$p < .05$. Difference between average mathematics scores for males and females is statistically significant and favors males.
NOTE: The United States met guidelines for sample participation rates only after substitute schools were included. The National Defined Population covered 90 percent to 95 percent of National Target Population (see appendix A). The standard errors of the estimates are shown in table E-12 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007

Both U.S. males and females' average scores, at the fourth and eighth grades, were higher in 2007 than in 1995 (figure 9). At grade four, the 2007 average scores of both males and females were higher than their average scores in both 1995 and 2003. U.S. fourth-grade males scored 12 points higher on average in mathematics in 2007 than in 1995 (532 v. 520), and U.S. fourth-grade females scored 10 points higher, on average (526 v. 516).

At grade eight in 2007, U.S. males and females had higher scores, on average, compared to their scores in 1995: by 15 scale score points among males (510 v. 495) and by 17 scale score points among females (507 v. 490; figure 9).

Figure 9. Trends in sex differences in average mathematics scores of U.S. fourth- and eighth-grade students: 1995, 1999, 2003, and 2007



*$p < .05$. Significantly different from 2007.
[1]No fourth-grade assessment was conducted in 1999.
NOTE: In 2007, the United States met guidelines for sample participation rates only after substitute schools were included. The National Defined Population covered 90 percent to 95 percent of National Target Population (see appendix A). The standard errors of the estimates are shown in table E-13 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995, 1999, 2003 and 2007.

Average scores of students of different races and ethnicities

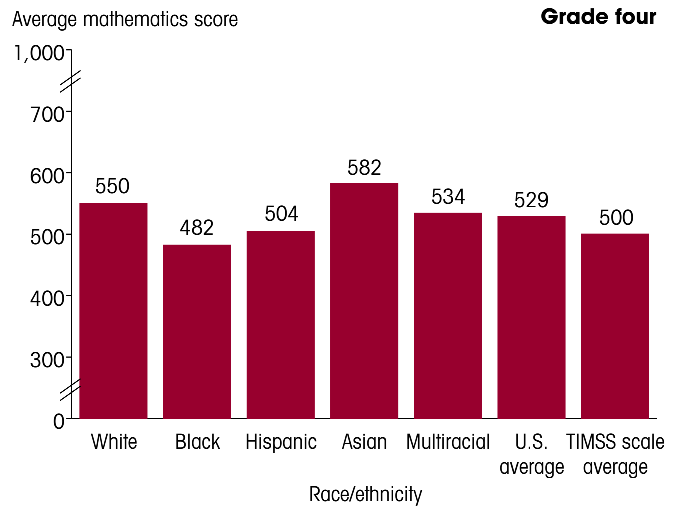
In 2007 U.S. non-Hispanic White, non-Hispanic Asian and multiracial fourth-graders scored higher on average than the TIMSS scale average in mathematics, while U.S. non-Hispanic Black fourth-graders scored lower (figure 10).[10] U.S. Hispanic fourth-graders' average score showed no measurable difference from the TIMSS scale average. In comparison to the U.S. national average, U.S. White and Asian fourth-graders scored higher, on average, while U.S. Black and Hispanic fourth-graders scored lower. U.S. multiracial fourth-graders did not score measurably different from the U.S. national average in mathematics.

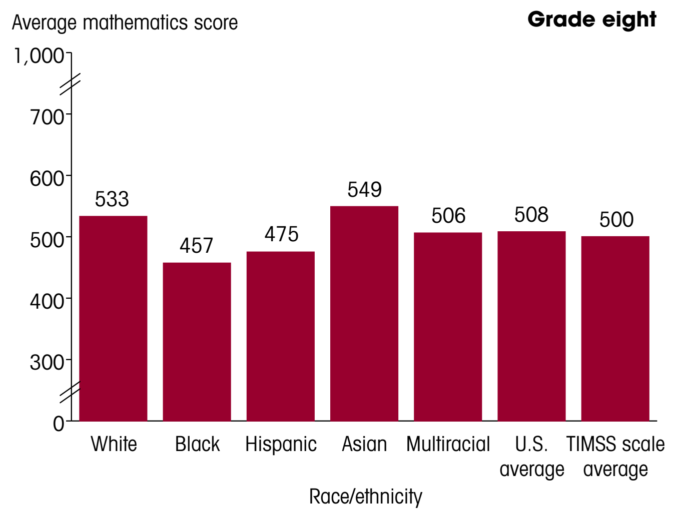
At grade eight, U.S. White, and Asian students scored higher, on average, than both the TIMSS scale average and the U.S. national average in mathematics. On the other hand, U.S. Black and Hispanic eighth-graders scored lower, on average, than the TIMSS scale average and U.S. national average. U.S. multiracial eighth-graders did not score measurably different from either the TIMSS scale average or the U.S. national average score in mathematics.

Over time, U.S. White, Black, Hispanic, and Asian students, in both fourth and eighth grades, have generally shown overall improvement in mathematics (figure 11). At grade four, U.S. White, Black, and Asian students had higher scores in 2007 than in 1995 or 2003; Hispanic students improved their average mathematics score over a shorter period of time, between 2003 and 2007, but not over the 12-year period since 1995.[11] Though in each of the data collection years the differences in the average scores of White fourth-graders and their Black peers were statistically significant, the gap in scores decreased between 1995 and 2007 (84 points v. 67 points). On the other hand, the difference in average scores between White and Asian fourth-graders has reversed and grown over the same period of time, from being in favor of Whites in 1995 (541 v. 525) to being in favor of Asians in 2007 (550 v. 582). There has been no detectable change in the size of the gap in scores between White fourth-graders and their Hispanic classmates.

At grade eight, U.S. White, Black, Hispanic, and Asian students improved in mathematics, on average, when 2007 scores are compared to those from 1995 (figure 11). Black and Hispanic eighth-graders also showed an increase in scores over a shorter period of time, when 2007 is compared to 1999. Though in each of the data collection years the differences in the average scores of White eighth-graders and their Black and Hispanic peers were statistically significant, the sizes of the gap in scores between these groups of students were smaller in 2007 than they were 12 years earlier in 1995 (White v. Black: 76 points v. 97 points; White v. Hispanic: 58 points v. 73 points). There has been no detectable change in the size of the gap in scores between White eighth-graders and their Asian peers.

Figure 10. Average mathematics scores of U.S. fourth- and eighth-grade students, by race/ethnicity: 2007

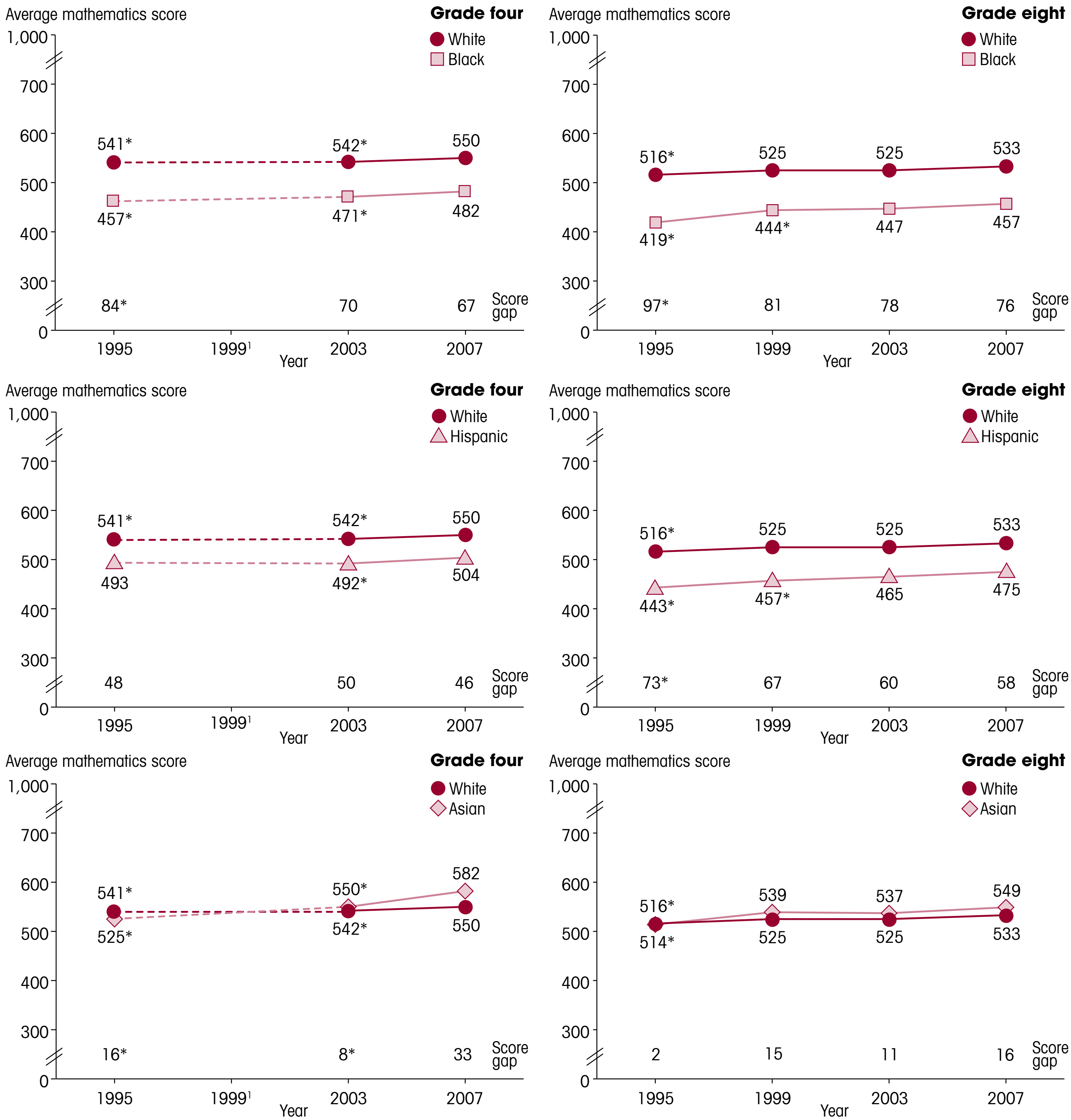




NOTE: Reporting standards were not met for American Indian/Alaska Native and Native Hawaiian/Other Pacific Islander. Black includes African American. Racial categories exclude Hispanic origin. Students who identified themselves as being of Hispanic origin were classified as Hispanic, regardless of their race. Although data for some race/ethnicities are not shown separately because the reporting standards were not met, they are included in the U.S. totals shown throughout the report. The United States met guidelines for sample participation rates only after substitute schools were included. The National Defined Population covered 90 percent to 95 percent of the National Target Population (see appendix A). See appendix A in this report for more information. The standard errors of the estimates are shown in table E-14 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

[10]Black includes African American and Hispanic includes Latino. Race categories exclude Hispanic origin.
[11]The large apparent difference is not statistically significant because of relatively large standard errors.

Figure 11. Trends in differences in average mathematics scores of U.S. fourth- and eighth-grade students, by selected race/ethnicity: 1995, 1999, 2003, and 2007



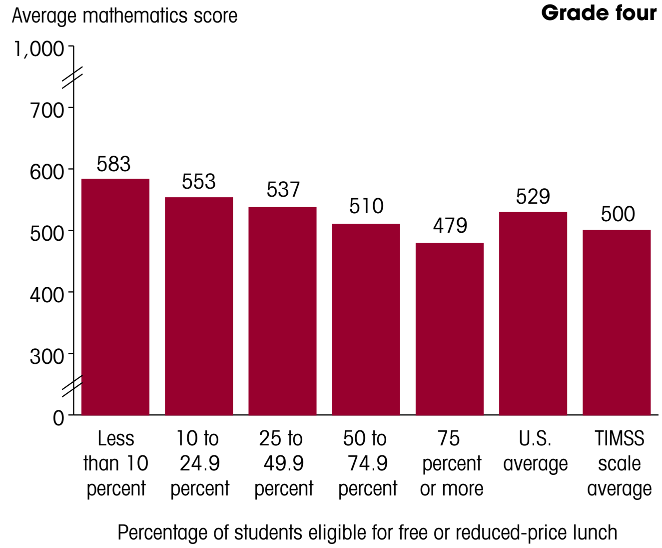
*$p < .05$. Significantly different from 2007.
[1]No fourth-grade assessment was conducted in 1999.
NOTE: Only the four numerically largest racial categories are shown. Multiracial data were not collected in 1995 and 1999. Reporting standards were not met for American Indian/Alaska Native and Native Hawaiian/Other Pacific Islander. Black includes African American. Racial categories exclude Hispanic origin. Students who identified themselves as being of Hispanic origin were classified as Hispanic, regardless of their race. Although data for some race/ethnicities are not shown separately because the reporting standards were not met, they are included in the U.S. totals shown throughout the report. In 2007, the United States met guidelines for sample participation rates only after substitute schools were included. The National Defined Population covered 90 percent to 95 percent of National Target Population (see appendix A). 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. See appendix A in this report for more information. The standard errors of the estimates are shown in table E-15 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1995, 1999, 2003, and 2007.

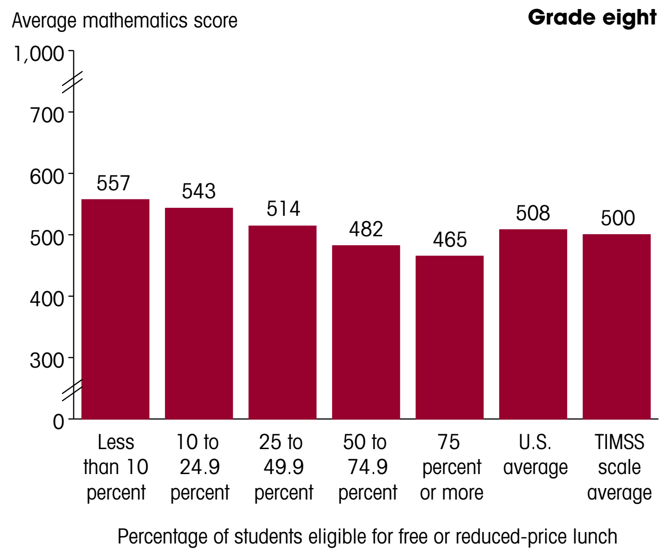
Average scores of students attending public schools of various poverty levels

The U.S. results are also arrayed by the concentration of low-income enrollment in the public schools, as measured by eligibility for free or reduced-price lunch, and shown in relation to the TIMSS scale average and the U.S. national average. In comparison to the TIMSS scale average, the average mathematics score of U.S. fourth graders in the highest poverty public schools (at least 75 percent of students eligible for free or reduced-price lunch) in 2007 was lower (479 v. 500); the average scores of fourth-graders in each of the other categories of school poverty was higher than the TIMSS scale average (figure 12). In comparison to the U.S. national average score, fourth-graders in schools with 50 percent or more students eligible for free or reduced-price lunch scored lower, on average, while those in schools with lower proportions of poor students scored higher, on average, than the U.S. national average.

On average, U.S. eighth-graders in public schools with at least 50 percent eligible for free and reduced price lunch scored lower than the TIMSS scale average in 2007 (482 and 465 v. 500). U.S. eighth-graders attending public schools with fewer than 50 percent of students eligible for the free or reduced-price lunch program scored higher than the TIMSS scale average in mathematics. In comparison to the U.S. national average, U.S. eighth-graders in public schools with fewer than 25 percent of students eligible scored higher in mathematics, on average, while students in public schools with at least 50 percent eligible scored lower, on average.

Figure 12. Average mathematics scores of U.S. fourth- and eighth-grade students, by percentage of students in public school eligible for free or reduced-price lunch: 2007



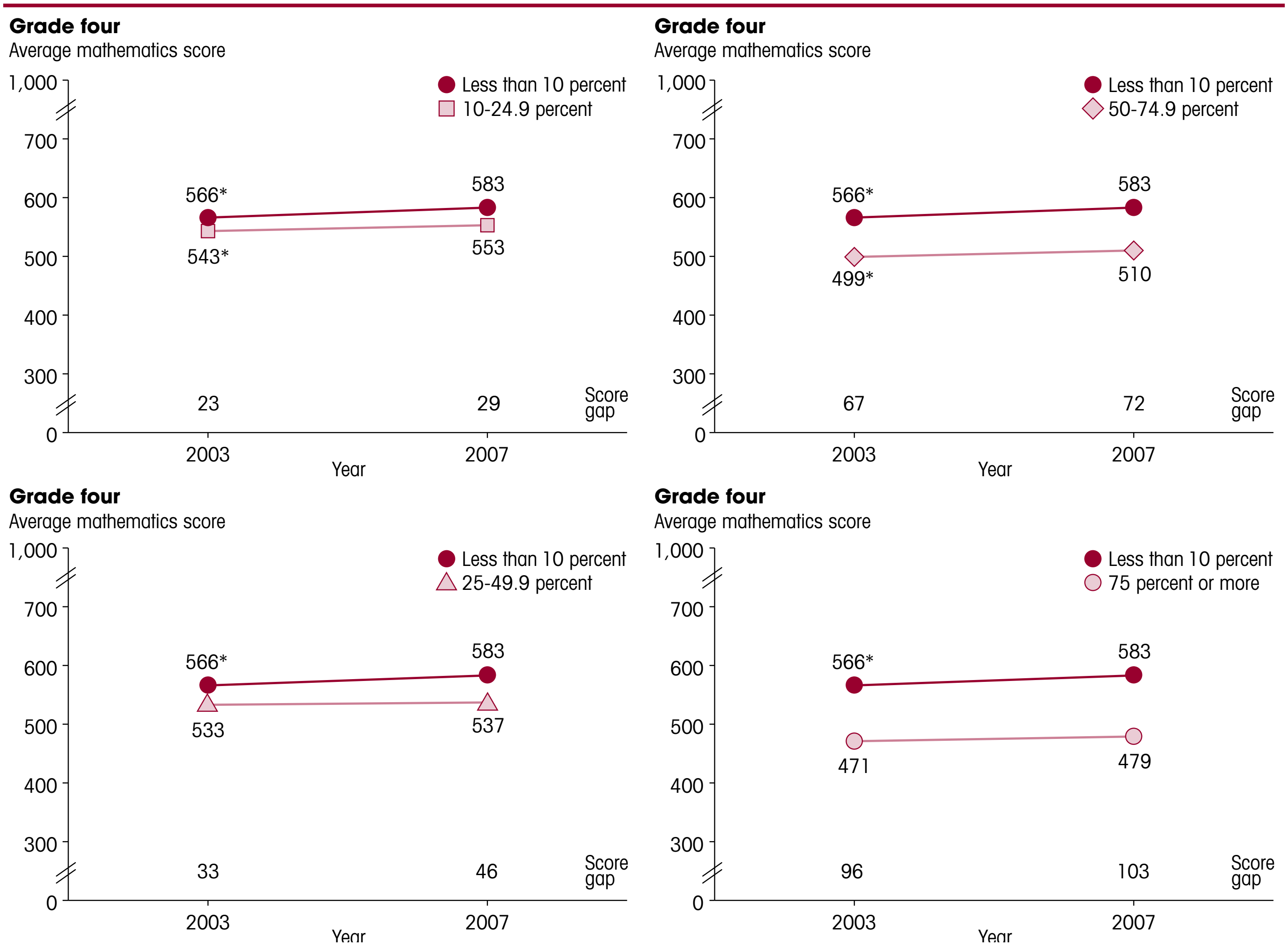


NOTE: Analyses are limited to public schools only, based on school reports of the percentage of students in public school eligible for the federal free or reduced-price lunch program. The United States met guidelines for sample participation rates only after substitute schools were included. The National Defined Population covered 90 percent to 95 percent of the National Target Population (see appendix A). The standard errors of the estimates are shown in table E-16 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Comparisons of scores in 2007 to 2003 showed an inconsistent pattern of improvement in mathematics among U.S. fourth-graders in public schools serving students from various levels of poverty (figure 13).[12] On the one hand, fourth graders in public schools with relatively lower levels of poverty (less than 10 percent to 24.9 percent eligible) and in public schools with relatively higher levels of poverty (50 to almost 75 percent eligible) had higher average mathematics scores in 2007 than in 2003. On the other hand, there was no measurable difference detected in the average scores of students in public schools serving students from medium and the highest level of poverty. Moreover, though the average mathematics scores were higher in 2007, the score gaps evident in the earlier data collections did not appear to diminish over time.[13]

Consistent with the lack of significant change between 1999 and 2007 in eighth-grade mathematics scores overall, students in different types of public schools categorized by poverty also did not show detectable change in performance generally. And, as at grade four, the score gaps evident in earlier data collections did not appear to diminish over time.

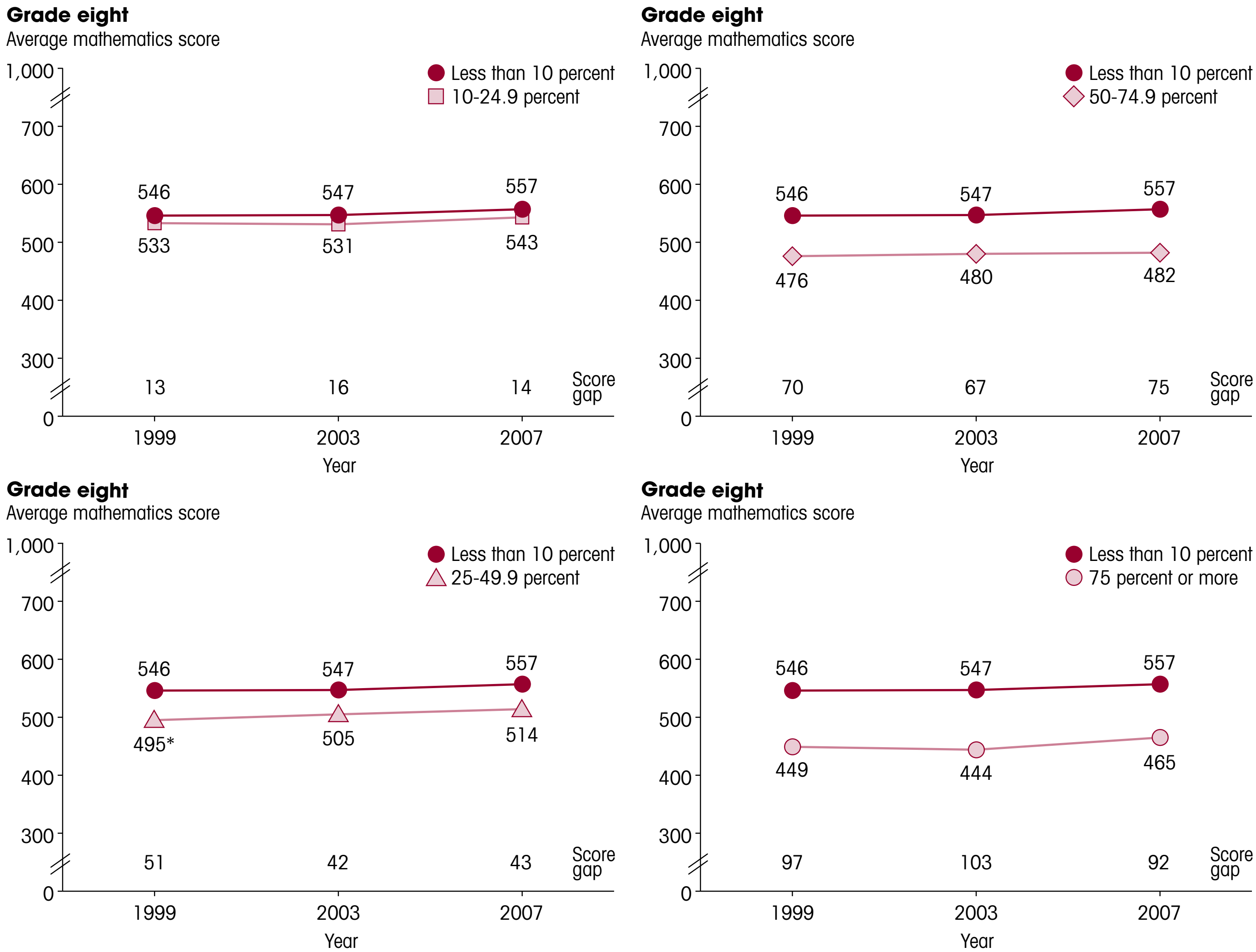
Figure 13. Trends in differences in average mathematics scores of U.S. fourth- and eighth-grade students, by school poverty level: 1999, 2003, and 2007



See notes at end of table.

[12]Information on the percentage of students eligible for the federal free or reduced-price lunch program was not collected in 1995 for either grade. Thus, comparisons over time on this measure are limited to an 8-year period.
[13]Large apparent differences are not statistically significant because of relatively large standard errors.

Figure 13. Trends in differences in average mathematics scores of U.S. fourth- and eighth-grade students, by school poverty level: 1999, 2003, and 2007—Continued



*$p < .05$. Significantly different from 2007.
NOTE: Information on the percentage of students in school eligible for free or reduced-price lunch was not collected in 1995. No fourth-grade assessment was conducted in 1999. Analyses are limited to public schools only, based on school reports of the percentage of students in public school eligible for the federal free or reduced-price lunch program. In 2007, the United States met guidelines for sample participation rates only after substitute schools were included. The National Defined Population covered 90 percent to 95 percent of the National Target Population (see appendix A). The standard errors of the estimates are shown in table E-17 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1999, 2003, and 2007.

Effect size of the difference in average scores

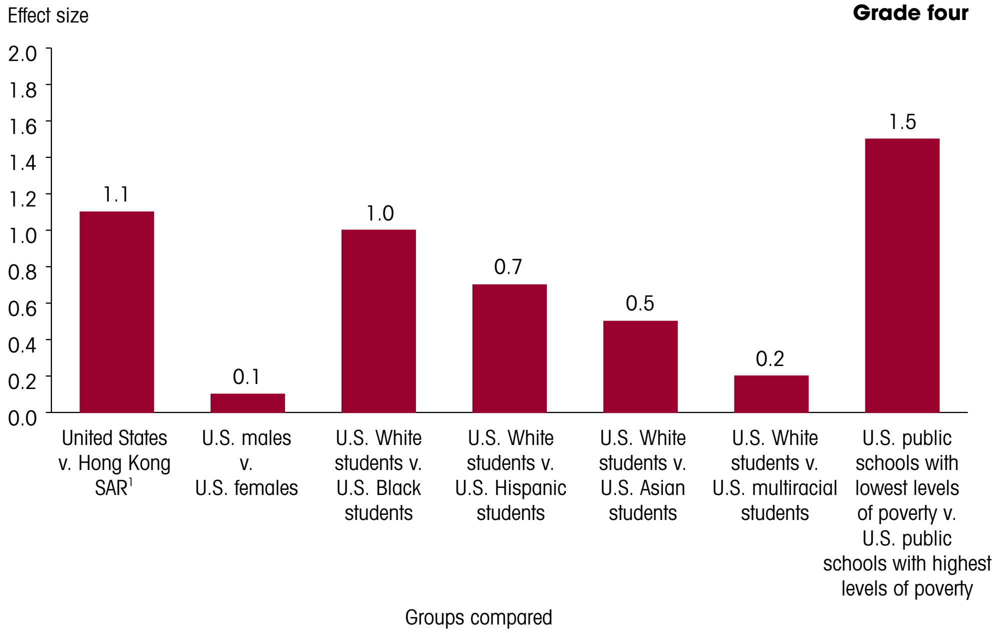
As noted in the introduction, this report includes effect sizes to provide the reader with a sense of the magnitude of the statistically significant differences reported thus far. Statistically significant results do not necessarily indicate those findings that are important or large enough to consider as informing policy or practice. Small differences may be statistically significant, but may not have much practical import.

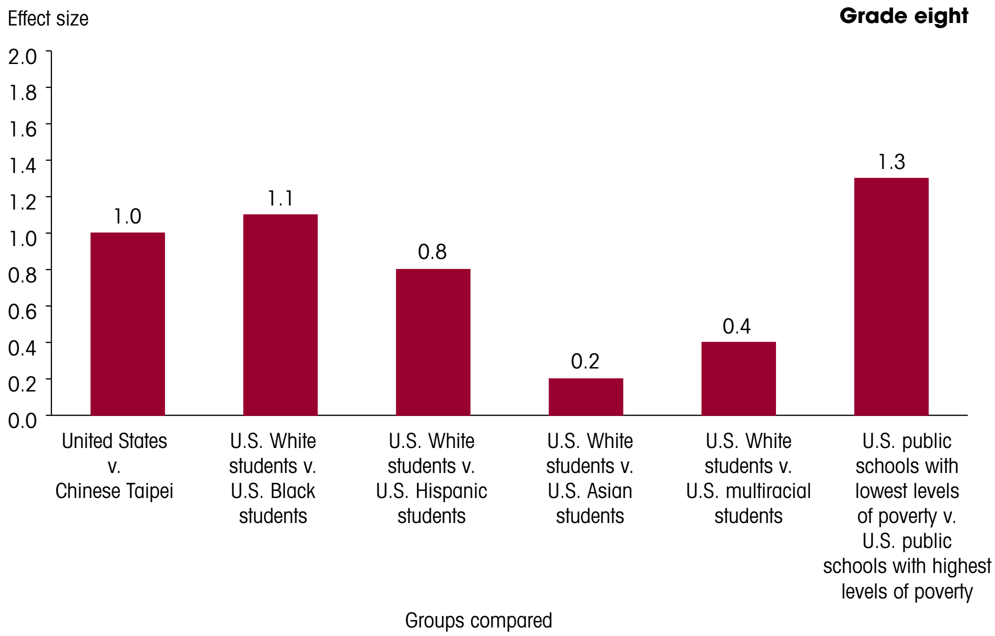
One way of looking at within-country differences in achievement between groups of students is to ask how large these differences are relative to across-country differences between the U.S. national average and an international benchmark, such as the national average for the country with the highest estimated score. As shown previously, the countries with the highest scores outpaced the United States on a number of measures. For example, the difference at grade four between the U.S. average mathematics score (529) and Hong Kong SAR average score (607) was 78 score points (see table 3). The gap between the United States and Hong Kong SAR is also apparent in the percentage of students scoring at the advanced level: 10 percent of U.S. fourth-graders met the advanced international benchmark compared with 40 percent in Hong Kong SAR (see figure 4). Are differences within the United States between groups of students (e.g., by race/ethnicity or poverty concentration in schools) bigger or smaller than these international differences? Effect sizes help make these comparisons. Figure 14 shows the effect size of the difference only for those groups with statistically significant score differences. Appendix A provides a discussion of how effect sizes were calculated.

As shown in figure 14, in grade four mathematics, the effect size of the difference between U.S. White and Black students is roughly the same as the effect size between the United States and Hong Kong SAR, the country with the highest estimated score, while the effect size between U.S. White and Hispanic students is roughly three-fifths the effect size between the United States and Hong Kong SAR. The largest effect size, between U.S. fourth-graders in schools with the lowest and highest poverty levels, is 1.4 times the effect size between the United States and Hong Kong SAR.

At grade eight, the effect size of the difference in mathematics scores between U.S. White and Black students is 1.1 times the effect size between the United States and Chinese Taipei, the country with the highest estimated score. The effect size between U.S. White and Hispanic students is four-fifths the effect size between the United States and Chinese Taipei. The largest effect size, between U.S. eighth-graders in schools with the lowest and highest poverty levels, is 1.3 times the effect size between the United States and Chinese Taipei.

Figure 14. Effect size of difference in average mathematics achievement of fourth- and eighth-grade, by country, sex, race/ethnicity, and school poverty level: 2007





[1]Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.
NOTE: Effect size is shown only for statistically significant differences between group means. Effect size is calculated by dividing the raw difference between group means by the pooled standard deviation (see appendix A). Black includes African American. Racial categories exclude Hispanic origin. Students who identified themselves as being of Hispanic origin were classified as Hispanic, regardless of their race. High-poverty schools are those in which 75 percent or more of students are eligible for the federal free or reduced-price lunch program. Low-poverty schools are those in which less than 10 percent of students are eligible. The United States met guidelines for sample participation rates only after substitute schools were included. The National Defined Population covered 90 percent to 95 percent of the National Target Population. See table E-18 (available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001) for standard deviations of the U.S. and other countries' student populations. See table E-19 (available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001) for standard deviations of U.S. student subpopulations.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Page intentionally left blank