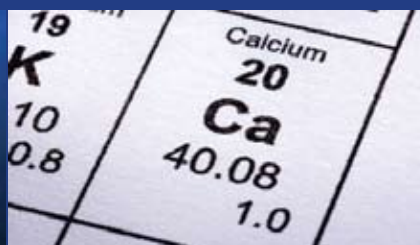


# U.S. Performance Across International Assessments of Student Achievement

Special Supplement to The **Condition of Education 2009**





# U.S. Performance Across International Assessments of Student Achievement

Special Supplement to The Condition of Education 2009

AUGUST 2009

**Stephen Provasnik**

**Patrick Gonzales**

National Center for Education Statistics

**David Miller**

American Institutes for Research

**U.S. Department of Education**

Arne Duncan  
*Secretary*

**Institute of Education Sciences**

John Easton  
*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 nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. 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 Education Statistics  
Institute of Education Sciences  
U.S. Department of Education  
1990 K Street NW  
Washington, DC 20006-5651

**August 2009**

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

**Suggested Citation**

Provasnik, S., Gonzales, P., and Miller, D. (2009). *U.S. Performance Across International Assessments of Student Achievement: Special Supplement to The Condition of Education 2009* (NCES 2009-083). 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**

Stephen Provasnik  
(202) 502-7480  
[Stephen.Provasnik@ed.gov](mailto:Stephen.Provasnik@ed.gov)

# Executive Summary

*The Condition of Education* summarizes important developments and trends in education using the latest available data. The report, which the National Center for Education Statistics (NCES) is required by law to produce, is an indicator report intended for a general audience of readers who are interested in education. The indicators represent a consensus of professional judgment on the most significant national measures of the condition and progress of education for which accurate data are available. For the 2009 edition, NCES prepared a special analysis to take a closer look at U.S. student performance on international assessments.

This special analysis looks at information gathered from recent international studies that U.S. students have participated in: the Progress in International Reading Literacy Study (PIRLS), the Program for International Student Assessment (PISA), and the Trends in International Mathematics and Science Study (TIMSS). PIRLS, sponsored by the International Association for the Evaluation of Educational Achievement (IEA) and first conducted in 2001, assesses the reading performance of 4th-graders every 5 years. PISA, sponsored by the Organization for Economic Cooperation and Development (OECD) and first conducted in 2000, assesses the reading, mathematics, and science literacy of 15-year-old students every 3 years. And TIMSS, sponsored by the IEA and first conducted in 1995, assesses the mathematics and science performance of both 4th- and 8th-graders every 4 years. Not all countries<sup>1</sup> have participated in all three studies or in all administrations of a single study's assessments. All three studies include both developed and developing countries; however, TIMSS and PIRLS have a larger proportion of developing countries participating than PISA does because PISA is principally a study of the member countries of the OECD—an intergovernmental organization of 30 developed countries.

This special analysis examines the performance of U.S. students in reading, mathematics, and science compared with the performance of their peers in other countries that participated in PIRLS, PISA, and TIMSS. It identifies which of these countries have outperformed the United States, in terms of students' average scores and the

percentage of students reaching internationally benchmarked performance levels, and which countries have done so consistently.

## Major findings include:

### Reading

- In PIRLS 2006, the average U.S. 4th-graders' reading literacy score (540) was above the PIRLS scale average of 500, but below that of 4th-graders in 10 of the 45 participating countries, including 3 Canadian provinces (Russian Federation, Hong Kong, Alberta, British Columbia, Singapore, Luxembourg, Ontario, Hungary, Italy, and Sweden).
- Among the 28 countries that participated in both the 2001 and 2006 PIRLS assessments, the average reading literacy score increased in 8 countries and decreased in 6 countries. In the rest of these countries, including the United States, there was no measurable change in the average reading literacy score between 2001 and 2006. The number of these countries that outperformed the United States increased from 3 in 2001 to 7 in 2006.

### Mathematics

- The 2007 TIMSS results showed that U.S. students' average mathematics score was 529 for 4th-graders and 508 for 8th-graders. Both scores were above the TIMSS scale average, which is set at 500 for every administration of TIMSS at both grades, and both were higher than the respective U.S. score in 1995.
- Fourth-graders in 8 of the 35 other countries that participated in 2007 (Hong Kong, Singapore, Chinese Taipei, Japan, Kazakhstan, Russian Federation, England, and Latvia) scored above their U.S. peers, on average; and 8th-graders in 5 of the 47 other countries that participated in 2007 (Chinese Taipei, Korea, Singapore, Hong Kong, and Japan) scored above their U.S. peers, on average.
- Among the 16 countries that participated in both the first TIMSS in 1995 and the most recent TIMSS in 2007, at grade 4, the average mathematics score increased in 8 countries, including in the United States, and decreased in 4 countries. Among the 20 countries that participated in both the 1995 and 2007 TIMSS at grade 8, the average mathematics score increased in 6 countries, including in the United States, and decreased in 10 countries.

<sup>1</sup> The term "country" is used throughout this special analysis as a common name for the range of political entities that have participated in each study. In most cases, participating countries represent an entire nation state, as in the case of the United States. However, in some studies, participating countries represent parts of nation states. For example, several Canadian provinces participated separately in PIRLS 2006, instead of Canada. Likewise, England and Scotland regularly participate separately (instead of the entire United Kingdom) and Belgium regularly participates as two units (Flemish-speaking and French-speaking Belgium) in PIRLS and TIMSS. Similarly, Hong Kong and Macao, which are special administrative regions (SAR) of China, also participate independently.

- In PISA 2006, U.S. 15-year-old students' average mathematics literacy score of 474 was lower than the OECD average of 498, and placed U.S. 15-year-olds in the bottom quarter of participating OECD nations, a relative position unchanged from 2003.
- Fifteen-year-old students in 23 of the 29 other participating OECD-member countries outperformed their U.S. peers.
- There was no measurable change in U.S. 15-year-olds' average mathematics literacy score between 2003 and 2006, in its relationship to the OECD average, or in its relative position to the countries whose scores increased or decreased.

### Science

- The 2007 TIMSS results showed that U.S. students' average science score was 539 for 4th-graders and 520 for 8th-graders. Both scores were above the TIMSS scale average, which is set at 500 for every administration of TIMSS at both grades, but neither was measurably different than the respective U.S. score in 1995.
- Fourth-graders in 4 of the 35 other countries that participated in 2007 (Singapore, Chinese Taipei, Hong Kong, and Japan) scored above their U.S. peers, on average; and 8th-graders

in 9 of the 47 other countries that participated in 2007 (Singapore, Chinese Taipei, Japan, Korea, England, Hungary, the Czech Republic, Slovenia, and the Russian Federation) scored above their U.S. peers, on average.

- While there was no measurable change in the average score of U.S. 4th-graders or 8th-graders in science between 1995 and 2007, among the other 15 countries that participated in the 1995 and 2007 TIMSS at grade 4, the average science score increased in 7 countries and decreased in 5 countries; and among the other 18 countries that participated in both the 1995 and 2007 TIMSS at grade 8, the average science score increased in 5 countries and decreased in 3 countries.

- In PISA 2006, U.S. 15-year-old students' average science literacy score of 489 was lower than the OECD average of 500, and placed U.S. 15-year-olds in the bottom third of participating OECD nations. Fifteen-year-old students in 16 of the 29 other participating OECD-member countries outperformed their U.S. peers in terms of average scores.

Technical notes about the data sources, methodology, and standard errors are included at the end of this report. Special analyses are available on the *The Condition of Education* website (<http://nces.ed.gov/programs/coe>).

# Contents

---

	Page
Executive Summary .....	iii
List of Tables.....	vi
List of Figures .....	vii
<b>Introduction .....</b>	<b>1</b>
<i>Are International Assessment Results Reliable, Valid, and Comparable? .....</i>	<i>6</i>
<b>How Do U.S. Students Compare With Their Peers in Other Countries? .....</b>	<b>7</b>
<b>Reading.....</b>	<b>8</b>
Reading results for 4th-graders .....	8
Reading results for 15-year-olds .....	10
Synthesis of reading results .....	14
<b>Mathematics.....</b>	<b>16</b>
Mathematics results for 4th- and 8th-graders.....	16
Mathematics results for 15-year-olds .....	20
Synthesis of mathematics results .....	21
<i>How Much Variation Is There Between Low and High Performers in Different Countries?.....</i>	<i>30</i>
<b>Science .....</b>	<b>32</b>
Science results for 4th- and 8th-graders.....	32
<i>How Much Does Performance Within the United States Vary by School Poverty? .....</i>	<i>40</i>
Science results for 15-year-olds.....	40
Synthesis of science results .....	44
<b>Summary .....</b>	<b>45</b>
<b>References .....</b>	<b>47</b>
<b>Supplemental Tables .....</b>	<b>49</b>
<b>Appendix A: Technical Notes.....</b>	<b>67</b>

# List of Tables

---

Table	Page
1. Participation in the most recent international assessments, by jurisdiction and grade or age.....	4
2. Average PIRLS scores of fourth-grade students on reading literacy scale and cutpoint scores for bottom and top 10 percent of students in each jurisdiction, by jurisdiction: 2006 .....	9
3. Average PISA scores of 15-year-old students on reading literacy scale and cutpoint scores for bottom and top 10 percent of students in each jurisdiction, by jurisdiction: 2003 .....	13
4. Average TIMSS scores of fourth-grade students in mathematics and cutpoint scores for bottom and top 10 percent of students in each jurisdiction, by jurisdiction: 2007 .....	18
5. Average TIMSS scores of eighth-grade students in mathematics and cutpoint scores for bottom and top 10 percent of students in each jurisdiction, by jurisdiction: 2007 .....	19
6. Average PISA scores of 15-year-old students on mathematics literacy scale and cutpoint scores for bottom and top 10 percent of students in each jurisdiction, by jurisdiction: 2006.....	26
7. Average TIMSS scores of fourth-grade students in science and cutpoint scores for bottom and top 10 percent of students in each jurisdiction, by jurisdiction: 2007.....	33
8. Average TIMSS scores of eighth-grade students in science and cutpoint scores for bottom and top 10 percent of students in each jurisdiction, by jurisdiction: 2007.....	34
9. Average PISA scores of 15-year-old students on science literacy scale and cutpoint scores for bottom and top 10 percent of students in each jurisdiction, by jurisdiction: 2006 .....	41
A-1. Average PIRLS scores of fourth-grade students on combined reading literacy scale, by jurisdiction: 2001 and 2006 .....	50
A-2. Average PISA scores of 15-year-old students on combined reading literacy scale, by jurisdiction: 2000, 2003, and 2006 .....	52
A-3. Average TIMSS mathematics scores of fourth-grade students on combined mathematics scale, by jurisdiction: 1995, 2003, and 2007.....	54
A-4. Average TIMSS mathematics scores of eighth-grade students on combined mathematics scale, by jurisdiction: 1995, 1999, 2003, and 2007.....	56
A-5. Average PISA scores of 15-year-old students on combined mathematics literacy scale, by jurisdiction: 2000, 2003, and 2006 .....	58
A-6. Average TIMSS science scores of fourth-grade students on combined science scale, by jurisdiction: 1995, 2003, and 2007 .....	60
A-7. Average TIMSS science scores of eighth-grade students on combined science scale, by jurisdiction: 1995, 1999, 2003, and 2007 .....	62
A-8. Average PISA scores of 15-year-old students on combined science literacy scale, by jurisdiction: 2000, 2003, and 2006 .....	64



# List of Figures

---

Figure	Page
1. Percentage of fourth-grade students reaching the PIRLS international benchmarks in reading, by jurisdiction: 2006 .....	11
2. Change in average PIRLS reading literacy scores of fourth-grade students in selected jurisdictions, by jurisdiction: 2001 to 2006 .....	12
3. Percentage distribution of 15-year-old students on PISA reading literacy scale, by proficiency level and jurisdiction: 2003 .....	15
4. Change in average PISA reading literacy scores of 15-year-old students in selected jurisdictions, by jurisdiction: 2000 to 2003 .....	17
5. Percentage of fourth-grade students reaching the TIMSS international benchmarks in mathematics, by jurisdiction: 2007 .....	22
6. Percentage of eighth-grade students reaching the TIMSS international benchmarks in mathematics, by jurisdiction: 2007 .....	23
7. Change in average TIMSS mathematics scores of fourth-grade students in selected jurisdictions, by jurisdiction: 1995 to 2007 .....	24
8. Change in average TIMSS mathematics scores of eighth-grade students in selected jurisdictions, by jurisdiction: 1995 to 2007 .....	25
9. Percentage distribution of 15-year-old students on PISA mathematics literacy scale, by proficiency level and jurisdiction: 2006 .....	28
10. Change in average PISA mathematics scores of 15-year-old students in selected jurisdictions, by jurisdiction: 2003 to 2006 .....	29
A-1. Distribution of PISA scores for 15-year-old students on the mathematics literacy scale at the 10th and 90th percentiles, by OECD jurisdiction: 2006 .....	30
A-2. Distribution of PISA scores for 15-year-old students on the mathematics literacy scale at the 10th and 90th percentiles, by non-OECD jurisdiction: 2006 .....	31
11. Percentage of fourth-grade students reaching the TIMSS international benchmarks in science, by jurisdiction: 2007 .....	36
12. Percentage of eighth-grade students reaching the TIMSS international benchmarks in science, by jurisdiction: 2007 .....	37
13. Change in average TIMSS science scores of fourth-grade students in selected jurisdictions, by jurisdiction: 1995 to 2007 .....	38
14. Change in average TIMSS science scores of eighth-grade students in selected jurisdictions, by jurisdiction: 1995 to 2007 .....	39
15. Percentage distribution of 15-year-old students on PISA science literacy scale, by proficiency level and jurisdiction: 2006 .....	43

---

*This page intentionally left blank.*

---

---

# Introduction

---

# Introduction

The National Center for Education Statistics (NCES) is congressionally mandated to report on the state of education in the United States and other countries.<sup>1</sup> To carry out this mission, NCES participates in several international assessments to measure how the performance of U.S. students and adults compares with that of their counterparts in other countries. This special analysis looks closely at the information NCES has gathered from recent international studies that U.S. students have participated in: the Progress in International Reading Literacy Study (PIRLS), the Program for International Student Assessment (PISA), and the Trends in International Mathematics and Science Study (TIMSS).<sup>2</sup>

This special analysis describes the most recent results from these international studies as well as trends in the results, when possible. It is organized by subject area into three parts—reading, mathematics, and science. For each subject area, the following topics are addressed:

- How does the performance of U.S. students compare with their peers in other countries?
- Which countries' students outperform U.S. students, and which have done so consistently?
- How has the performance of U.S. students changed over time?
- To what extent has the performance of U.S. students changed over time relative to their peers in high-performing countries?

The three international studies examined in this special analysis periodically measure one or more dimensions of the performance of students at different ages or grade levels. PIRLS, sponsored by the International Association for the Evaluation of Educational Achievement (IEA) and first conducted in 2001, assesses the reading performance of 4th-graders every 5 years. PISA, sponsored by the Organization for Economic Cooperation and Development (OECD) and first conducted in 2000, assesses the reading, mathematics, and science literacy of 15-year-old students every 3 years.<sup>3</sup> And TIMSS,

sponsored by the IEA and first conducted in 1995, assesses the mathematics and science performance of both 4th- and 8th-graders every 4 years.<sup>4</sup> Although organized and run by two different international organizations, these three assessments all provide score results on a scale of 0 to 1,000, with a standard deviation of 100.<sup>5</sup> However, scores from different assessment studies (e.g., PISA and TIMSS) cannot be compared with each other directly because of differences in each study's purpose, subject matter, and assessed grade or age. Thus all comparisons in this special analysis are between countries that participated in the same study.

It is important to point out here that the term “country” is used for simplicity's sake throughout this special analysis as a common name for the range of political entities that have participated in each study. In most cases, participating countries represent an entire nation state, as in the case of the United States. However, in some studies, participating countries represent parts of nation states. For example, several Canadian provinces participated separately in PIRLS 2006, instead of Canada. Likewise, England and Scotland regularly participate separately (instead of the entire United Kingdom) and Belgium regularly participates as two units (Flemish-speaking and French-speaking Belgium) in PIRLS and TIMSS. Similarly, Hong Kong and Macao, which are special administrative regions (SAR) of China, also participate independently.<sup>6</sup>

Not all countries have participated in all three studies or in all administrations of a single study's assessments.<sup>7</sup> Table 1 lists the participating countries in the most recent administration of each assessment, and the supplemental tables 1–8 list participating countries in all administrations of the assessments. All three studies include both developed and developing countries; however, TIMSS and PIRLS have a larger proportion of developing countries participating than PISA does because PISA is principally a study of the member countries of the OECD—an intergovernmental organization of 30 developed countries.

<sup>1</sup> Most recently mandated in the Education Sciences Reform Act of 2002.

<sup>2</sup> This special analysis does not examine the results of international assessments of adult literacy, in which the United States has also participated. The reason for this is that the results of the 2002 Adult Literacy and Lifeskills Survey (ALL), the last assessment of adult literacy, have already been described in *The Condition of Education 2006* special analysis (see <http://nces.ed.gov/programs/coe/2006/analysis/index.asp>), and the next assessment, the Program for the International Assessment of Adult Competencies (PIAAC), is not scheduled to be conducted until 2011.

<sup>3</sup> While PISA assesses each subject area every 3 years, each assessment cycle focuses on one particular subject. In 2000, the focus was on reading literacy; in 2003, on mathematics literacy; in 2006, on science literacy. In 2009, the focus is on reading literacy again.

<sup>4</sup> In 1995, TIMSS also assessed students at the end of secondary school: in some countries, these were students in grade 10, while in others these were students in grade 14. In the United States, 12th-graders were assessed.

<sup>5</sup> For details about scale scores, see appendix A.

<sup>6</sup> In some assessments, subnational units such as states and regions have been benchmarking participants either instead of or in addition to the entire nation-state. For a list of U.S. states that have participated in international assessments, independent of the nation as a whole, see appendix A.

Note that official designation of participating entities may differ between assessments. For example, in TIMSS, the official designation for Hong Kong is “Hong Kong SAR,” while in PISA, it is “Hong Kong-China.” In the text of this special analysis, shortened forms of official designations are used; but in the figures and tables, the assessment's full official designations are used.

<sup>7</sup> Countries vary over time in the assessments in which they participate for a variety of reasons, including individual countries' perceptions of the benefits and costs of each assessment, and the specific logistic challenges of administering the assessments.

Differences in the set of countries that participate in an assessment can affect how well the United States appears to do internationally when results are released. One reason for this is that average student performance in developed countries tends to be higher than in developing countries. As a result, the extent to which developing countries participate in an assessment can affect the international average of participating countries as well as the relative position of one country compared with the others.<sup>8</sup> To deal with this problem, none of the international assessments calculates an international “average” score based on results of all participating countries. Instead, PISA calculates an *OECD average*, for each PISA subject area, that is based only on the results of the OECD-member countries. All OECD-member countries participate in PISA; therefore, PISA ostensibly calculates this average based on a consistent group of

countries.<sup>9</sup> TIMSS and PIRLS, on the other hand, do not calculate an average based on the results of any of the participating countries; they report results relative to the midpoint of each assessment’s reporting scale, which they call the “scale average.”<sup>10</sup>

All differences reported in this special analysis are statistically “measurable” or significant at a 95 percent level of confidence. All *t*-tests supporting this special analysis were done without adjustments for multiple comparisons. It is also important to note that the purpose of this special analysis is to provide descriptive information; thus, complex interactions and causal relationships have not been explored. Readers are cautioned not to make causal inferences based on the results presented here.

---

<sup>8</sup> Specifically, as more developing countries participate in a study, the lower the international average tends to be and the higher the participating developed countries appear to be ranked.

---

<sup>9</sup> While all OECD-member countries’ results are used to calculate PISA’s OECD average, the number of countries used to calculate this average has actually increased. For example, in 2000, results for The Netherlands were not used to calculate the average because of its low response rates. In addition, between 2000 and 2003, the total number of countries in the OECD increased from 28 to 30 when the Slovak Republic and Turkey joined the OECD.

<sup>10</sup> Although the IEA uses the label “scale average,” this is not actually a calculated average: it equals 500 because that is the “average” value on the assessment’s 1,000-point scale. For a more detailed explanation of scale scores and scale averages, see appendix A.

**Table 1. Participation in the most recent international assessments, by jurisdiction and grade or age**

Jurisdiction	PIRLS 2006	PISA 2006	TIMSS 2007 <sup>1</sup>	
	4th grade	15-year-olds	4th grade	8th grade
<b>OECD<sup>3</sup></b>				
Australia		•	•	•
Austria	•	•	•	
Belgium		•		
Flemish	•			
French	•			
Canada		•		
Alberta	•		○	
British Columbia	•		○	○
Nova Scotia	•			
Ontario	•		○	○
Quebec	•		○	○
Czech Republic		•	•	•
Denmark	•	•	•	
Finland		•		
France	•	•		
Germany	•	•	•	
Greece		•		
Hungary	•	•	•	•
Iceland	•	•		
Ireland		•		
Italy	•	•	•	•
Japan		•	•	•
Korea, Republic of		•		•
Luxembourg	•	•		
Mexico		•		
Netherlands	•	•	•	
New Zealand	•	•	•	
Norway	•	•	•	•
Poland	•	•		
Portugal		•		
Slovak Republic	•	•	•	
Spain	•	•		
Sweden	•	•	•	•
Switzerland		•		
Turkey		•		•
United Kingdom		•		
England	•		•	•
Scotland	•		•	•
<b>United States<sup>4</sup></b>	•	•	•	•
<b>OECD country total</b>	<b>19</b>	<b>30</b>	<b>16</b>	<b>12</b>
<b>Total OECD jurisdictions</b>	<b>25</b>	<b>30</b>	<b>16</b>	<b>12</b>
	PIRLS 2006	PISA 2006	TIMSS 2007 <sup>2</sup>	
	4th grade	15-year-olds	4th grade	8th grade
<b>Non-OECD</b>				
Algeria			•	•
Argentina		•		
Armenia			•	•
Azerbaijan		•		
Bahrain				•
Bosnia and Herzegovina				•
Botswana				•
Brazil		•		
Bulgaria	•	•		•
Chile		•		
Chinese Taipei	•	•	•	•
Colombia		•	•	•
Croatia		•		
Cyprus				•

See notes at end of table.

**Table 1. Participation in the most recent international assessments, by jurisdiction and grade or age—Continued**

	PIRLS 2006	PISA 2006	TIMSS 2007 <sup>2</sup>	
	4th grade	15-year-olds	4th grade	8th grade
<b>Non-OECD—Continued</b>				
Egypt				•
El Salvador			•	•
Estonia		•		
Georgia	•		•	•
Ghana				•
Hong Kong-China	•	•	•	•
Indonesia	•	•		•
Iran, Islamic Republic of	•		•	•
Israel	•	•		•
Jordan		•		•
Kazakhstan			•	
Kyrgyz Republic		•		
Kuwait	•		•	•
Latvia	•	•	•	
Lebanon				•
Liechtenstein		•		
Lithuania	•	•	•	•
Macedonia	•			
Macao-China		•		
Malaysia				•
Malta				•
Moldova, Republic of	•			
Montenegro, Republic of		•		
Morocco	•		•	
Oman				•
Palestinian National Authority				•
Qatar	•	•	•	•
Romania	•	•		•
Russian Federation	•	•	•	•
Saudi Arabia				•
Serbia, Republic of		•		•
Singapore	•		•	•
Slovenia	•	•	•	•
South Africa	•			
Syrian Arab Republic				•
Thailand		•		•
Trinidad and Tobago	•			
Tunisia		•	•	•
Ukraine			•	•
Uruguay		•		
Yemen			•	
<b>Non-OECD country total</b>	<b>20</b>	<b>27</b>	<b>20</b>	<b>36</b>
<b>Total jurisdictions</b>	<b>45</b>	<b>57</b>	<b>36</b>	<b>48</b>

<sup>1</sup> Four Canadian provinces (Alberta, British Columbia, Ontario, and Quebec), as well as the Basque region of Spain, Dubai of the United Arab Emirates, and Massachusetts and Minnesota of the United States, participated in TIMSS 2007 as benchmarking participants and are not included in the total counts shown.

<sup>2</sup> Although Mongolia and Morocco participated at both grades, the quality of the data for Mongolia was not well documented at both grades and there was a problem with the participation rates for Morocco at the eighth grade. For more information, see Olson, J.F., Martin, M.O., & Mullis, I.V.S. (Eds.). (2008). *TIMSS 2007 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

<sup>3</sup> There are a total of 30 countries in the Organization for Economic Cooperation and Development (OECD). An OECD country is counted in the OECD total if it participated as a single/entire country (which is generally the case) or if it participated as one or more component jurisdictions of the country (e.g., England and Scotland as representing the United Kingdom).

<sup>4</sup> PISA 2006 reading literacy results were not reported for the United States because of an error in printing the test booklets.

NOTE: A bullet "•" indicates participation in the particular assessment. An open bullet "◦" indicates jurisdictions that participated as "benchmarking participants."

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), 2006; IEA, Trends in International Mathematics and Science Study (TIMSS), 2007; Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

## Are International Assessment Results Reliable, Valid, and Comparable?

Since the United States began participating in comparative international assessments in the 1960s, the number and scope of international assessments have grown. In addition, the quality of the data they collect has improved because of the international adoption of ever more rigorous technical standards and monitoring, along with growing expertise in the international community relating to assessment design (National Research Council 2002, p. 9). The international organizations that sponsor international student assessments—the OECD and the International Association for the Evaluation of Educational Achievement (IEA)—go to great lengths to ensure that their assessment results are reliable, valid, and comparable among participating countries.<sup>11</sup>

For each study, the sponsoring international organization verifies that all participating countries select a nationally representative sample of schools and, from those schools, randomly select either classrooms of a particular grade or students of the particular age or grade targeted by the assessment. To ensure comparability, target grades or ages are clearly defined. For example, in TIMSS, at the upper grade level, countries are required to sample students in the grade that corresponds to the end of 8 years of formal schooling, providing that the mean age of the students at the time of testing is at least 13.5 years. Moreover, comparisons by age are carefully chosen to ensure that students at the target age are enrolled in school at comparable rates across countries. For example, PISA elected to study 15-year-old students because 15 is the oldest age at which enrollment rates remain around 90 percent or higher in most developed countries, including the United States (OECD 2008, table C2.1). For students 16 and older, attendance is not universally compulsory.

Not all selected schools and students choose to participate in the assessment; and certain students, such as some with mental or physical disabilities, may not be able to take the assessment. Thus the sponsoring international organizations check each country's *participation rates* (for schools and students) and *exclusion rates* (at the school level and within schools) to ensure they meet established target rates in order for the country's results to be reported.<sup>12</sup>

In addition to international requirements and verification to ensure valid samples, the sponsoring international organizations require compliance with standardized

procedures for the preparation, administration, and scoring of assessments. Countries are required to send quality-control monitors to visit schools and scoring centers to report on compliance with the standardized procedures. Furthermore, independent international quality-control monitors visit a sample of schools in each country to ensure that the international standards are met.

Results for countries that fail to meet the required participation rates or other international requirements are footnoted with explanations of the specific failures (e.g., “only met guidelines for sample participation rates after substitute schools were included”), shown separately in the international reports (e.g., listed in a separate section at the bottom of a table), or omitted from the international reports and datasets (as happened to The Netherlands' PISA results in 2000, the United Kingdom's PISA results in 2003, and Morocco's TIMSS 2007 results at grade 8). For more details on international requirements, see appendix A.

Every participating country is involved in a thorough process of developing the assessment. The national representatives from each country review every test item to be included in the assessment to ensure that each item adheres to the internationally agreed-upon *framework* (the outline of the topics and skills that should be assessed in a particular subject area), and that each item is culturally appropriate for their country. Each country translates the assessment into their own language or languages, and external translation companies independently review each country's translations.

A “field test” (a small-scale, trial run of the assessment) is then conducted in the participating countries to see if any items were biased because of national, social, or cultural differences. Statistical analyses of the item data are also conducted to check for evidence of differences in student performance across countries that could indicate a linguistic or conceptual translation problem. Problematic items may be dropped from the final pool of items or scaled differently.

When this process is complete, the main assessment instruments are created. Each assessment “instrument” consists of the instructions, the same number of “blocks” of items (each block is a small set of selected items from the final pool of items), and a student background questionnaire. (Additional questionnaires are often prepared and administered to the students' teachers, parents, and/or school principal.) The instruments are then administered to the sampled students in each of the participating countries at comparable times.

For more details on the development and administration of the international assessments, see the Technical Reports produced for each assessment.

<sup>11</sup> For complete details on the methods instituted to ensure data quality and comparability, see OECD 2008; Martin et al. 2007; and Olson, Martin, and Mullis 2008.

<sup>12</sup> The United States also conducts its own nonresponse bias analysis if school participation rates are below 85 percent. For more details about nonresponse bias analysis, see appendix A.



# **How Do U.S. Students Compare With Their Peers in Other Countries?**

# U.S. Students Compared to Peers in Other Countries

This section presents key findings from PIRLS, PISA, and TIMSS and is organized, by subject area, into three parts: reading, mathematics, and science. For each subject area, the assessments in that subject are described and their similarities and differences are highlighted. Then for each assessment in that subject,

- the U.S. average (mean) student score is compared with those of the other participating countries;
- the threshold or *cutpoint* score for “the top 10 percent” of U.S. students (technically the score of U.S. students at the 90th percentile) is compared with the cutpoint score for the top 10 percent of students in other countries;
- the cutpoint score for the bottom 10 percent of U.S. students is compared with the cutpoint score for the bottom 10 percent of students in other countries;
- the percentage of students reaching the highest international benchmark or highest level of proficiency set by each assessment is compared; and
- changes in these measures over time for the United States and for the top-scoring countries are examined, when possible.

These data are described to provide a broader understanding of the performance of U.S. students compared to their peers around the world than is gained by just knowing average scores. Specifically, knowing the cutpoint scores for the top and bottom 10 percent of students tells us how well the highest and lowest performing students do in each country and how wide a range there is in student performance within each country. This range, in turn, provides important contextual information to understand whether a country that outperforms the United States scores higher on account of the performance of its students overall, of mostly its top-performing students, or of mostly its low-performing students. In contrast, comparing the percentage of students who reach the same international benchmarks or levels of proficiency provides information on the extent to which a country’s education system brings student performance up to standardized levels that have been internationally established.

After these data have been described for each assessment, you will find references for more detailed information and a brief synthesis of all the assessment results in the subject area.

## Reading

Both PIRLS and PISA assess aspects of reading skills, but they differ in terms of whom they assess and what they assess.

PIRLS assesses 4th-graders and is designed to reflect the curriculum of participating countries. PIRLS asks students to read two texts—either two literary texts (narrative fiction, generally drawn from children’s books), two informational texts (typically excerpts from biographies, step-by-step instructions, or scientific or non-fiction materials), or one of each type. It then asks students about a dozen questions (both multiple-choice and open-ended “constructed response”) about the texts that range from identifying the place, time, and actions of the main characters or events to interpreting how characters might feel, why events occurred, or what the passage means overall (e.g., does the story teach a lesson?).<sup>13</sup>

PISA assesses 15-year-old students and does not explicitly focus on curricular outcomes; rather it focuses on cognitive skills and the application of reading to problems within a real-life context. Thus it presents students with a range of texts that they are likely to encounter as young adults, such as excerpts from government forms, brochures, newspaper articles, instruction manuals, books, and magazines. For each text, it then usually asks each student 3–5 questions (both multiple choice and constructed response) to measure the extent to which students can retrieve information, interpret a text, reflect on a text, and evaluate its author’s rhetorical choices.<sup>14</sup> In years when PISA focuses on reading, students receive between 12 and 24 reading texts (depending on the particular cluster of items in their particular test booklet); when PISA focuses on mathematics or science, students receive about 7 reading texts.

## Reading results for 4th-graders

In PIRLS 2006, the average U.S. 4th-graders’ reading literacy score (540) was above the PIRLS scale average of 500, but below that of 4th-graders in 10 of the 45 participating countries, including 3 Canadian provinces (Russian Federation, Hong Kong, Alberta, British Columbia, Singapore, Luxembourg, Ontario, Hungary, Italy, and Sweden)<sup>15</sup> (table 2). The top 10 percent of U.S. 4th-graders scored 631 or higher, a cutpoint score below that of the top 10 percent of students in 8 countries. The bottom 10 percent of U.S. 4th-graders scored 441 or lower, a cutpoint score below that of the bottom 10 percent of students in 13 countries.

<sup>13</sup> Examples of PIRLS items can be viewed at [http://nces.ed.gov/pubs2008/2008017\\_2.pdf](http://nces.ed.gov/pubs2008/2008017_2.pdf).

<sup>14</sup> Examples of PISA reading items can be viewed at <http://www.oecd.org/dataoecd/30/17/39703267.pdf>, pages 288 to 291.

<sup>15</sup> Countries are listed in rank order from highest to lowest score for countries outperforming the United States.

**Table 2. Average PIRLS scores of fourth-grade students on reading literacy scale and cutpoint scores for bottom and top 10 percent of students in each jurisdiction, by jurisdiction: 2006**

Jurisdiction	Average score	Cutpoint score	
		Bottom 10 percent	Top 10 percent
<b>All jurisdictions</b>	<b>500</b> ▼	<b>398</b> ▼	<b>597</b> ▼
Russian Federation	565 ▲	474 ▲	649 ▲
Hong Kong, SAR <sup>1</sup>	564 ▲	486 ▲	637
Canada, Alberta	560 ▲	472 ▲	645 ▲
Canada, British Columbia	558 ▲	467 ▲	645 ▲
Singapore	558 ▲	456	652 ▲
Luxembourg	557 ▲	470 ▲	641 ▲
Canada, Ontario	555 ▲	463 ▲	644 ▲
Hungary	551 ▲	459	637
Italy	551 ▲	462 ▲	637
Sweden	549 ▲	465 ▲	627
Germany	548	463 ▲	627
Belgium (Flemish) <sup>2</sup>	547	474 ▲	616 ▼
Bulgaria	547	437	647 ▲
Netherlands <sup>2</sup>	547	478 ▲	613 ▼
Denmark	546	454	629
Canada, Nova Scotia	542	442	634
Latvia	541	460 ▲	619 ▼
<b>United States<sup>2</sup></b>	<b>540</b>	<b>441</b>	<b>631</b>
England	539	423 ▼	645 ▲
Austria	538	454	617 ▼
Lithuania	537	461 ▲	608 ▼
Chinese Taipei	535	451	613 ▼
Canada, Quebec	533	450	611 ▼
New Zealand	532 ▼	415 ▼	637
Slovak Republic	531 ▼	433	617 ▼
Scotland <sup>2</sup>	527 ▼	420 ▼	624
France	522 ▼	433	605 ▼
Slovenia	522 ▼	427 ▼	608 ▼
Poland	519 ▼	417 ▼	612 ▼
Spain	513 ▼	420 ▼	599 ▼
Israel	512 ▼	369 ▼	626
Iceland	511 ▼	417 ▼	594 ▼
Belgium (French)	500 ▼	409 ▼	585 ▼
Moldova, Republic of	500 ▼	406 ▼	584 ▼
Norway <sup>3</sup>	498 ▼	409 ▼	579 ▼
Romania	489 ▼	362 ▼	597 ▼
Georgia	471 ▼	369 ▼	565 ▼
Macedonia, Republic of	442 ▼	305 ▼	571 ▼
Trinidad and Tobago	436 ▼	295 ▼	563 ▼
Iran, Islamic Republic of	421 ▼	295 ▼	539 ▼
Indonesia	405 ▼	301 ▼	504 ▼
Qatar	353 ▼	228 ▼	479 ▼
Kuwait	330 ▼	186 ▼	476 ▼
Morocco	323 ▼	181 ▼	468 ▼
South Africa	302 ▼	141 ▼	500 ▼

▲ Score is higher than U.S. score.

▼ Score is lower than U.S. score.

<sup>1</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>2</sup> Met Progress in International Reading Literacy Study (PIRLS) guidelines for sample participation rates only after substitute schools were included.

<sup>3</sup> Did not meet guidelines for sample participation rates after substitute schools were included.

NOTE: Jurisdictions are ordered on the basis of average scores, from highest to lowest. Reading literacy scores are reported on a scale from 0 to 1,000. A cutpoint score is the threshold score for an established level of performance. The cutpoint scores for students in the top 10 percent is the 90th percentile score within the jurisdiction. The cutpoint score for students in the bottom 10 percent is the 10th percentile score within the jurisdiction. 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.

SOURCE: Mullis, I.V.S., Martin, M.O., Kennedy, A.M., and Foy, P. (2007). *PIRLS 2006 International Report: IEA's Progress in International Reading Literacy Study in Primary Schools in 40 Countries*, exhibit 1.1. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College; and International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), 2006, unpublished tabulations (November 2008).

PIRLS has developed four international benchmarks to help analyze the range of students' performance in reading within each participating country, with the highest, or *Advanced*, benchmark set at 625 score points.<sup>16</sup> For PIRLS 2006, students reaching the Advanced benchmark could interpret figurative language; integrate ideas across a text to provide interpretations of a character's traits, intentions, and feelings; and provide full text-based support for their interpretations.<sup>17</sup>

In 2006, twelve percent of U.S. 4th-graders reached this benchmark (figure 1). Eight participating countries, including 3 Canadian provinces, had a higher percentage of 4th-graders reaching this benchmark, ranging from 19 to 15 percent: Singapore, Russian Federation, Alberta, Bulgaria, British Columbia, Ontario, England, and Luxembourg. Among the countries with a greater percentage of students than the United States reaching the Advanced benchmark, two did not have average student scores higher than the United States: Bulgaria and England.<sup>18</sup>

### *Change over time*

Among the 28 countries that participated in both the 2001 and 2006 PIRLS assessments, the average reading literacy score increased in 8 countries and decreased in 6 countries (figure 2). In the rest of these countries, including the United States, there was no measurable change in the average reading literacy score between 2001 and 2006. The number of these countries that outperformed the United States increased from 3 in 2001 to 7 in 2006.<sup>19</sup> Three of the countries that outperformed the United States in 2006 (Hong Kong, the Russian Federation, and Singapore) had scored below the United States, on average, in 2001. In contrast, in 2 of the 6 countries where 4th-graders showed measurable declines (England and The Netherlands), 4th-graders outperformed their U.S. peers in 2001, but were not measurably different than their U.S. peers in 2006.

PIRLS will be offered again in 2011. Results from the PIRLS 2006 assessment can be found in Baer et al. (2007;

available at <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2008017>) and Mullis et al. (2007; available at [http://timss.bc.edu/pirls2006/intl\\_rpt.html](http://timss.bc.edu/pirls2006/intl_rpt.html)). For more information on PIRLS, see <http://nces.ed.gov/surveys/pirls/>.

### Reading results for 15-year-olds

PISA 2006 reading literacy results are not reported for the United States because of an error introduced when the test booklets were printing.<sup>20</sup> Thus the reading literacy results described here come from the PISA 2000 and 2003.

In PISA 2003, U.S. 15-year-old students' average reading literacy score of 495 was not measurably different than the OECD average of 494, and placed U.S. 15-year-olds in the middle third of participating OECD nations (table 3). Fifteen-year-old students in 9 of the 29 other participating OECD-member countries outperformed their U.S. peers (as did 15-year-olds in 2 of the 11 non-OECD countries that participated) in terms of average scores. U.S. 15-year-olds in the top 10 percent scored 622 or higher, a cutpoint score below that of the top 10 percent of students in 7 countries (all OECD countries). The bottom 10 percent of U.S. 15-year-olds scored 361 or lower, a cutpoint score below that of the bottom 10 percent of students in 9 OECD countries and 3 non-OECD countries.

PISA has developed five levels of proficiency to help analyze the range of students' performance in reading within each participating country.<sup>21</sup> The highest level of proficiency identifies students who can complete sophisticated reading tasks, such as managing information that is difficult to find in unfamiliar texts; showing detailed understanding of such texts and inferring which information in the text is relevant to the task; and being able to evaluate critically and build hypotheses, draw on specialized knowledge, and accommodate concepts that may be contrary to expectations. For PISA 2003, the highest level of proficiency corresponds with a score at or above 625 score points.<sup>22</sup>

<sup>16</sup> See figure 1 for the cut scores established for the other three international benchmarks. For details about all the international benchmarks, see Mullis et al. (2007), chapter 2.

<sup>17</sup> The IEA set international benchmarks for PIRLS based on an analysis of score points. The score points for each benchmark remain the same across assessments; however, the configuration of items that define what students reaching a benchmark can do may vary slightly from one assessment to the next. For more details on the IEA's benchmarks and how they differ from PISA's levels of proficiency, see appendix A.

<sup>18</sup> There was no measurable difference between the average student scores in the United States and in Bulgaria and England.

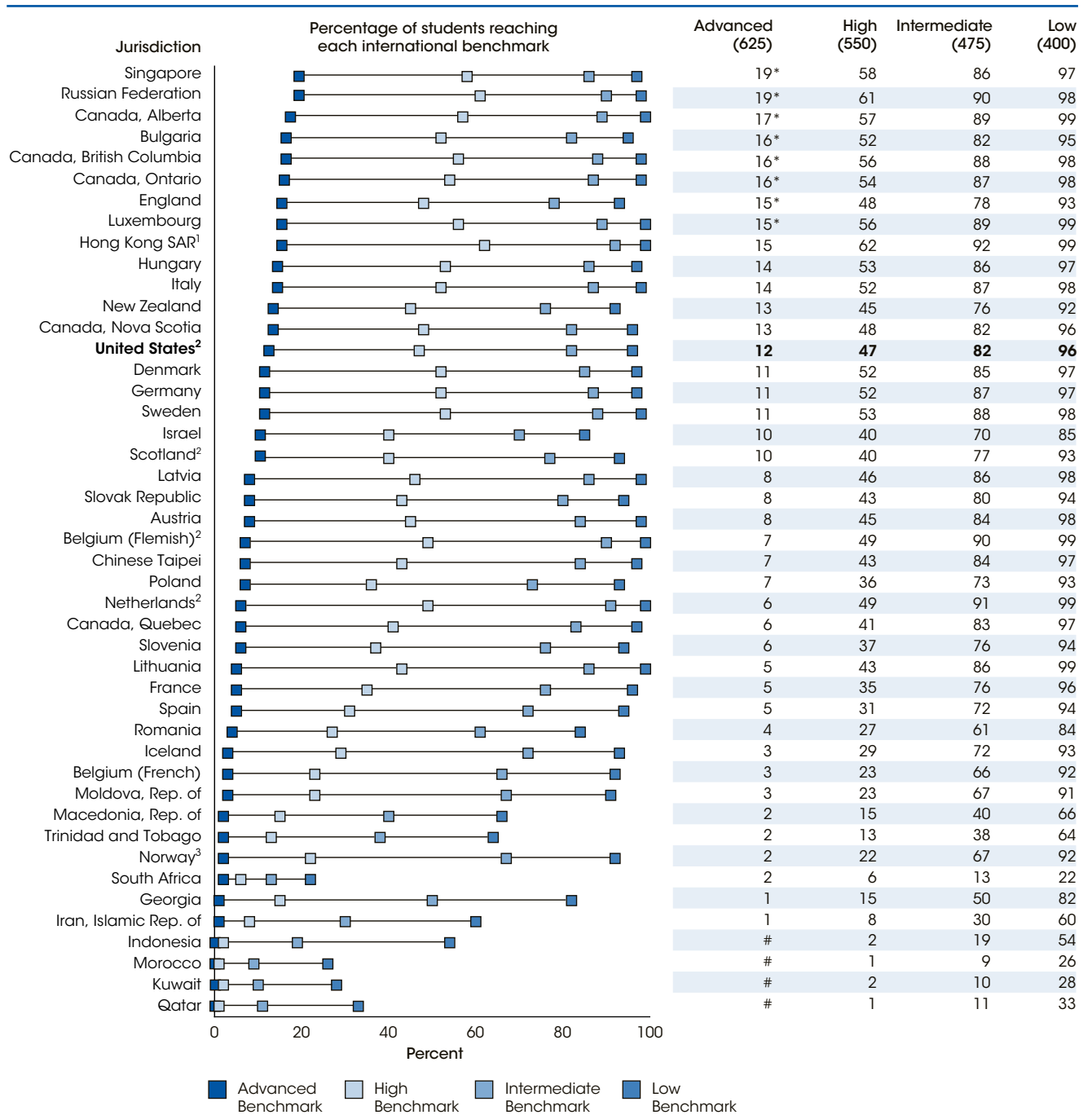
<sup>19</sup> Luxembourg and the Canadian provinces of Alberta and British Columbia also outperformed the United States in 2006, but they did not participate in 2001.

<sup>20</sup> In various parts of the U.S. PISA 2006 reading literacy assessment test booklet, students were incorrectly instructed to refer to the passage on the "opposite page" when students actually needed to turn back to the previous page to see the necessary passage.

<sup>21</sup> See figure 3 for the cut scores for all five levels of proficiency in 2003. For details about all five levels, see OECD 2004, pp.272–79.

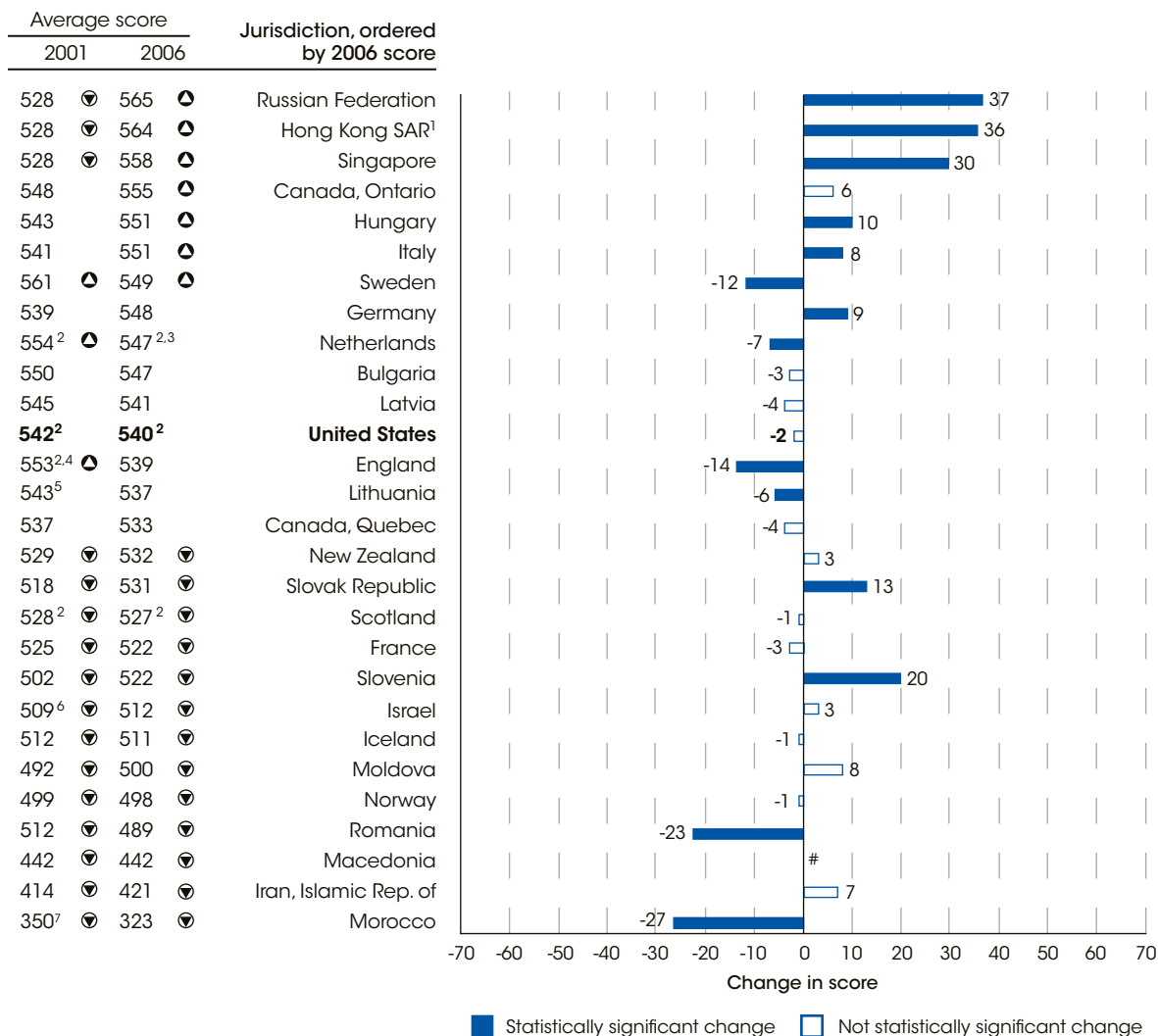
<sup>22</sup> PISA has defined levels of proficiency based on specific student proficiencies. These specific student proficiencies remain the same across assessments; however, the score point threshold for students who demonstrate these specific student proficiencies may vary slightly from assessment to assessment. For more details on PISA's levels of proficiency and how they differ from the IEA's benchmarks, see appendix A.

**Figure 1. Percentage of fourth-grade students reaching the PIRLS international benchmarks in reading, by jurisdiction: 2006**



\*  $p < .05$ . Percentage reaching advanced benchmark is measurably higher than U.S. percentage.  
 # Rounds to zero.  
<sup>1</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.  
<sup>2</sup> Met guidelines for sample participation rates only after substitute schools were included.  
<sup>3</sup> Did not meet guidelines for sample participation rates after substitute schools were included.  
 NOTE: Jurisdictions are ordered by the percentage of students who reached the advanced international benchmark on a scale of 0 to 1,000. The International Association for the Evaluation of Educational Achievement (IEA) set international benchmarks for the Progress in International Reading Literacy Study (PIRLS) based on an analysis of score points. Students who did not reach benchmarks are not shown nor are students whose placement relative to the benchmarks was uncertain. Because IEA benchmarks are anchored by scores, it is possible to calculate the percentage of students who reach each benchmark's score threshold with accuracy but not the percentage distribution of students by benchmarks. Apparent differences may not be statistically significant.  
 SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), 2006.

**Figure 2. Change in average PIRLS reading literacy scores of fourth-grade students in selected jurisdictions, by jurisdiction: 2001 to 2006**



▲ Average score was higher than the U.S. average score.

▼ Average score was lower than the U.S. average score.

# Rounds to zero.

<sup>1</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>2</sup> Met guidelines for sample participation rates only after substitute schools were included.

<sup>3</sup> Did not meet guidelines for sample participation rates after substitute schools were included.

<sup>4</sup> National Defined Population covers less than 95 percent of National Target Population.

<sup>5</sup> National Target Population does not cover all of International Target Population because coverage falls below 65 percent.

<sup>6</sup> National Defined Population covers less than 80 percent of National Target Population.

<sup>7</sup> Nearly satisfied guidelines for sample participation rates after substitute schools were included.

NOTE: Selected jurisdictions include those that participated in both PIRLS 2001 and PIRLS 2006. Average reading literacy scores are reported on a scale from 0 to 1,000. Tests for statistical significance take into account the standard errors for scores in both years. Because the size of standard errors can vary, a small difference between the scores of one jurisdiction may be significant while a larger difference between the scores of another jurisdiction may not be significant. Detail may not sum to totals due to rounding.

SOURCE: Baer, J., Baldi, S., Ayotte, K., and Green, P. (2007). *The Reading Literacy of U.S. Fourth-Grade Students in an International Context: Results From the 2001 and 2006 Progress in International Reading Literacy Study (PIRLS)* (NCES 2008-017), figure 3 and table 2. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC; Ogle, L., Sen, A., Pahlke, E., Jocelyn, L., Kastberg, D., Roey, S., and Williams, T. (2003). *International Comparisons in Fourth-Grade Reading Literacy: Findings from the Progress in International Reading Literacy Study (PIRLS) of 2001* (NCES 2003-073), figure 3. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC.

**Table 3. Average PISA scores of 15-year-old students on reading literacy scale and cutpoint scores for bottom and top 10 percent of students in each jurisdiction, by jurisdiction: 2003**

Jurisdiction	Average score	Cutpoint score	
		Bottom 10 percent	Top 10 percent
<b>All OECD jurisdictions</b>	<b>494</b>	<b>361</b>	<b>617</b>
<b>OECD</b>			
Finland	543 ▲	437 ▲	641 ▲
Korea, Republic of	534 ▲	428 ▲	634 ▲
Canada	528 ▲	410 ▲	636 ▲
Australia	525 ▲	395 ▲	644 ▲
New Zealand	522 ▲	381 ▲	652 ▲
Ireland	515 ▲	401 ▲	622
Sweden	514 ▲	390 ▲	631 ▲
Netherlands	513 ▲	400 ▲	621
Belgium	507 ▲	355	635 ▲
Norway	500	364	625
Switzerland	499	373	615
Japan	498	355	624
Poland	497	374	616
France	496	367	614
<b>United States</b>	<b>495</b>	<b>361</b>	<b>622</b>
Denmark	492	376 ▲	600 ▼
Iceland	492	362	612 ▼
Germany	491	341 ▼	624
Austria	491	354	617
Czech Republic	489	362	607 ▼
Hungary	482 ▼	361	597 ▼
Spain	481 ▼	354	597 ▼
Luxembourg	479 ▼	344 ▼	601 ▼
Portugal	478 ▼	351	592 ▼
Italy	476 ▼	341 ▼	598 ▼
Greece	472 ▼	333 ▼	599 ▼
Slovak Republic	469 ▼	348	587 ▼
Turkey	441 ▼	324 ▼	562 ▼
Mexico	400 ▼	274 ▼	521 ▼
<b>Non-OECD</b>			
Liechtenstein	525 ▲	405 ▲	636
Hong Kong-China	510 ▲	397 ▲	608 ▼
Macao-China	498	409 ▲	583 ▼
Latvia	491	372	603 ▼
Russian Federation	442 ▼	319 ▼	558 ▼
Uruguay	434 ▼	272 ▼	587 ▼
Thailand	420 ▼	322 ▼	520 ▼
Serbia and Montenegro <sup>1</sup>	412 ▼	306 ▼	516 ▼
Brazil	403 ▼	256 ▼	542 ▼
Indonesia	382 ▼	282 ▼	478 ▼
Tunisia	375 ▼	251 ▼	497 ▼

▲ Score is higher than U.S. score.

▼ Score is lower than U.S. score.

<sup>1</sup> The Republics of Serbia and Montenegro were a united jurisdiction for the 2003 Program for International Student Assessment (PISA). NOTE: Because PISA is principally an Organization for Economic Cooperation and Development (OECD) study, the results for non-OECD jurisdictions are displayed separately from those of the OECD countries. The United Kingdom is not included: because of low response rates, 2003 data for the United Kingdom are omitted from PISA reports. Jurisdictions are ordered on the basis of average scores, from highest to lowest within the OECD countries and the non-OECD jurisdictions. Reading literacy scores are reported on a scale from 0 to 1,000. A cutpoint score is the threshold score for an established level of performance. The cutpoint scores for students in the top 10 percent is the 90th percentile score within the jurisdiction. The cutpoint score for students in the bottom 10 percent is the 10th percentile score within the jurisdiction. 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.

SOURCE: Organization for Economic Cooperation and Development (OECD). (2004). *Learning for Tomorrow's World: First Results from PISA 2003*, table 6.2. Paris: Author.

In 2003, nine percent of U.S. 15-year-old students performed at this level (figure 3). The same 7 countries whose top 10 percent of students outperformed U.S. students had a higher percentage of 15-year-old students who performed at this level. The percentages of students performing at this level in countries that outperformed the United States ranged from 16 percent in New Zealand to 11 percent in Sweden.

### *Change over time*

Among the 32 countries that participated in both PISA 2000 and PISA 2003, the average reading literacy score increased in 4 countries and decreased in 5 countries (figure 4).<sup>23</sup> For the United States, there was no measurable change in the average reading literacy score between 2000 and 2003 or in its relationship to the OECD average.<sup>24</sup> However, among the countries that participated in both PISA 2000 and PISA 2003, the number that outperformed the United States increased from 7 to 9.<sup>25</sup>

The 4 countries whose scores increased (Latvia, Liechtenstein, Luxembourg, and Poland) were all outperformed by the United States in 2000; although, in 2003, Luxembourg scored lower than the United States, Latvia and Poland were not measurably different than the United States, and Liechtenstein outscored the United States. Among the countries whose scores decreased between 2003 and 2000, only Japan saw its score fall relative to the United States: Japan outperformed the United States in reading literacy in 2000 but was not measurably different in 2003. The score for Sweden was not measurably different between 2000 and 2003, but in 2003 its score was measurably higher than the U.S. score whereas in 2000 it was not.

From 2000 to 2006, the average PISA reading literacy score increased in 4 countries that participated in both the 2000 and 2006 PISA (Korea, Poland, Liechtenstein, and Latvia) (OECD 2007). At the same time, 10 countries had lower reading literacy scores in 2006 than in 2000 (9 OECD countries, Spain, Japan, Iceland, Norway, Italy, France, Australia, Greece, and Mexico, and one non-OECD country, the Russian Federation.)

<sup>23</sup> Note that statistical comparisons between PISA 2000 scores and PISA 2003 scores need to properly account for linking error. See appendix A for details. Because of low response rates, data for The Netherlands were not discussed for PISA 2000 nor were data for the United Kingdom for PISA 2003.

<sup>24</sup> Large standard errors for the United States in 2000 may account at least in part for the fact that the U.S. reading literacy score was not measurably different between 2000 and 2003 as well as the fact that the score was not different from the OECD average in 2000.

<sup>25</sup> The total number of countries that outperformed the United States in 2000 includes the United Kingdom, which had no official score in 2003 because of low response rates. The total number of countries that outperformed the United States in 2003 includes The Netherlands, which had no official score in 2000 because of technical problems, and Hong Kong, which did not participate in 2000.

PISA is being repeated in 2009. More detailed information on the PISA 2003 results can be found in Lemke et al. (2004; available at <http://nces.ed.gov/pubs2005/2005003.pdf>); and in OECD 2004 (available at <http://www.oecd.org/dataoecd/1/60/34002216.pdf>). More detailed information on the PISA 2006 results can be found in OECD 2007 (available at <http://www.oecd.org/dataoecd/30/17/39703267.pdf>). For more information on PISA, see <http://nces.ed.gov/surveys/pisa/>.

### Synthesis of reading results

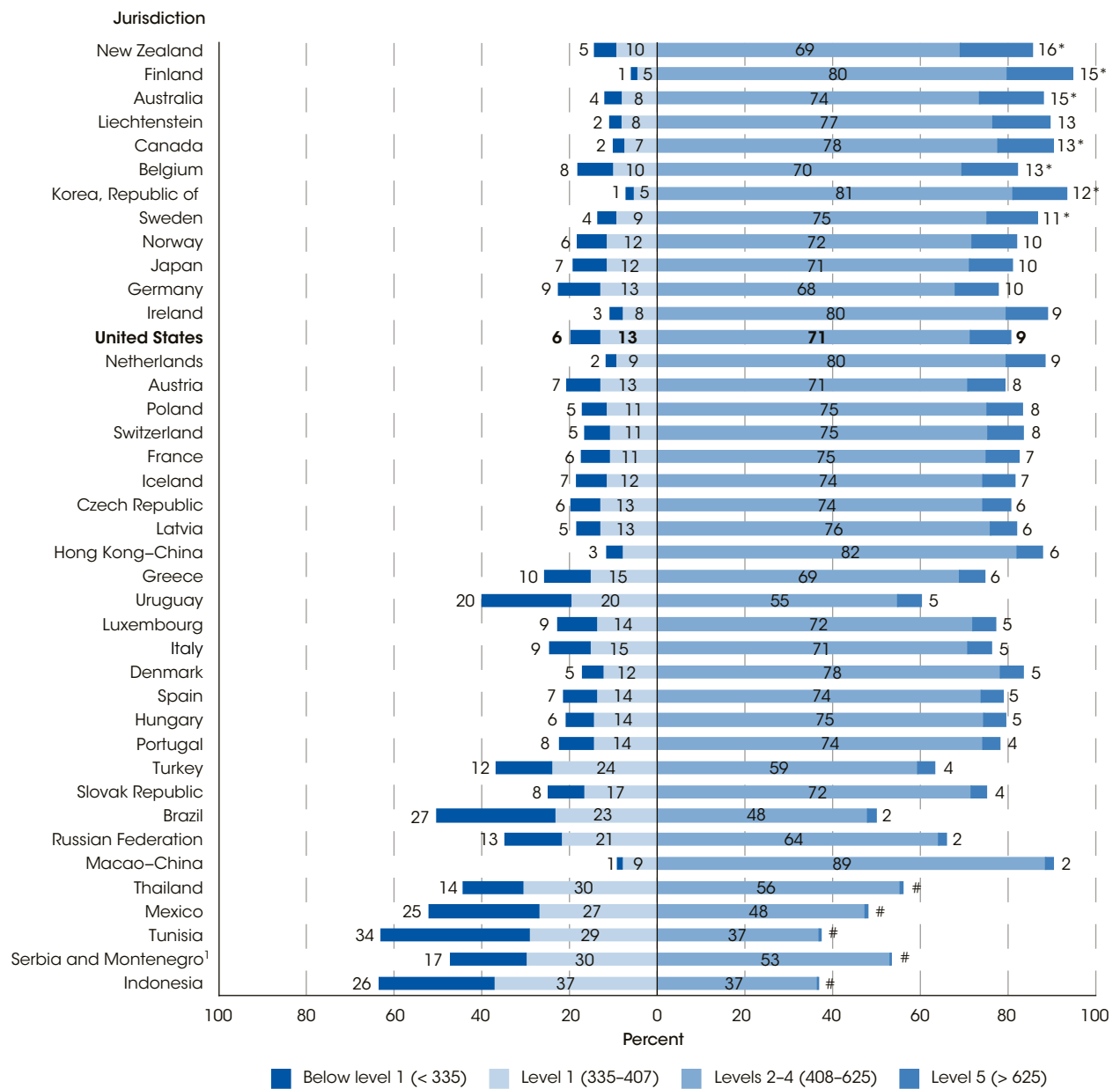
Only one country, Sweden, has consistently outperformed the United States in reading at grade 4 (see supplemental table A-1). However, between 2001 and 2006, when U.S. 4th-graders' scores did not measurably change, the reading scores of their peers improved in 8 countries. The gains made by 5 of these countries (Hong Kong, Hungary, Italy, the Russian Federation, and Singapore) brought the total number of countries that outperformed the United States in 2006 to 10. These gains were evident even among the best-performing students. Between 2001 and 2006, the percentage of U.S. 4th-graders who reached the Advanced benchmark did not measurably change. In contrast, the percentage of students who reached the Advanced benchmark increased in the Russian Federation from 5 to 19 percent, in Hong Kong from 5 to 15 percent, in Singapore from 12 to 19 percent, and in Hungary from 10 to 14 percent (Mullis et al. 2007, exhibit 2.2).

Six countries outperformed U.S. 15-year-old students in reading in both 2000 and 2003 (Australia, Canada, Finland, Ireland, Korea, and New Zealand). In addition, although the scores of U.S. 15-year-old students did not measurably change between 2000 and 2003, U.S. students fell behind their peers in 3 countries where scores improved (Liechtenstein) or did not measurably change (Belgium and Sweden) (figure 4), which brought the total number of countries outperforming the United States to 9 (among countries that participated in both 2000 and 2003) (see supplemental table A-2). Because of the lack of valid 2006 PISA reading scores for the United States, we do not know how many countries outperformed the United States in 2006, nor do we know whether the reading achievement of U.S. 15-year-olds changed between 2003 and 2006 or, once again, remained stable. However, it is clear that the performance of some countries has steadily improved since 2000, both on average and among the top 10 percent of students.<sup>26</sup>

<sup>26</sup> For example, in Korea, the cutpoint score for the top 10 percent of 15-year-olds improved by 26 points between 2000 and 2003 (when there was no measurable change in the cutpoint score of the top 10 percent of U.S. 15-year-olds), and by 55 points between 2000 and 2006 (OECD 2007, table 6.3c).



**Figure 3. Percentage distribution of 15-year-old students on PISA reading literacy scale, by proficiency level and jurisdiction: 2003**



\* $p < .05$ . Percentage at proficiency Level 5 is measurably higher than U.S. percentage.

# Rounds to zero.

<sup>1</sup> The Republics of Serbia and Montenegro were a unified jurisdiction for the 2003 Program of International Student Assessment (PISA).

NOTE: Jurisdictions are ordered by the percentage of students scoring 625 or above on a scale from 0 to 1,000. Students with a score of 407 or lower appear on the left side of the percentage distribution. These students performed at proficiency level 1 or below. At level 1, students can complete only the simplest reading tasks, such as locating a single piece of information, identifying the main theme of a text or making a simple connection with everyday knowledge. The Program for International Student Assessment (PISA) has defined five levels of proficiency based on specific student proficiencies. These specific student proficiencies remain the same across all PISA assessments; however, the score point threshold for students who demonstrate these specific student proficiencies may vary slightly from assessment to assessment. Because OECD proficiency levels are anchored by specific student proficiencies (i.e., by items not scores), the percentage distribution by level can be reported. Apparent differences may not be statistically significant.

SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2003.

## Mathematics

TIMSS and PISA both assess aspects of mathematical skills; however, they differ in terms of whom they assess and what they measure. TIMSS assesses 4th- and 8th-graders' knowledge of specific mathematical topics and cognitive skills that are closely linked to the curricula of the participating countries. PISA assesses 15-year-old students' *mathematics literacy*, which it defines as

An individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgments and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned, and reflective citizen (OECD, 2006, p. 12).

On account of these different aims, the two assessments ask students to perform different tasks. TIMSS asks 4th- and 8th-graders to complete a range of multiple-choice and short constructed response questions that test their knowledge of specific mathematics topics or content domains—*numbers* (manipulating whole numbers and place values; performing addition, subtraction, multiplication, and division; and using fractions and decimals), *geometric shapes and measures*, and *data display* at grade 4; and *numbers, algebra, geometry, and data and chance* at grade 8.<sup>27</sup>

In contrast, PISA does not focus exclusively on outcomes that can be directly linked to curricula, but instead emphasizes real world applications of mathematical knowledge. PISA's content domains are defined in terms of the manner in which mathematical knowledge is likely to be encountered in the larger world: *space and shape, change and relationships, and uncertainty*. Thus, PISA presents students with a variety of situations or problems in which, as young adults, they are likely to encounter numbers and mathematical concepts. These can range from estimating an area or comparing the best buy for a product to interpreting the statistics in a news report or government document. Most questions are multiple-choice, but there are some constructed response questions as well which could ask students to explain a mathematics principle, show their calculations, or explain their reasoning.<sup>28</sup> PISA complements information obtained from studies such as TIMSS because it addresses whether students can apply what they have learned both in and out of school.<sup>29</sup>

It is important to note that PISA's mathematics assessment was revised in 2003 to broaden the content domains covered. (The PISA 2000 assessment did not cover uncertainty.) Because of this change, comparing mathematics literacy scores from PISA 2000 with those of later PISA assessments is inappropriate.

### Mathematics results for 4th- and 8th-graders

The 2007 TIMSS results showed that U.S. students' average mathematics score was 529 for 4th-graders and 508 for 8th-graders (tables 4 and 5). Both scores were above the TIMSS scale average, which is set at 500 for every administration of TIMSS at both grades. The U.S. 4th-grade average score reflects the fact that U.S. 4th-graders performed above the TIMSS scale average in all three mathematics content domains (numbers, geometric shapes and measures, and data display). The U.S. 8th-grade average score reflects the fact that U.S. 8th-graders performed above the TIMSS scale average in two of the four mathematics content domains (numbers and data and chance) in 2007 (Gonzales et al. 2008, table 6). In algebra they did not score measurably different from the TIMSS scale average, and in geometry they scored 20 score points below the TIMSS scale average (Gonzales et al. 2008, table 7).

Comparing the performance of U.S. students with the performance of their peers in other countries, 4th-graders in 8 countries (Hong Kong, Singapore, Chinese Taipei, Japan, Kazakhstan, Russian Federation, England, and Latvia) scored above their U.S. peers, on average (table 4). The top 10 percent of U.S. 4th-graders scored 625 or higher, a cutpoint score below that of the top 10 percent of students in 7 countries (all of which had higher average scores than the United States), while the bottom 10 percent scored 430 or lower, a cutpoint score below that of the bottom 10 percent of students in 6 countries (5 of which had higher average scores than the United States).

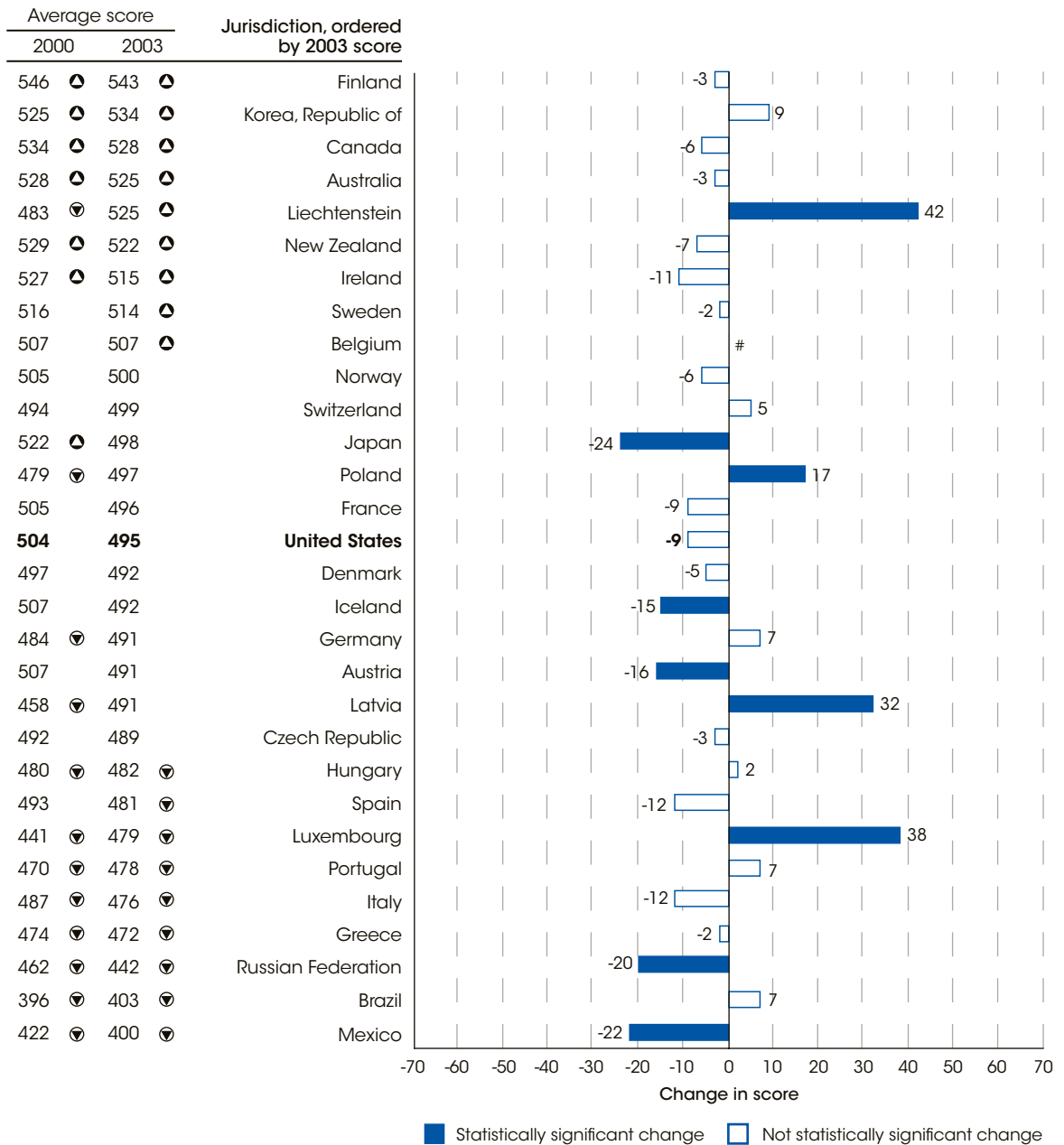
Eighth-graders in 5 countries (Chinese Taipei, Korea, Singapore, Hong Kong, and Japan) scored above their U.S. peers, on average (table 5). The top 10 percent of U.S. 8th-graders scored 607 or higher, a cutpoint score below that of the top 10 percent of students in 6 countries, including the 5 countries with average scores higher than the U.S. average scores. The bottom 10 percent of U.S. 8th-graders scored 408 or lower, a cutpoint score below that of the bottom 10 percent of students in 4 countries (all of which had average scores higher than the United States).

<sup>27</sup> Examples of TIMSS mathematics items can be viewed at <http://timss.bc.edu/timss2007/items.html>.

<sup>28</sup> Examples of PISA mathematics items can be found in OECD 2007, pp. 306–311 (available at <http://www.oecd.org/dataoecd/30/17/39703267.pdf>).

<sup>29</sup> For a more detailed comparison of TIMSS and PISA, see Neidorf et al. 2006.

**Figure 4. Change in average PISA reading literacy scores of 15-year-old students in selected jurisdictions, by jurisdiction: 2000 to 2003**



▲ Average score was higher than the U.S. average score.

▼ Average score was lower than the U.S. average score.

# Rounds to zero.

NOTE: Selected jurisdictions include those that participated in both PISA 2000 and PISA 2003. The Netherlands participated in the Program for International Student Assessment (PISA) in 2000, but its data are omitted from the PISA report because of technical problems with its sample. Average reading literacy scores are reported on a scale from 0 to 1,000. Tests for statistical significance take into account the standard errors for scores in both years and the linking error across assessments. For more information on linking error, see appendix A. Because the size of standard errors can vary, a small difference between the scores of one jurisdiction may be significant while a larger difference between the scores of another jurisdiction may not be significant. Detail may not sum to totals due to rounding.

SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2000 and 2003.

**Table 4. Average TIMSS scores of fourth-grade students in mathematics and cutpoint scores for bottom and top 10 percent of students in each jurisdiction, by jurisdiction: 2007**

Jurisdiction	Average score	Cutpoint score	
		Bottom 10 percent	Top 10 percent
<b>All jurisdictions</b>	<b>500</b> ♣	<b>366</b> ♣	<b>576</b> ♣
Hong Kong, SAR <sup>1</sup>	607 ▲	520 ▲	691 ▲
Singapore	599 ▲	487 ▲	702 ▲
Chinese Taipei	576 ▲	488 ▲	663 ▲
Japan	568 ▲	471 ▲	663 ▲
Kazakhstan <sup>2</sup>	549 ▲	435	653 ▲
Russian Federation	544 ▲	436	647 ▲
England	541 ▲	429	647 ▲
Latvia <sup>2</sup>	537 ▲	444 ▲	628
Netherlands <sup>3</sup>	535	454 ▲	612 ♣
Lithuania <sup>2</sup>	530	430	624
<b>United States<sup>4,5</sup></b>	<b>529</b>	<b>430</b>	<b>625</b>
Germany	525	440	607 ♣
Denmark <sup>4</sup>	523	431	611 ♣
Australia	516 ♣	408 ♣	620
Hungary	510 ♣	389 ♣	620
Italy	507 ♣	406 ♣	601 ♣
Austria	505 ♣	416 ♣	590 ♣
Sweden	503 ♣	417 ♣	586 ♣
Slovenia	502 ♣	408 ♣	589 ♣
Armenia	500 ♣	385 ♣	617
Slovak Republic	496 ♣	389 ♣	597 ♣
Scotland <sup>4</sup>	494 ♣	389 ♣	592 ♣
New Zealand	492 ♣	377 ♣	598 ♣
Czech Republic	486 ♣	392 ♣	576 ♣
Norway	473 ♣	372 ♣	566 ♣
Ukraine	469 ♣	356 ♣	573 ♣
Georgia <sup>2</sup>	438 ♣	322 ♣	549 ♣
Iran, Islamic Republic of	402 ♣	290 ♣	508 ♣
Algeria	378 ♣	261 ♣	493 ♣
Colombia	355 ♣	238 ♣	470 ♣
Morocco	341 ♣	223 ♣	466 ♣
El Salvador	330 ♣	212 ♣	448 ♣
Tunisia	327 ♣	178 ♣	469 ♣
Kuwait <sup>6</sup>	316 ♣	184 ♣	443 ♣
Qatar	296 ♣	179 ♣	413 ♣
Yemen	224 ♣	81 ♣	371 ♣

▲ Score is higher than U.S. score.

♣ Score is lower than U.S. score.

<sup>1</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>2</sup> National Target Population did not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS).

<sup>3</sup> Nearly satisfied guidelines for sample participation rates only after substitute schools were included.

<sup>4</sup> Met guidelines for sample participation rates only after substitute schools were included.

<sup>5</sup> National Defined Population covered 90 percent to 95 percent of National Target Population.

<sup>6</sup> Kuwait tested the same cohort of students as other countries, but later in the year, at the beginning of the next school year.

NOTE: Jurisdictions are ordered on the basis of average scores, from highest to lowest. Mathematics scores are reported on a scale from 0 to 1,000. A cutpoint score is the threshold score for an established level of performance. The cutpoint scores for students in the top 10 percent is the 90th percentile score within the jurisdiction. The cutpoint score for students in the bottom 10 percent is the 10th percentile score within the jurisdiction. 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.

SOURCE: 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), tables 3 and 9. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

**Table 5. Average TIMSS scores of eighth-grade students in mathematics and cutpoint scores for bottom and top 10 percent of students in each jurisdiction, by jurisdiction: 2007**

Jurisdiction	Average score	Cutpoint score	
		Bottom 10 percent	Top 10 percent
<b>All jurisdictions</b>	<b>500</b> ♣	<b>339</b> ♣	<b>559</b> ♣
Chinese Taipei	598 ▲	448 ▲	721 ▲
Korea, Republic of	597 ▲	475 ▲	711 ▲
Singapore	593 ▲	463 ▲	706 ▲
Hong Kong SAR <sup>1,2</sup>	572 ▲	438	681 ▲
Japan	570 ▲	460 ▲	677 ▲
Hungary	517	405	624 ▲
England <sup>1</sup>	513	400	618
Russian Federation	512	402	617
<b>United States<sup>1,3</sup></b>	<b>508</b>	<b>408</b>	<b>607</b>
Lithuania <sup>4</sup>	506	402	609
Czech Republic	504	408	599
Slovenia	501 ♣	409	594 ♣
Armenia	499 ♣	390 ♣	601
Australia	496 ♣	394	600
Sweden	491 ♣	399	582 ♣
Malta	488 ♣	359 ♣	597 ♣
Scotland <sup>1</sup>	487 ♣	381 ♣	590 ♣
Serbia <sup>3,4</sup>	486 ♣	368 ♣	597
Italy	480 ♣	381 ♣	574 ♣
Malaysia	474 ♣	372 ♣	578 ♣
Norway	469 ♣	382 ♣	552 ♣
Cyprus	465 ♣	347 ♣	575 ♣
Bulgaria	464 ♣	324 ♣	586 ♣
Israel <sup>5</sup>	463 ♣	328 ♣	584 ♣
Ukraine	462 ♣	346 ♣	572 ♣
Romania	461 ♣	328 ♣	587 ♣
Bosnia and Herzegovina	456 ♣	352 ♣	552 ♣
Lebanon	449 ♣	354 ♣	549 ♣
Thailand	441 ♣	327 ♣	562 ♣
Turkey	432 ♣	297 ♣	581 ♣
Jordan	427 ♣	290 ♣	556 ♣
Tunisia	420 ♣	336 ♣	508 ♣
Georgia <sup>4</sup>	410 ♣	280 ♣	532 ♣
Iran, Islamic Republic of	403 ♣	295 ♣	516 ♣
Bahrain	398 ♣	289 ♣	505 ♣
Indonesia <sup>4</sup>	397 ♣	286 ♣	509 ♣
Syrian Arab Republic	395 ♣	290 ♣	502 ♣
Egypt	391 ♣	258 ♣	521 ♣
Algeria	387 ♣	311 ♣	465 ♣
Colombia	380 ♣	281 ♣	477 ♣
Oman	372 ♣	245 ♣	492 ♣
Palestinian National Authority	367 ♣	233 ♣	498 ♣
Botswana	364 ♣	264 ♣	460 ♣
Kuwait <sup>6</sup>	354 ♣	252 ♣	455 ♣
El Salvador	340 ♣	248 ♣	433 ♣
Saudi Arabia	329 ♣	231 ♣	429 ♣
Ghana	309 ♣	192 ♣	428 ♣
Qatar	307 ♣	186 ♣	427 ♣

▲ Score is higher than U.S. score.

♣ Score is lower than U.S. score.

<sup>1</sup> Met guidelines for sample participation rates only after substitute schools were included.

<sup>2</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>3</sup> National Defined Population covered 90 percent to 95 percent of National Target Population.

<sup>4</sup> National Target Population did not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS).

<sup>5</sup> National Defined Population covered less than 90 percent of National Target Population (but at least 77 percent).

<sup>6</sup> Kuwait tested the same cohort of students as other countries, but later in 2007, at the beginning of the next school year.

NOTE: Jurisdictions are ordered on the basis of average scores, from highest to lowest. Mathematics scores are reported on a scale from 0 to 1,000. A cutpoint score is the threshold score for an established level of performance. The cutpoint scores for students in the top 10 percent is the 90th percentile score within the jurisdiction. The cutpoint score for students in the bottom 10 percent is the 10th percentile score within the jurisdiction. 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.

SOURCE: 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), tables 3 and 9. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC.

TIMSS has developed four international benchmarks to help analyze the range of students' performance in mathematics within each participating country.<sup>30</sup> The Advanced benchmark is set at 625 score points for both grades.<sup>31</sup>

Fourth-graders reaching the Advanced benchmark demonstrate a developing understanding of fractions and decimals and the relationship between them. They can select appropriate information to solve multi-step word problems involving proportions. They can formulate or select a rule for a relationship. They show understanding of area and can use measurement concepts to solve a variety of problems. They show some understanding of rotation. They can organize, interpret, and represent data to solve problems.

Eighth-graders reaching the Advanced benchmark can organize information, make generalizations, solve non-routine problems, and draw and justify conclusions from data. They can compute percentage change and apply their knowledge of numeric and algebraic concepts and relationships to solve problems. They can solve simultaneous linear equations and model simple situations algebraically. They can apply their knowledge of measurement and geometry in complex problem situations. They can interpret data from a variety of tables and graphs, including interpolation and extrapolation.

In 2007, ten percent of U.S. 4th-graders and 6 percent of U.S. 8th-graders reached the Advanced benchmark (figures 5 and 6). In comparison, 7 participating countries had a higher percentage of 4th-graders who reached this benchmark (ranging from 41 to 16 percent): Singapore, Hong Kong, Chinese Taipei, Japan, Kazakhstan, England, and the Russian Federation (the same 7 countries with higher cutpoints for their top 10 percent of students). A slightly different set of 7 participating countries had a measurably higher percentage of 8th-graders who reached this benchmark (ranging from 45 to 8 percent): Chinese Taipei, Korea, Singapore, Hong Kong, Japan, Hungary, and the Russian Federation (6 of these 7 countries had higher cutpoints for their top 10 percent of students).

### *Change over time*

Among the 16 countries that participated in both the first TIMSS in 1995 and the most recent TIMSS in 2007, at grade 4, the average mathematics score increased in 8 countries, including in the United States, and decreased

in 4 countries (figure 7). Among the 20 countries that participated in both the 1995 and 2007 TIMSS at grade 8, the average mathematics score increased in 6 countries, including in the United States, and decreased in 10 countries (figure 8).

Between 1995 and 2007 the average score of U.S. 4th-graders increased 11 score points (from 518 to 529). In 4 countries, the average score of 4th-graders increased more than in the United States during this time: England, Hong Kong, Slovenia, and Latvia. Increases in England (57 points) and Latvia (38 points) moved their 4th-graders from scoring below their U.S. peers in 1995 to scoring higher than their U.S. peers in 2007. Increases in Slovenia (40 points) and Hong Kong (50 points) did not change their standing relative to the United States.

Scores decreased during this time for 4th-graders in Hungary (12 points), The Netherlands (14 points), Austria (25 points), and the Czech Republic (54 points). As a result, the performance of U.S. 4th-graders showed improvement relative to their peers in these countries.

At grade 8, the U.S. average score increased 16 score points (from 492 to 508) between 1995 and 2007. In 2 countries, the average score of 8th-graders increased more than in the United States during this time: Colombia (47 points) and Lithuania (34 points). Neither of these countries outperformed the United States in 2007.

Scores decreased during this time for 8th-graders in 10 countries, with the decreases ranging from 10 points in Hungary to 63 points in Bulgaria. The decreases in Australia (13 points), Sweden (48 points), and Bulgaria (63 points) were large enough that their 8th-graders' average scores in 2007 were below those of their U.S. peers, whereas in 1995 their students outperformed their U.S. peers.

The next TIMSS assessment will be administered in 2011. More detailed results for TIMSS 2007 can be found in Gonzales et al. (2008; available at <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001>) and Mullis et al. (2008a; available at <http://timss.bc.edu/TIMSS2007/mathreport.html>). For more information on TIMSS, see <http://nces.ed.gov/timss/>.

### **Mathematics results for 15-year-olds**

In PISA 2006, U.S. 15-year-old students' average mathematics literacy score of 474 was lower than the OECD average of 498 (table 6), and placed U.S. 15-year-olds in the bottom quarter of participating OECD nations. Fifteen-year-old students in 23 of the 29 other participating OECD-member countries outperformed their U.S. peers (as did 15-year-olds in 8 of the 27 non-OECD countries that participated) in terms of average scores.

<sup>30</sup> See figures 5 and 6 for the cut scores established for all the international benchmarks. For details about the international benchmarks, see Mullis et al. (2008a), chapter 2.

<sup>31</sup> The IEA set international benchmarks for TIMSS based on an analysis of score points. The score points for each benchmark remain the same across assessments; however, the configuration of items that define what students reaching a benchmark can do may vary slightly from one assessment to the next. For more details, see appendix A.

A comparable pattern is evident when looking at the results of U.S. 15-year-olds in the top 10 percent of performance. The top 10 percent of 15-year-olds in the same 23 OECD countries and in 6 of the 8 non-OECD countries scored higher than the top 10 percent of U.S. 15-year-olds, who scored 593 or higher. In comparison, students in the top 10 percent in Chinese Taipei scored 677 or higher; in Hong Kong, 665 or higher; and in Korea, 664 or higher. Comparing the performances of the bottom 10 percent of students in each country, 18 OECD countries and 8 non-OECD countries scored higher than the United States, where the bottom 10 percent of 15-year-olds scored 358 or lower.

PISA has developed six levels of student achievement to help analyze the range of student performance in mathematics within each participating country.<sup>32</sup> The highest level of proficiency in mathematics (above 669 score points for PISA 2006) identifies students who are capable of advanced mathematical thinking and reasoning and who demonstrate a mastery of symbolic and formal mathematical operations and relationships. They can conceptualize, generalize, and use information based on their investigations and modeling of complex problem situations. They can link different information sources and representations and can flexibly translate among them. They can develop new approaches and strategies for attacking novel situations.

In 2006, one percent of U.S. 15-year-olds performed at this level (figure 9). Twenty-seven countries had a higher percentage of 15-year-olds who performed at this level; Chinese Taipei had the largest percentage of students at this level (12 percent). The percentage of students at this level in Korea, Hong Kong, Switzerland, Belgium, Finland, the Czech Republic, Liechtenstein, New Zealand, The Netherlands, and Japan ranged from 9 to 5 percent.

### *Change over time*

Because of the addition of a content domain to the PISA mathematics assessment in 2003, only the results of PISA 2003 and PISA 2006 are compared here. Among the 40 countries that participated in both PISA 2003 and PISA 2006, average mathematics literacy scores increased in 4 countries and decreased in 4 countries (figure 10).<sup>33</sup> There was no measurable change in the average U.S. mathematics literacy score between 2003 and 2006, in its relationship to the OECD average, or in its relationship to the countries whose scores increased or decreased.

<sup>32</sup> See figure 9 for the cut scores for all six levels of proficiency. For details about all six levels, see OECD 2007, pp. 312–15.

<sup>33</sup> Note that statistical comparisons between PISA 2003 scores and PISA 2006 scores properly need to account for linking error. See appendix A for details.

However, three countries whose scores were not measurably different than the United States in 2003 outperformed the United States in 2006: Latvia, Hungary, and Poland.

More detailed information on the PISA 2006 results can be found in Baldi et al. (2007; available at <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2008016>) and OECD (2007; available at <http://www.oecd.org/dataoecd/30/17/39703267.pdf>). For more information on PISA, see <http://nces.ed.gov/surveys/pisa/>.

## Synthesis of mathematics results

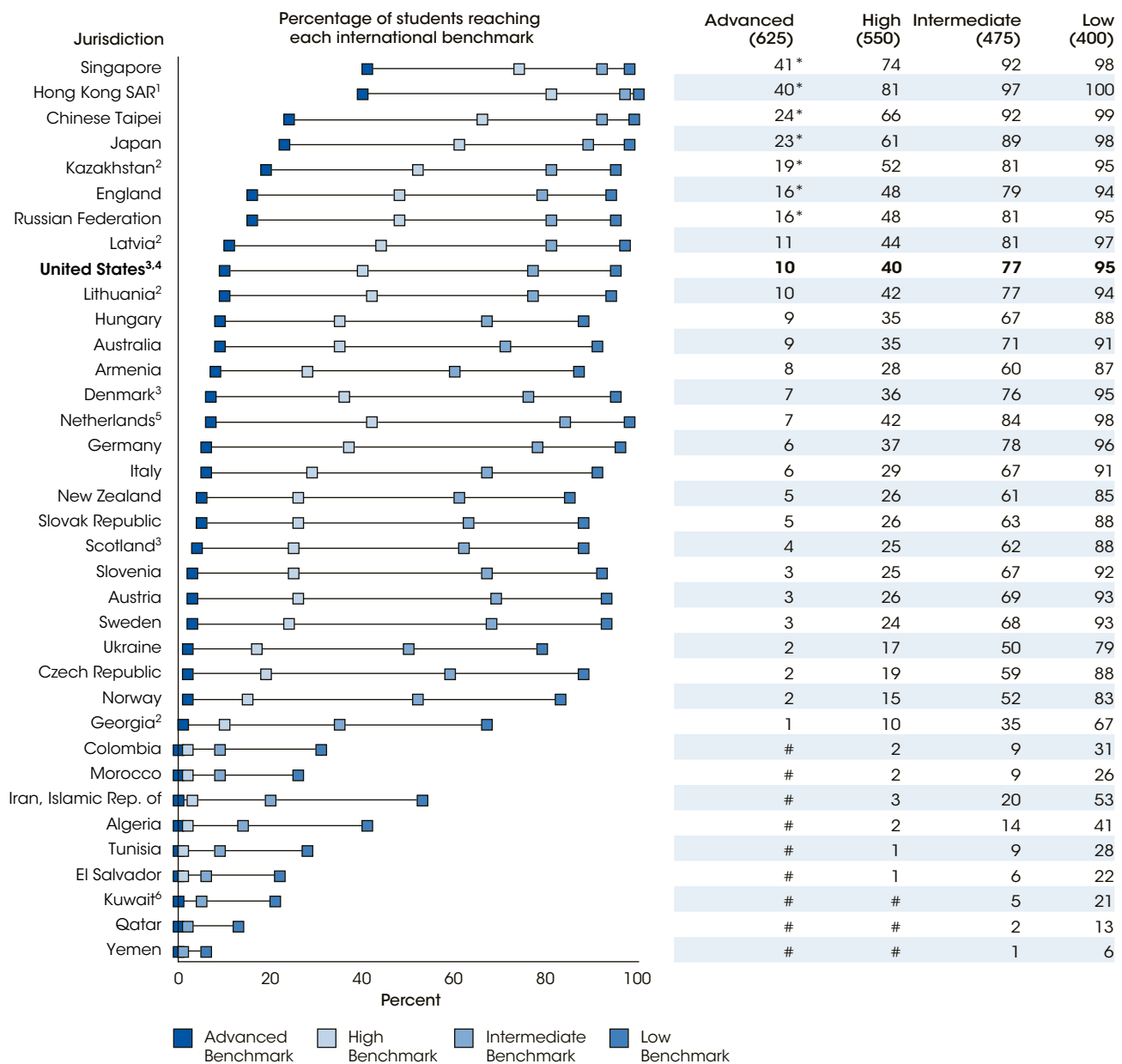
While U.S. 4th-graders' average scores increased between 1995 and 2007, 4th-graders in Hong Kong, Japan, and Singapore consistently outperformed their U.S. peers in mathematics, as did 4th-graders in the Russian Federation and Chinese Taipei, both of whom only participated in TIMSS 2003 and 2007 (see supplemental table A-3). U.S. 4th-graders lost ground relative to their peers in England and Latvia who improved at a faster rate between 1995 and 2007, but 4th-graders gained ground relative to their peers in Hungary, The Netherlands, Austria, and the Czech Republic (figure 7).

The same three Asian countries that consistently outperformed the United States in mathematics at grade 4 (Hong Kong, Japan, and Singapore) also consistently outperformed the United States in mathematics at grade 8 between 1995 and 2007 (see supplemental table A-4). In addition, Korea and Chinese Taipei also outperformed the United States in mathematics at grade 8 each time they participated in TIMSS. No other participating countries, however, outperformed the United States in 2007, including 6 of the countries that outperformed the United States at grade 8 in 1995: Australia, Bulgaria, the Czech Republic, Hungary, the Russian Federation, and Sweden. Thus while U.S. 8th-graders have not caught up with their Asian peers in terms of their average mathematics score, they have improved their standing relative to their peers in these 6 countries.

Still, when the ability of 15-year-old students to apply mathematics knowledge and skills to real-world tasks was assessed, 23 countries outperformed the United States in both 2003 and 2006 in terms of average scores (see supplemental table A-5). In addition, 15-year-olds in Hungary, Latvia, and Poland—three countries that did not outperform the United States in 2003—outperformed their U.S. peers in 2006.<sup>34</sup>

<sup>34</sup> The 2003 average scores of Latvia, Hungary, and Poland were not statistically different than the 2003 U.S. average score; and there was no measurable increase in the scores of the United States, Latvia, Hungary, and Poland in 2006 (i.e., they were not statistically different than their own scores in 2003). Nevertheless, the latter three countries' scores were statistically different than the United States in 2006 on account of differences in the amount of their respective standard errors in 2003 compared with 2006.

**Figure 5. Percentage of fourth-grade students reaching the TIMSS international benchmarks in mathematics, by jurisdiction: 2007**



\* $p < .05$ . Percentage reaching advanced benchmark is measurably higher than U.S. percentage.

# Rounds to zero.

<sup>1</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>2</sup> National Target Population does not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS).

<sup>3</sup> Met guidelines for sample participation rates only after substitute schools were included.

<sup>4</sup> National Defined Population covers 90 percent to 95 percent of National Target Population.

<sup>5</sup> Nearly satisfied guidelines for sample participation rates only after substitute schools were included.

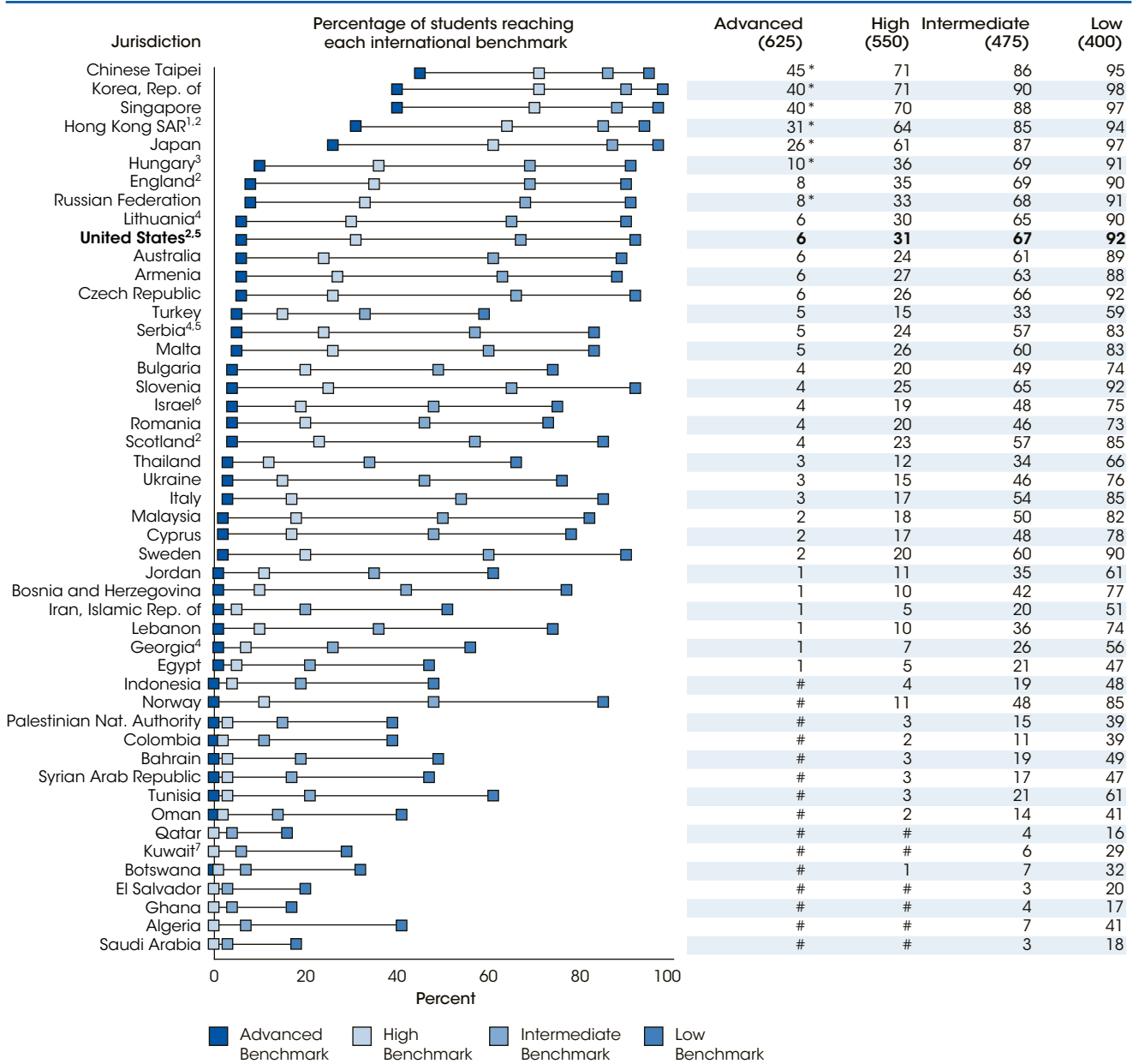
<sup>6</sup> Kuwait tested the same cohort of students as other countries, but later in 2007, at the beginning of the next school year.

NOTE: Jurisdictions are ordered by the percentage of students who reached the advanced international benchmark on a scale of 0 to 1,000. The International Association for the Evaluation of Educational Achievement (IEA) set international benchmarks for TIMSS based on an analysis of score points. Students who did not reach benchmarks are not shown nor are students whose placement relative to the benchmarks was uncertain. Because IEA benchmarks are anchored by scores, it is possible to calculate the percentage of students who reach each benchmark's score threshold with accuracy but not the percentage distribution of students by benchmarks. Apparent differences may not be statistically significant.

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



**Figure 6. Percentage of eighth-grade students reaching the TIMSS international benchmarks in mathematics, by jurisdiction: 2007**



\*  $p < .05$ . Percentage reaching advanced benchmark is measurably higher than U.S. percentage.

# Rounds to zero.

<sup>1</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>2</sup> Met guidelines for sample participation rates only after substitute schools were included.

<sup>3</sup> In Hungary, the percentage of eighth graders who had scores of 625 or higher was 9.825, which is reported as 10 percent when rounded. However, the top ten percent of eighth-grade performers actually had scores of 624 or higher.

<sup>4</sup> National Target Population does not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS).

<sup>5</sup> National Defined Population covers 90 percent to 95 percent of National Target Population.

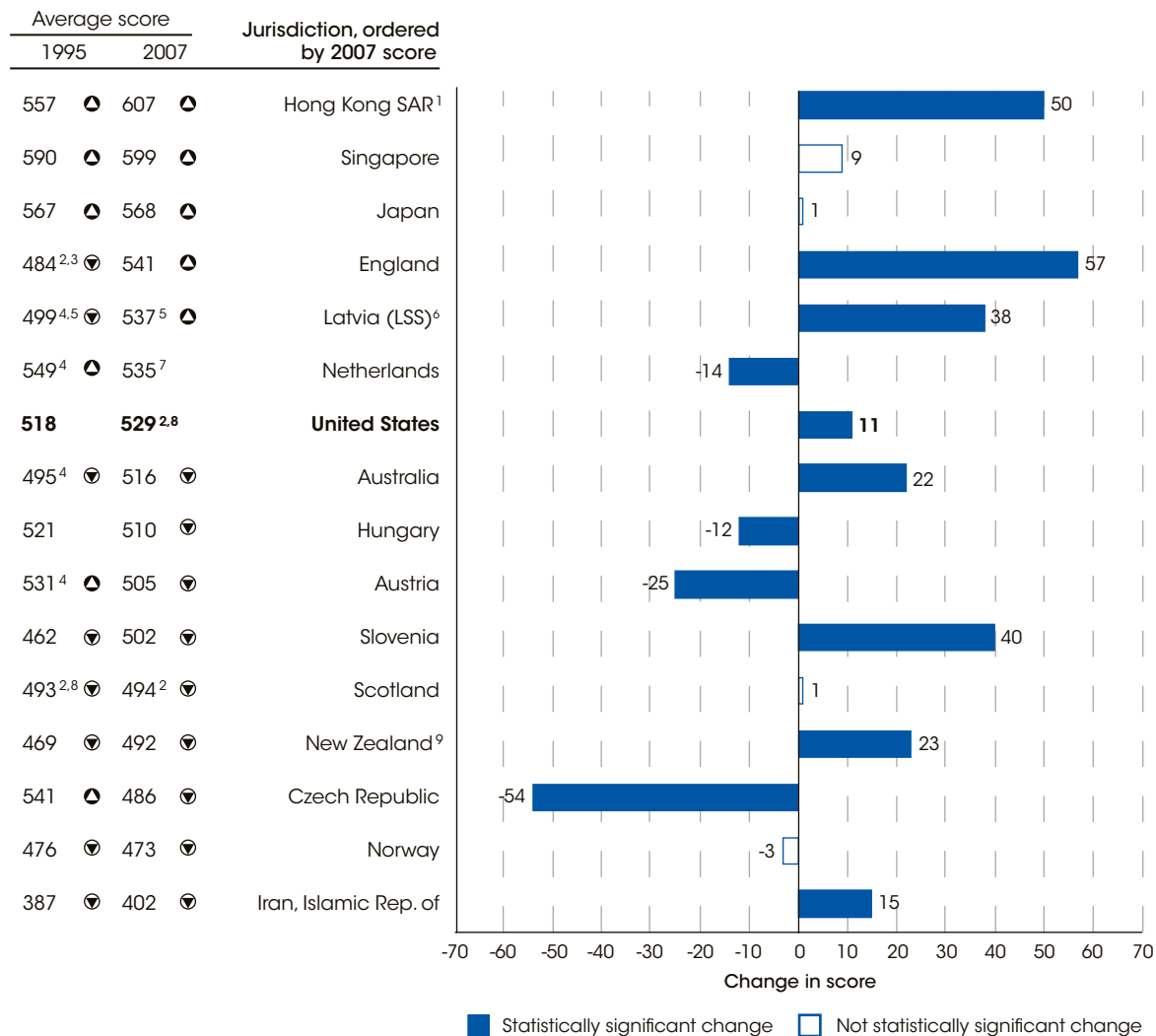
<sup>6</sup> National Defined Population covers less than 90 percent of National Target Population (but at least 77 percent).

<sup>7</sup> Kuwait tested the same cohort of students as other countries, but later in 2007, at the beginning of the next school year.

NOTE: Jurisdictions are ordered by the percentage of students who reached the advanced international benchmark on a scale of 0 to 1,000. The International Association for the Evaluation of Educational Achievement (IEA) set international benchmarks for TIMSS based on an analysis of score points. Students who did not reach benchmarks are not shown nor are students whose placement relative to the benchmarks was uncertain. Because IEA benchmarks are anchored by scores, it is possible to calculate the percentage of students who reach each benchmark's score threshold with accuracy but not the percentage distribution of students by benchmarks. Apparent differences may not be statistically significant.

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

**Figure 7. Change in average TIMSS mathematics scores of fourth-grade students in selected jurisdictions, by jurisdiction: 1995 to 2007**



▲ Average score was higher than the U.S. average score.

▼ Average score was lower than the U.S. average score.

<sup>1</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>2</sup> Met guidelines for sample participation rates only after substitute schools were included.

<sup>3</sup> National Defined Population covers less than 90 percent of National Target Population (but at least 77 percent).

<sup>4</sup> Did not satisfy sampling participation rates.

<sup>5</sup> National Target Population did not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS).

<sup>6</sup> Only Latvian-speaking schools (LSS) are included.

<sup>7</sup> Nearly satisfied guidelines for sample participation rates only after substitute schools were included.

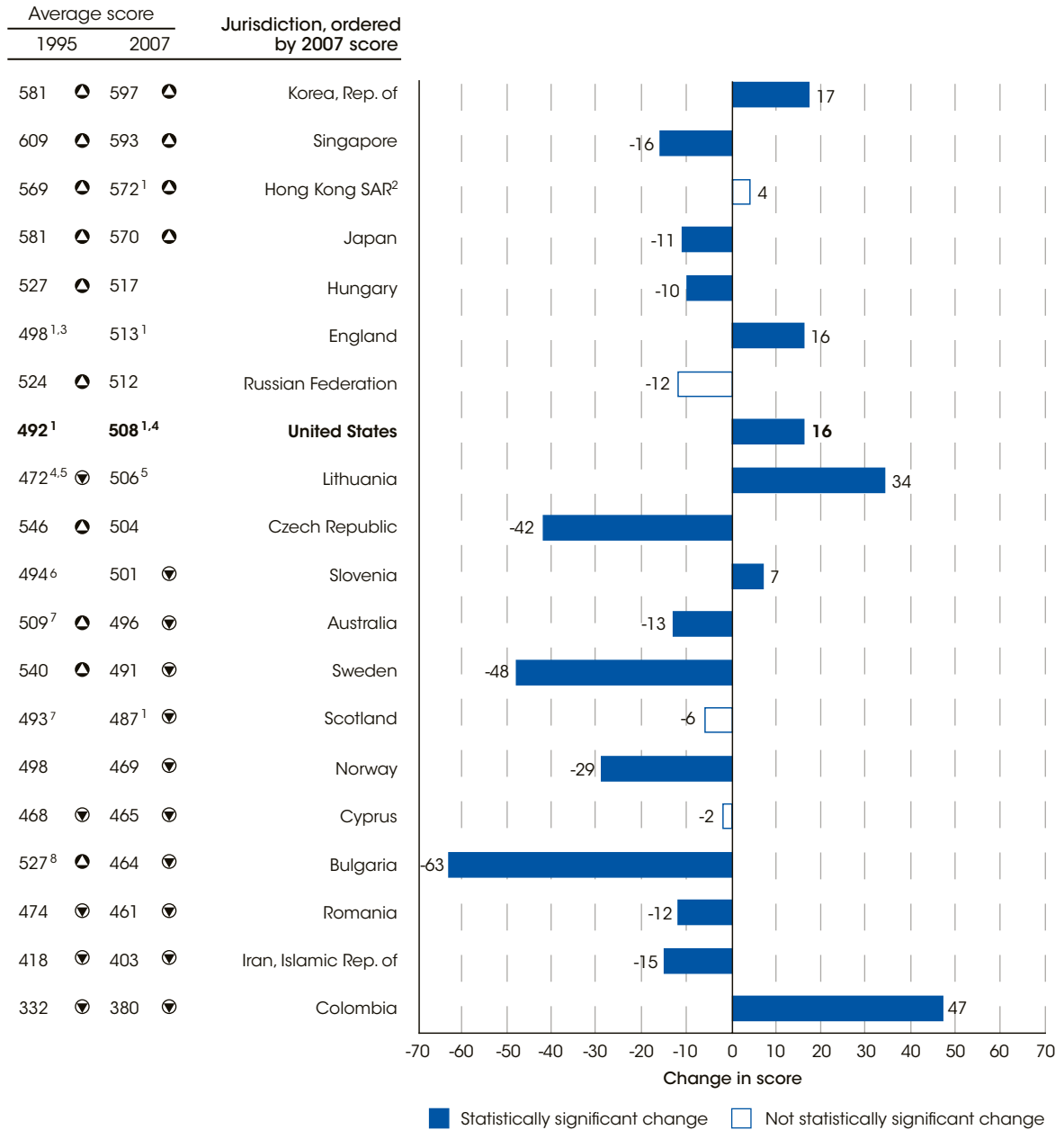
<sup>8</sup> National Defined Population covers 90 percent to 95 percent of National Target Population.

<sup>9</sup> Estimates for New Zealand have been computed for students taught in English only, which represents 98 to 99 percent of the student population.

NOTE: Selected jurisdictions include those that participated in both TIMSS 1995 and TIMSS 2007. Average mathematics scores are reported on a scale from 0 to 1,000. Tests for statistical significance take into account the standard errors for scores in both years. Because the size of standard errors can vary, a small difference between the scores of one jurisdiction may be significant while a larger difference between the scores of another jurisdiction may not be significant. Detail may not sum to totals due to rounding.

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

**Figure 8. Change in average TIMSS mathematics scores of eighth-grade students in selected jurisdictions, by jurisdiction: 1995 to 2007**



▲ Average score was higher than the U.S. average score.

▼ Average score was lower than the U.S. average score.

<sup>1</sup> Met guidelines for sample participation rates only after substitute schools were included.

<sup>2</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>3</sup> National Defined Population covers less than 90 percent of National Target Population (but at least 77 percent).

<sup>4</sup> National Defined Population covers 90 percent to 95 percent of National Target Population.

<sup>5</sup> National Target Population did not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS).

<sup>6</sup> Sampling issues identified by TIMSS. See TIMSS 1995 report for details.

<sup>7</sup> Nearly satisfied guidelines for sample participation rates only after substitute schools were included.

<sup>8</sup> Did not satisfy guidelines for sample participation rates.

NOTE: Selected jurisdictions include those that participated in both TIMSS 1995 and TIMSS 2007. Average mathematics scores are reported on a scale from 0 to 1,000. Tests for statistical significance take into account the standard errors for scores in both years. Because the size of standard errors can vary, a small difference between the scores of one jurisdiction may be significant while a larger difference between the scores of another jurisdiction may not be significant. Detail may not sum to totals due to rounding.

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

**Table 6. Average PISA scores of 15-year-old students on mathematics literacy scale and cutpoint scores for bottom and top 10 percent of students in each jurisdiction, by jurisdiction: 2006**

Jurisdiction	Average score	Cutpoint score	
		Bottom 10 percent	Top 10 percent
<b>All jurisdictions</b>	<b>498 ▲</b>	<b>379 ▲</b>	<b>615 ▲</b>
<b>OECD</b>			
Finland	548 ▲	444 ▲	652 ▲
Korea, Republic of	547 ▲	426 ▲	664 ▲
Netherlands	531 ▲	412 ▲	645 ▲
Switzerland	530 ▲	401 ▲	652 ▲
Canada	527 ▲	416 ▲	635 ▲
Japan	523 ▲	404 ▲	638 ▲
New Zealand	522 ▲	401 ▲	643 ▲
Belgium	520 ▲	381 ▲	650 ▲
Australia	520 ▲	406 ▲	633 ▲
Denmark	513 ▲	404 ▲	621 ▲
Czech Republic	510 ▲	376 ▲	644 ▲
Iceland	506 ▲	391 ▲	618 ▲
Austria	505 ▲	373	630 ▲
Germany	504 ▲	375	632 ▲
Sweden	502 ▲	387 ▲	617 ▲
Ireland	501 ▲	396 ▲	608 ▲
France	496 ▲	369	617 ▲
United Kingdom	495 ▲	381 ▲	612 ▲
Poland	495 ▲	384 ▲	610 ▲
Slovak Republic	492 ▲	370	611 ▲
Hungary	491 ▲	377 ▲	609 ▲
Luxembourg	490 ▲	368	610 ▲
Norway	490 ▲	373 ▲	609 ▲
Spain	480	366	593
<b>United States</b>	<b>474</b>	<b>358</b>	<b>593</b>
Portugal	466	348	583
Italy	462 ▼	341 ▼	584
Greece	459 ▼	341 ▼	575 ▼
Turkey	424 ▼	316 ▼	550 ▼
Mexico	406 ▼	299 ▼	514 ▼
<b>Non-OECD</b>			
Chinese Taipei	549 ▲	409 ▲	677 ▲
Hong Kong–China	547 ▲	423 ▲	665 ▲
Macao–China	525 ▲	416 ▲	632 ▲
Liechtenstein	525 ▲	402 ▲	643 ▲
Estonia	515 ▲	411 ▲	618 ▲
Slovenia	504 ▲	390 ▲	623 ▲
Lithuania	486 ▲	369	602
Latvia	486 ▲	378 ▲	590
Azerbaijan	476	419 ▲	536 ▼
Russian Federation	476	363	592
Croatia	467	361	576 ▼
Israel	442 ▼	304 ▼	581
Serbia, Republic of	435 ▼	318 ▼	553 ▼
Uruguay	427 ▼	296 ▼	551 ▼
Thailand	417 ▼	317 ▼	524 ▼
Romania	415 ▼	307 ▼	523 ▼
Bulgaria	413 ▼	287 ▼	543 ▼
Chile	411 ▼	302 ▼	527 ▼
Montenegro, Republic of	399 ▼	291 ▼	510 ▼
Indonesia	391 ▼	293 ▼	498 ▼
Jordan	384 ▼	279 ▼	489 ▼
Argentina	381 ▼	249 ▼	508 ▼

See notes at end of table.

**Table 6. Average PISA scores of 15-year-old students on mathematics literacy scale and cutpoint scores for bottom and top 10 percent of students in each jurisdiction, by jurisdiction: 2006—Continued**

Jurisdiction	Average score	Cutpoint score	
		Bottom 10 percent	Top 10 percent
<b>All jurisdictions</b>	<b>498</b> ▲	<b>379</b> ▲	<b>615</b> ▲
<b>Non-OECD—Continued</b>			
Colombia	370 ▼	258 ▼	482 ▼
Brazil	370 ▼	255 ▼	487 ▼
Tunisia	365 ▼	250 ▼	488 ▼
Qatar	318 ▼	212 ▼	438 ▼
Kyrgyz Republic	311 ▼	204 ▼	423 ▼

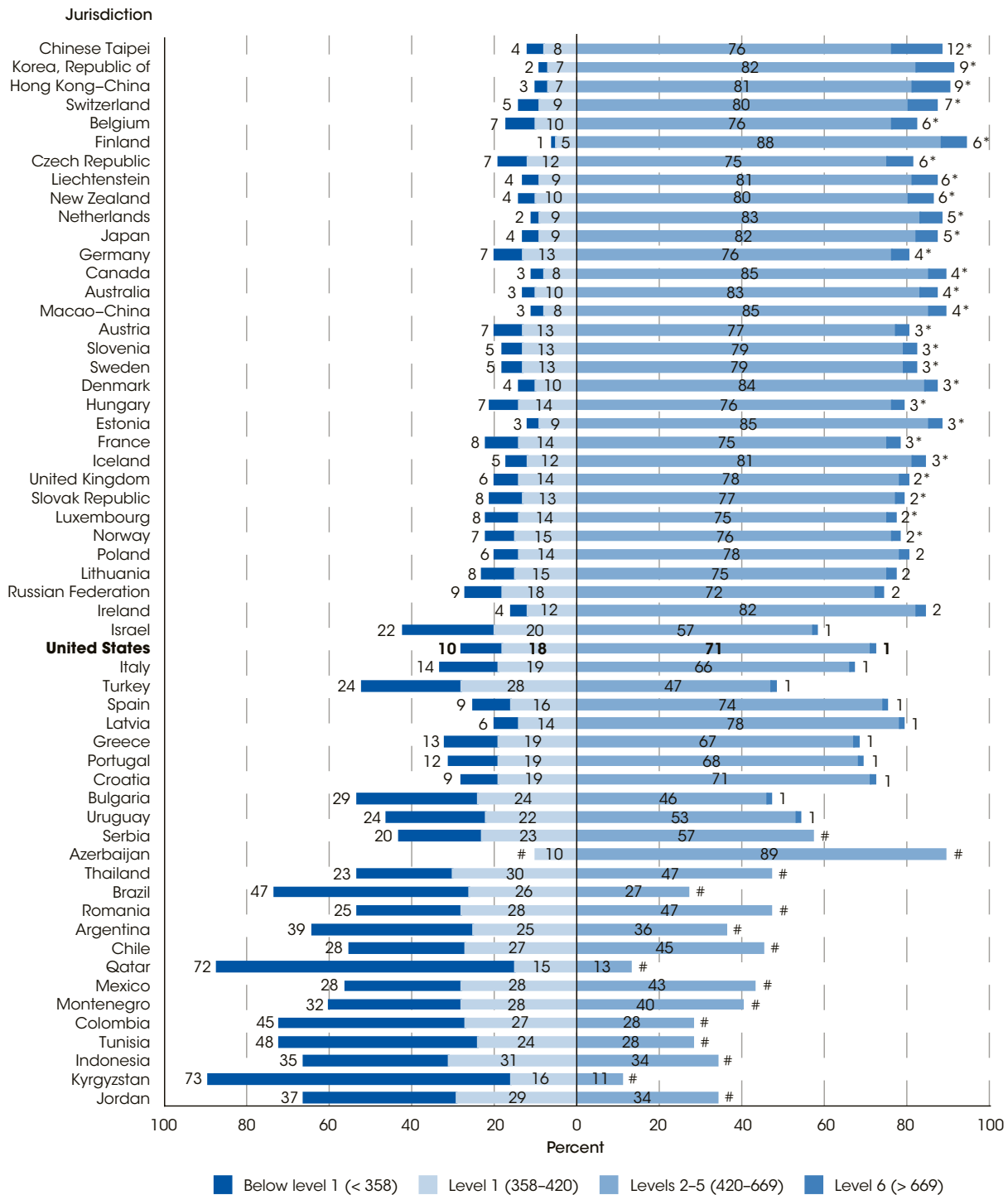
▲ Score is higher than U.S. score.

▼ Score is lower than U.S. score.

NOTE: Selected jurisdictions included all those that participated in the Program for International Student Assessment (PISA) in 2006. Because PISA is principally an Organization for Economic Cooperation and Development (OECD) study, the results for non-OECD jurisdictions are displayed separately from those of the OECD countries. Jurisdictions are ordered on the basis of average scores, from highest to lowest within the OECD countries and the non-OECD jurisdictions. Mathematics literacy scores are reported on a scale from 0 to 1,000. A cutpoint score is the threshold score for an established level of performance. The cutpoint scores for students in the top 10 percent is the 90th percentile score within the jurisdiction. The cutpoint score for students in the bottom 10 percent is the 10th percentile score within the jurisdiction. 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.

SOURCE: Organization for Economic Cooperation and Development (OECD). (2007). *PISA 2006: Science Competencies for Tomorrow's World, Volume 2: Data*, table 6.2c. Paris: Author.

**Figure 9. Percentage distribution of 15-year-old students on PISA mathematics literacy scale, by proficiency level and jurisdiction: 2006**



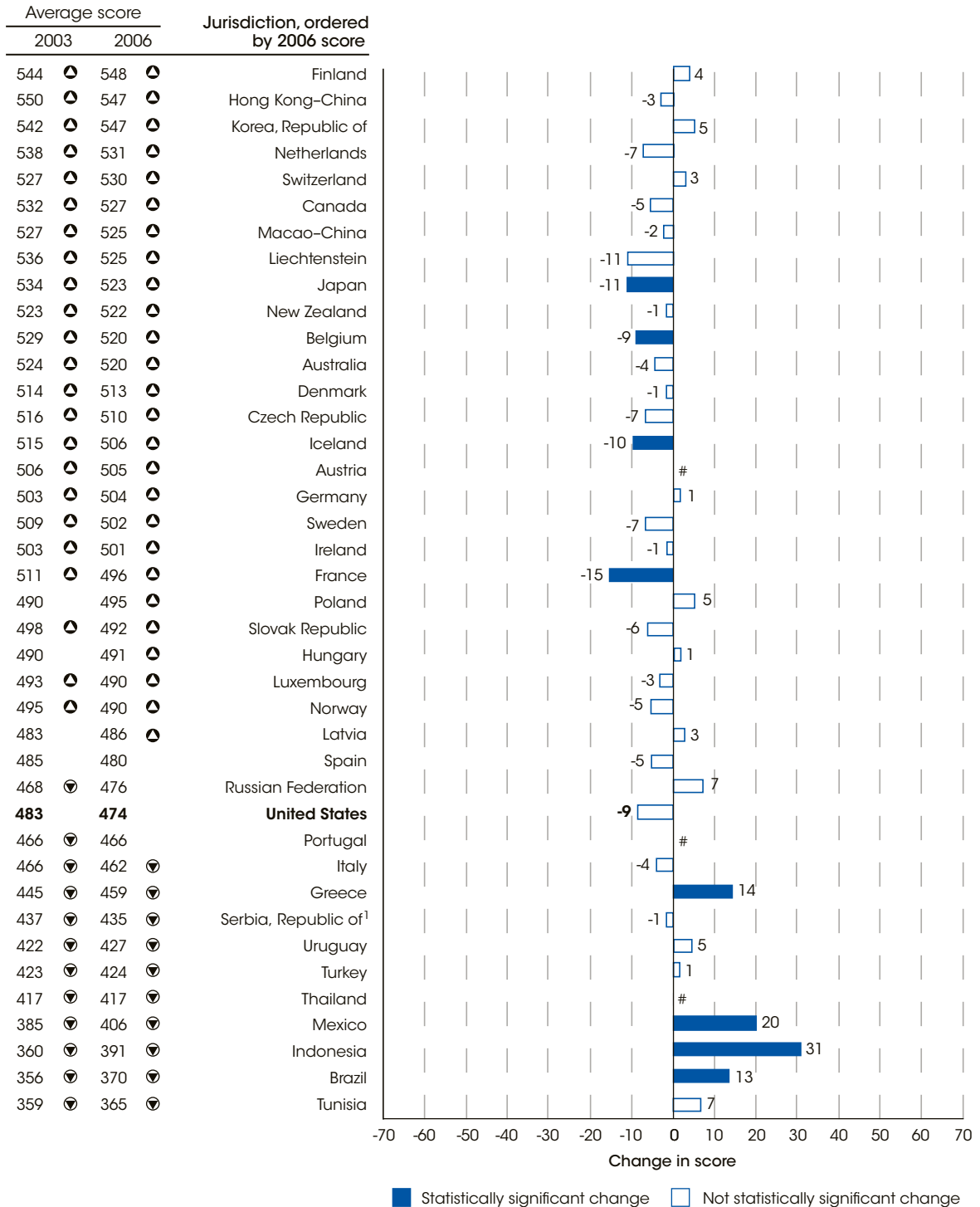
\*  $p < .05$ . Percentage at proficiency Level 6 is measurably higher than U.S. percentage.

# Rounds to zero.

NOTE: Jurisdictions are ordered by the percentage of students scoring 669.30 or above on a scale from 0 to 1000. Students with a score of 420.07 or lower appear on the left side of the percentage distribution. These students performed at proficiency level 1 or below. At Level 1 students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. They are able to identify information and to carry out routine procedures according to direct instructions in explicit situations. They can perform actions that are obvious and follow immediately from the given stimuli. The Program for International Student Assessment (PISA) has defined six levels of proficiency based on specific student proficiencies. These specific student proficiencies remain the same across all PISA assessments; however, the score point threshold for students who demonstrate these specific student proficiencies may vary slightly from assessment to assessment. Because OECD proficiency levels are anchored by specific student proficiencies (i.e., by items not scores), the percentage distribution by level can be reported. Apparent differences may not be statistically significant.

SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

**Figure 10. Change in average PISA mathematics scores of 15-year-old students in selected jurisdictions, by jurisdiction: 2003 to 2006**



▲ Average score was higher than the U.S. average score.

▼ Average score was lower than the U.S. average score.

# Rounds to zero.

<sup>1</sup> The Republics of Montenegro and Serbia were a united jurisdiction for the PISA 2003 assessment.

NOTE: Selected jurisdictions include those that participated in both PISA 2003 and PISA 2006. Mathematics results from PISA 2000 cannot be compared with mathematics results from PISA 2006 due to changes in the assessment framework; as such, mathematics results from PISA 2003 served as the basis of comparison. Average mathematics literacy scores are reported on a scale from 0 to 1,000. Tests for statistical significance take into account the standard errors for scores in both years and the linking error across assessments. For more information on linking error, see appendix A. Because the size of standard errors can vary, a small difference between the scores of one jurisdiction may be significant while a larger difference between the scores of another jurisdiction may not be significant. Detail may not sum to totals due to rounding.

SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2003 and 2006.

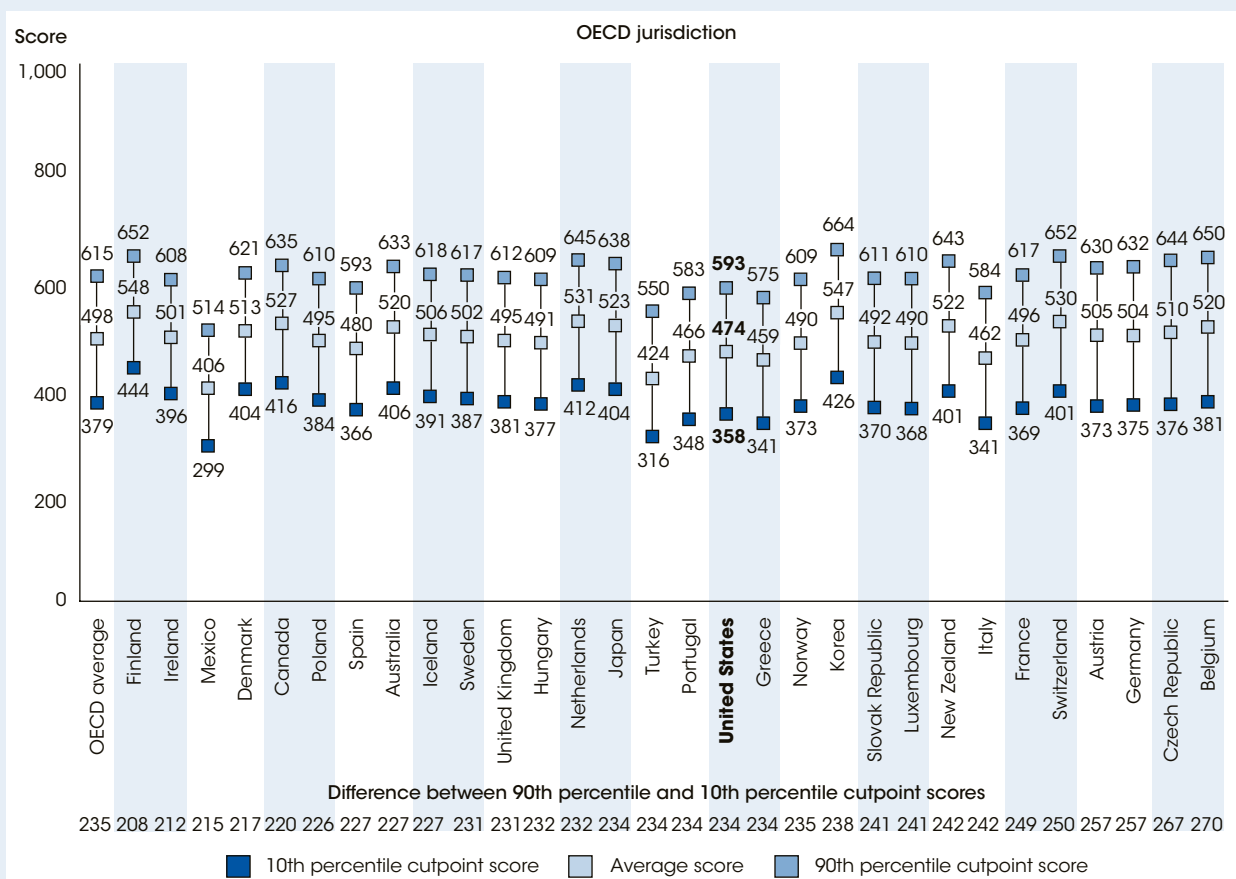
## How Much Variation Is There Between Low and High Performers in Different Countries?

The variation between low and high performers within countries provides important contextual information to understand average assessment results by providing a measure of the range or inequality of scores within a country. A common way to examine such variation is to measure the difference between cutpoint scores at the 10th and 90th percentiles for a particular subject area. The cutpoint score at the 10th percentile is the highest score achieved by the bottom 10 percent of students and the cutpoint score at the 90th percentile is the lowest score achieved by the top 10 percent of students.

Figures A-1 and A-2 show the dispersion of PISA 2006 mathematics literacy scores for 15-year-olds at the 10th and 90th percentiles. These are arranged by jurisdiction from smallest to largest gap and are shown separately for the OECD and non-OECD jurisdictions. Among the OECD countries, the differences in cutpoint scores ranged from 208 in Finland to 270 in Belgium, with the United States at 234 and an OECD average of 235. Among the non-OECD jurisdictions, the differences in cutpoint scores ranged from 117 in Azerbaijan to 277 in Israel.

As shown in figures A-1 and A-2, there is no consistent relationship between a country's average score and the variation between its low-performing 15-year-olds (i.e., those scoring at or below the 10th percentile) and

**Figure A-1. Distribution of PISA scores for 15-year-old students on the mathematics literacy scale at the 10th and 90th percentiles, by OECD jurisdiction: 2006**



NOTE: From left to right, jurisdictions appear in ascending order by the difference between their 90th percentile and 10th percentile cutpoint scores. Statistical tests were conducted, and the correlation coefficient between the jurisdiction's average score and the difference between its 90th percentile and 10th percentile cutpoint scores is 0.108.

SOURCE: Organization for Economic Cooperation and Development (OECD). (2007). *PISA 2006: Science Competencies for Tomorrow's World, Volume 2: Data*, table 6.2c. Paris: Author.

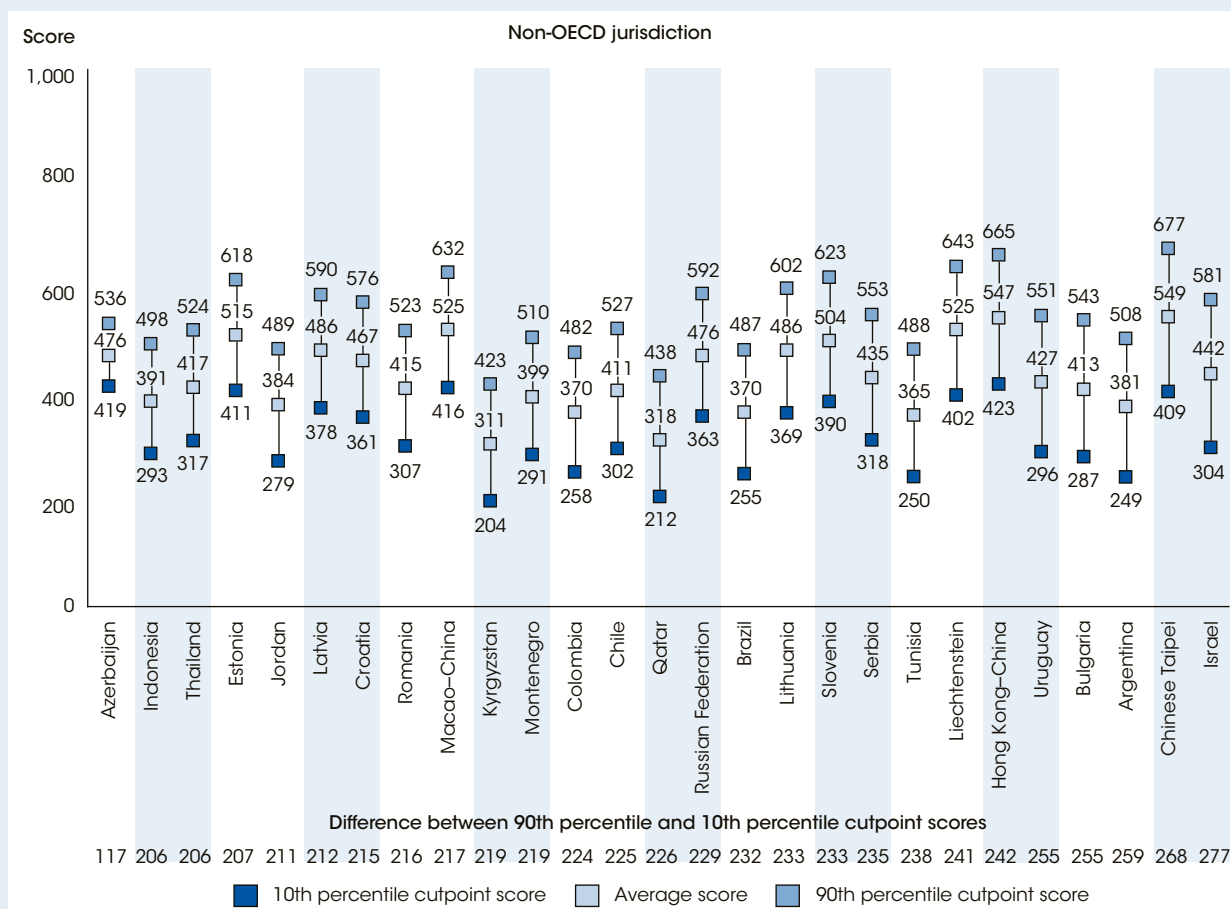


high-performing 15-year-olds (i.e., those scoring at or above the 90th percentile) in mathematics literacy. Some countries with relatively high average scores have a relatively large gap between their low and high performers (e.g., Chinese Taipei and Switzerland), while others have a relatively small gap (e.g., Finland and Canada). Similarly, some countries with relatively low average scores have a relatively large gap between their low and high performers (e.g., Argentina and Bulgaria), while others have a relatively small gap (e.g., Indonesia and Mexico).

Finland and Chinese Taipei were among the highest-performing countries in mathematics literacy. Fifteen-year-olds in these two countries (along with Korea and Hong Kong–China) scored higher, on average, than all

other countries on the mathematics literacy scale but did not measurably differ from each other. However, the relationship of low and high performers in each country was different. In Finland, the cutpoint scores at the 10th and 90th percentile were 444 and 652, respectively, for a difference of 208 points. In Chinese Taipei, the cutpoint scores at the 10th and 90th percentile were 409 and 677, respectively, for a difference of 268 points. Thus, relative to Finland, the high overall average score of 15-year-olds in Chinese Taipei can be attributed more to the performance of its very high performing students; whereas Finland’s high average score can be attributed more to the performance of students across the distribution of low to high performing students.

**Figure A-2. Distribution of PISA scores for 15-year-old students on the mathematics literacy scale at the 10th and 90th percentiles, by non-OECD jurisdiction: 2006**



SOURCE: Organization for Economic Cooperation and Development (OECD). (2007). PISA 2006: Science Competencies for Tomorrow's World, Volume 2: Data, table 6.2c. Paris: Author.

## Science

Two international assessments measure aspects of science skills: TIMSS, which focuses on students' content knowledge of the science that they are likely to have been taught in school by grades 4 and 8, and PISA, which focuses on the cognitive skills or abilities of 15-year-old students to apply science knowledge and skills to a variety of materials with a real-life context. Whereas TIMSS is closely linked to the curricula of the participating countries, PISA assesses 15-year-olds' *scientific literacy*, which it defines as

An individual's scientific knowledge and use of that knowledge to identify questions, to acquire new knowledge, to explain scientific phenomena, and to draw evidence-based conclusions about science-related issues, understanding of the characteristic features of science as a form of human knowledge and enquiry, awareness of how science and technology shape our material, intellectual, and cultural environments, and willingness to engage in science-related issues, and with the ideas of science, as a reflective citizen (OECD, 2006, p. 5).

On account of these different aims, the two assessments ask students to perform different tasks. TIMSS asks 4th- and 8th-graders to complete a range of multiple-choice and constructed response questions that test their knowledge of specific science topics or content domains—*life science*, *physical science*, and *Earth science* at grade 4 and *biology*, *chemistry*, *physics*, and *Earth science* at grade 8.<sup>35</sup> In contrast, PISA poses multiple-choice questions and constructed response questions that ask students to identify scientific issues (e.g., recognize issues that are possible to investigate scientifically), explain phenomena scientifically (e.g., describe or interpret phenomena scientifically and predict changes), and use scientific evidence (e.g., identify the assumptions, evidence, and reasoning behind conclusions). PISA presents students with a range of exercises, based on materials that they are likely to encounter as young adults, such as a discussion of acid rain, a picture of erosion at the Grand Canyon, or the results of a controlled experiment.<sup>36</sup>

It is important to note that PISA's science assessment was revised in 2006 to (1) more clearly distinguish knowledge about science as a form of human inquiry from knowledge of science, and (2) add to the framework components on the relationship between science and technology. In addition, to more clearly distinguish scientific literacy from reading literacy, the PISA 2006 science test items required less reading, on average,

than did the science items used in earlier PISA surveys. Because of these changes, it is not possible to compare science learning outcomes from PISA 2006 with those of earlier PISA assessments as is done for reading and mathematics. The differences in science performance that readers may observe when comparing PISA 2006 science scores with science scores from earlier PISA assessments may be attributable to changes in the nature of the science assessment as much as to changes in actual student performance.

### Science results for 4th- and 8th-graders

The 2007 TIMSS results showed that U.S. students' average science score was 539 for 4th-graders and 520 for 8th-graders (tables 7 and 8). Both scores were above the TIMSS scale average, which is set at 500 for every administration of TIMSS at both grades. The 4th-grade average score reflects the fact that U.S. 4th-graders performed above the TIMSS scale average in all three science content domains (life science, physical science, and Earth science) in 2007 (Gonzales et al. 2008, table 14). The 8th-grade average score reflects the fact that U.S. 8th-graders performed above the TIMSS scale average in three of the four science content domains (biology, chemistry, and Earth science) but in physics they scored not measurably different from the TIMSS scale average in 2007 (Gonzales et al. 2008, table 15).

Fourth-graders in 4 countries (Singapore, Chinese Taipei, Hong Kong, and Japan) scored above their U.S. peers, on average (table 7). The top 10 percent of U.S. 4th-graders scored 643 or higher, a cutpoint score below that of the top 10 percent of 4th-graders in Singapore and Chinese Taipei, while the bottom 10 percent scored 427 or lower, a cutpoint score below that of the bottom 10 percent of students in 7 countries. These 7 countries include the 4 with higher average scores, 2 countries with average scores that are not measurably different than the U.S. score, and 1 country with an average score lower than the U.S. score.

Eighth-graders in 9 countries (Singapore, Chinese Taipei, Japan, Korea, England, Hungary, the Czech Republic, Slovenia, and the Russian Federation) scored above their U.S. peers, on average (table 8). The top 10 percent of U.S. 8th-graders scored 623 or higher, a cutpoint score below that of the top 10 percent of 8th-graders in 6 countries (all of which had higher average scores), while the bottom 10 percent scored 410 or lower, a cutpoint score below that of the bottom 10 percent of students in 8 countries (all of which had higher average scores).

TIMSS has developed four international benchmarks to help analyze the range of students' performance in science within each participating country.<sup>37</sup> As in mathematics,

<sup>35</sup> Examples of TIMSS science items can be viewed at <http://timss.bc.edu/timss2007/items.html>.

<sup>36</sup> Examples of PISA science items can be viewed at <http://www.oecd.org/dataoecd/13/33/38709385.pdf>.

<sup>37</sup> See figures 11 and 12 for the cut scores established for all the international benchmarks. For details about the international benchmarks, see Mullis et al. (2008b), chapter 2.

**Table 7. Average TIMSS scores of fourth-grade students in science and cutpoint scores for bottom and top 10 percent of students in each jurisdiction, by jurisdiction: 2007**

Jurisdiction	Average score	Cutpoint score	
		Bottom 10 percent	Top 10 percent
<b>All jurisdictions</b>	<b>500</b> ▼	<b>359</b> ▼	<b>586</b> ▼
Singapore	587 ▲	464 ▲	701 ▲
Chinese Taipei	557 ▲	457 ▲	653 ▲
Hong Kong SAR <sup>1</sup>	554 ▲	466 ▲	637
Japan	548 ▲	459 ▲	633 ▼
Russian Federation	546	443 ▲	646
Latvia <sup>2</sup>	542	454 ▲	625 ▼
England	542	438	641
<b>United States<sup>3,4</sup></b>	<b>539</b>	<b>427</b>	<b>643</b>
Hungary	536	425	637
Italy	535	429	636
Kazakhstan <sup>2</sup>	533	433	623 ▼
Germany	528 ▼	427	623 ▼
Australia	527 ▼	423	626 ▼
Slovak Republic	526 ▼	416	627 ▼
Austria	526 ▼	423	620 ▼
Sweden	525 ▼	429	617 ▼
Netherlands <sup>5</sup>	523 ▼	445 ▲	598 ▼
Slovenia	518 ▼	416 ▼	610 ▼
Denmark <sup>3</sup>	517 ▼	417	610 ▼
Czech Republic	515 ▼	416 ▼	610 ▼
Lithuania <sup>2</sup>	514 ▼	428	595 ▼
New Zealand	504 ▼	382 ▼	614 ▼
Scotland <sup>3</sup>	500 ▼	400 ▼	593 ▼
Armenia	484 ▼	336 ▼	640
Norway	477 ▼	374 ▼	570 ▼
Ukraine	474 ▼	364 ▼	576 ▼
Iran, Islamic Republic of	436 ▼	304 ▼	558 ▼
Georgia <sup>2</sup>	418 ▼	306 ▼	524 ▼
Colombia	400 ▼	271 ▼	522 ▼
El Salvador	390 ▼	267 ▼	507 ▼
Algeria	354 ▼	220 ▼	483 ▼
Kuwait <sup>6</sup>	348 ▼	182 ▼	505 ▼
Tunisia	318 ▼	119 ▼	497 ▼
Morocco	297 ▼	139 ▼	465 ▼
Qatar	294 ▼	121 ▼	464 ▼
Yemen	197 ▼	20 ▼	379 ▼

▲ Score is higher than U.S. score.

▼ Score is lower than U.S. score.

<sup>1</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>2</sup> National Target Population does not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS).

<sup>3</sup> Met guidelines for sample participation rates only after substitute schools were included.

<sup>4</sup> National Defined Population covers 90 percent to 95 percent of National Target Population.

<sup>5</sup> Nearly satisfied guidelines for sample participation rates only after substitute schools were included.

<sup>6</sup> Kuwait tested the same cohort of students as other countries, but later in 2007, at the beginning of the next school year.

NOTE: Jurisdictions are ordered on the basis of average scores, from highest to lowest. Science scores are reported on a scale from 0 to 1,000. A cutpoint score is the threshold score for an established level of performance. The cutpoint scores for students in the top 10 percent is the 90th percentile score within the jurisdiction. The cutpoint score for students in the bottom 10 percent is the 10th percentile score within the jurisdiction. 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.

SOURCE: 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), tables 11 and 17. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

**Table 8. Average TIMSS scores of eighth-grade students in science and cutpoint scores for bottom and top 10 percent of students in each jurisdiction, by jurisdiction: 2007**

Jurisdiction	Average score	Cutpoint score	
		Bottom 10 percent	Top 10 percent
<b>All jurisdictions</b>	<b>500</b> ♣	<b>352</b> ♣	<b>573</b> ♣
Singapore	567 ▲	421	694 ▲
Chinese Taipei	561 ▲	439 ▲	665 ▲
Japan	554 ▲	454 ▲	648 ▲
Korea, Republic of	553 ▲	452 ▲	646 ▲
England <sup>1</sup>	542 ▲	427 ▲	649 ▲
Hungary	539 ▲	437 ▲	635 ▲
Czech Republic	539 ▲	447 ▲	630
Slovenia	538 ▲	442 ▲	628
Hong Kong SAR <sup>1,2</sup>	530	419	625
Russian Federation	530 ▲	427 ▲	627
<b>United States<sup>1,3</sup></b>	<b>520</b>	<b>410</b>	<b>623</b>
Lithuania <sup>4</sup>	519	414	616
Australia	515	410	617
Sweden	511 ♣	405	608 ♣
Scotland <sup>1</sup>	496 ♣	388 ♣	597 ♣
Italy	495 ♣	393 ♣	590 ♣
Armenia	488 ♣	366 ♣	612
Norway	487 ♣	389 ♣	578 ♣
Ukraine	485 ♣	374 ♣	588 ♣
Jordan	482 ♣	349 ♣	601 ♣
Malaysia	471 ♣	357 ♣	581 ♣
Thailand	471 ♣	363 ♣	578 ♣
Serbia <sup>3,4</sup>	470 ♣	359 ♣	571 ♣
Bulgaria <sup>5</sup>	470 ♣	330 ♣	595 ♣
Israel <sup>5</sup>	468 ♣	329 ♣	591 ♣
Bahrain	467 ♣	351 ♣	575 ♣
Bosnia and Herzegovina	466 ♣	359 ♣	565 ♣
Romania	462 ♣	345 ♣	572 ♣
Iran, Islamic Republic of	459 ♣	355 ♣	566 ♣
Malta	457 ♣	298 ♣	595 ♣
Turkey	454 ♣	336 ♣	577 ♣
Syrian Arab Republic	452 ♣	355 ♣	546 ♣
Cyprus	452 ♣	339 ♣	556 ♣
Tunisia	445 ♣	367 ♣	524 ♣
Indonesia	427 ♣	330 ♣	520 ♣
Oman	423 ♣	293 ♣	541 ♣
Georgia <sup>4</sup>	421 ♣	309 ♣	527 ♣
Kuwait <sup>6</sup>	418 ♣	298 ♣	530 ♣
Colombia	417 ♣	319 ♣	514 ♣
Lebanon	414 ♣	284 ♣	539 ♣
Egypt	408 ♣	275 ♣	537 ♣
Algeria	408 ♣	327 ♣	488 ♣
Palestinian National Authority	404 ♣	255 ♣	543 ♣
Saudi Arabia	403 ♣	300 ♣	503 ♣
El Salvador	387 ♣	298 ♣	477 ♣
Botswana	355 ♣	220 ♣	478 ♣
Qatar	319 ♣	146 ♣	480
Ghana	303 ♣	163 ♣	445 ♣

▲ Score is higher than U.S. score.

♣ Score is lower than U.S. score.

<sup>1</sup> Met guidelines for sample participation rates only after substitute schools were included.

<sup>2</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>3</sup> National Defined Population covers 90 percent to 95 percent of National Target Population.

<sup>4</sup> National Target Population does not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS).

<sup>5</sup> National Defined Population covers less than 90 percent of National Target Population (but at least 77 percent).

<sup>6</sup> Kuwait tested the same cohort of students as other countries, but later in 2007, at the beginning of the next school year.

NOTE: Jurisdictions are ordered on the basis of average scores, from highest to lowest. Science scores are reported on a scale from 0 to 1,000.

A cutpoint score is the threshold score for an established level of performance. The cutpoint scores for students in the top 10 percent is the 90th percentile score within the jurisdiction. The cutpoint score for students in the bottom 10 percent is the 10th percentile score within the jurisdiction.

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.

SOURCE: 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), tables 11 and 17. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

the Advanced benchmark is set at 625 score points for both grades.<sup>38</sup>

Fourth-graders reaching the Advanced benchmark demonstrate the knowledge and skills for beginning scientific inquiry. They demonstrate some understanding of Earth's features and processes and of the solar system. They can communicate their understanding of structure, function, and life processes in organisms and classify organisms according to major physical and behavioral features. They demonstrate some understanding of physical phenomena and properties of common materials.

Eighth-graders reaching the Advanced benchmark demonstrate a grasp of some complex and abstract scientific concepts. They can apply knowledge of the solar system and of Earth features, processes, and conditions, and apply understanding of the complexity of living organisms and how they relate to their environment. They show understanding of electricity, thermal expansion, and sound, as well as the structure of matter and physical and chemical properties and changes. They show understanding of environmental and resource issues. Students at this level understand some fundamentals of scientific investigation and can apply basic physical principles to solve some quantitative problems. They can provide written explanations to communicate scientific knowledge.

In 2007, fifteen percent of U.S. 4th-graders and 10 percent of U.S. 8th-graders reached the Advanced benchmark (figures 11 and 12). In comparison, 2 participating countries had a higher percentage of 4th-graders reaching this benchmark (Singapore, with 36 percent reaching this benchmark, and Chinese Taipei, with 19 percent); and 6 participating countries had a higher percentage of 8th-graders reaching this benchmark (ranging from 32 to 13 percent): Singapore, Chinese Taipei, Japan, England, Korea, and Hungary.

### *Change over time*

Among the 16 countries that participated in both the first TIMSS in 1995 and the most recent TIMSS in 2007, at grade 4, the average science score increased in 7 countries and decreased in 5 countries (figure 13). Among the 19 countries that participated in both the 1995 and 2007 TIMSS at grade 8, the average science score increased in 5 countries and decreased in 3 countries (figure 14).

Between 1995 and 2007, there was no measurable change in average score of U.S. 4th-graders in science. Average scores increased during this time in England, Hungary, Hong Kong, Slovenia, Iran, Latvia, and Singapore. Increases in Singapore (63 points) and Hong Kong (46 points) moved their 4th-graders from scoring below their U.S. peers in 1995 to scoring higher than their U.S. peers in 2007. Increases in Latvia (56 points), Hungary (28 points), and England (14 points) moved their 4th-graders from scoring below their U.S. peers in 1995 to being not measurably different than their U.S. peers in 2007.

Scores decreased during this time for 4th-graders in Japan (5 points), Austria (12 points), Scotland (14 points), the Czech Republic (17 points), and Norway (27 points). Of these countries, only Austria changed its position relative to the United States; its 4th-graders moved from being not measurably different from their U.S. peers in 1995 to scoring below their U.S. peers in 2007.

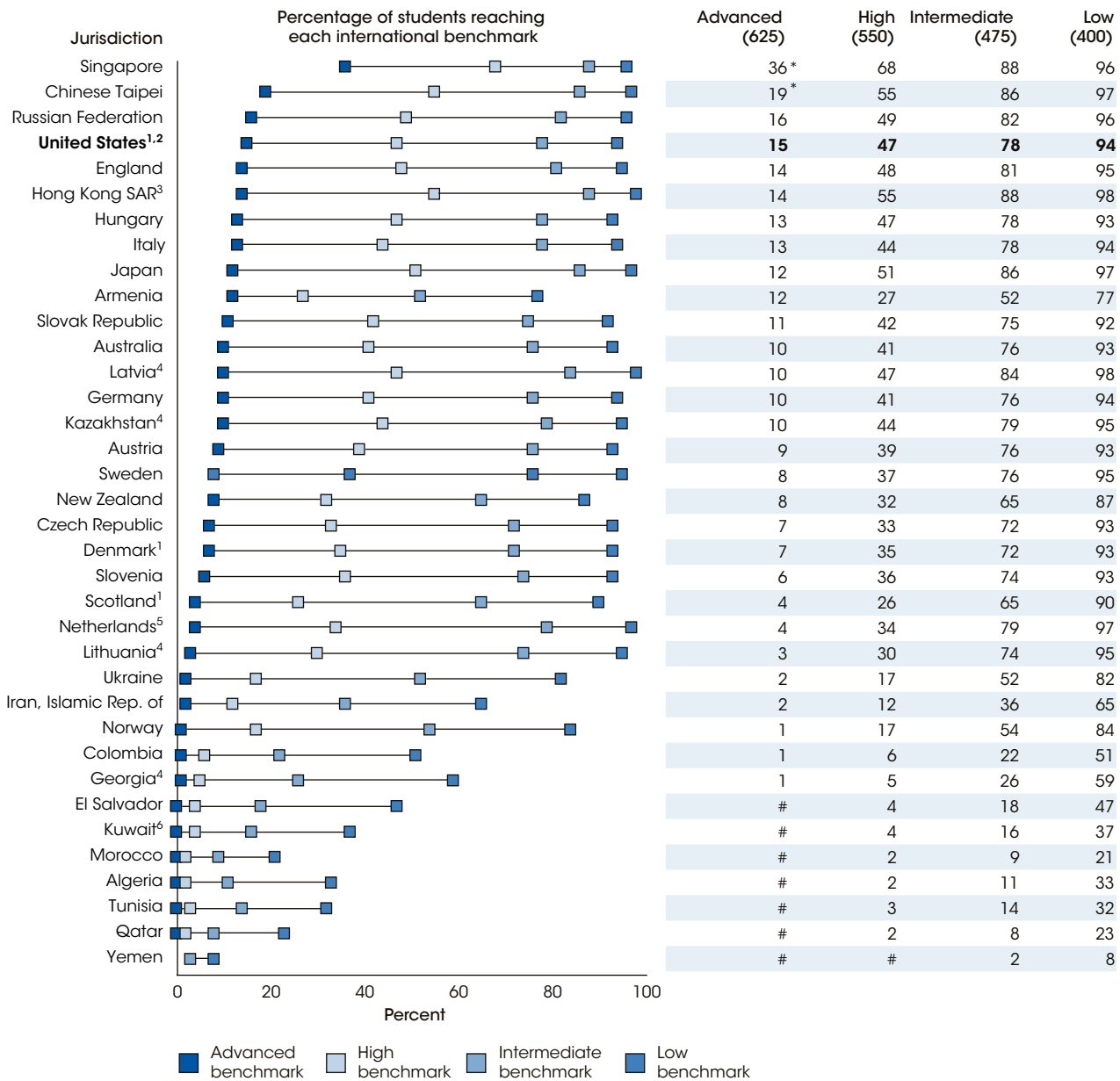
At grade 8, the U.S. average score in science did not measurably change between 1995 and 2007. Average scores increased during this time in Korea, Hong Kong, Slovenia, Colombia, and Lithuania. Two of the countries with increases in the average scores of their 8th-graders changed their position relative to the United States: Lithuania and Slovenia. An increase in Lithuania (55 points) moved their 8th-graders from scoring below their U.S. peers in 1995 to being not measurably different from their U.S. peers in 2007. An increase in Slovenia (24 points) moved their 8th-graders from being not measurably different from their U.S. peers in 1995 to scoring higher than their U.S. peers in 2007.

Scores decreased during this time for 8th-graders in the Czech Republic (16 points), Norway (28 points), and Sweden (42 points). The decrease in Norway moved their 8th-graders from being not measurably different from their U.S. peers in 1995 to scoring below their U.S. peers in 2007. The decrease in Sweden moved their 8th-graders from scoring above their U.S. peers in 1995 to scoring below their U.S. peers in 2007.

The next TIMSS assessment will be administered in 2011. More detailed results for TIMSS 2007 can be found in Gonzales et al. (2008; available at <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001>) and Mullis et al. (2008b; available at <http://timss.bc.edu/TIMSS2007/sciencereport.html>). For more information on TIMSS, see <http://nces.ed.gov/timss/>.

<sup>38</sup> The IEA set international benchmarks for TIMSS based on an analysis of score points. The score points for each benchmark remain the same across assessments; however, the configuration of items that define what students reaching a benchmark can do may vary slightly from one assessment to the next. For more details, see appendix A.

**Figure 11. Percentage of fourth-grade students reaching the TIMSS international benchmarks in science, by jurisdiction: 2007**



\*  $p < .05$ . Percentage reaching advanced benchmark is measurably higher than U.S. percentage.

# Rounds to zero.

<sup>1</sup> Met guidelines for sample participation rates only after substitute schools were included.

<sup>2</sup> National Defined Population covers 90 percent to 95 percent of National Target Population.

<sup>3</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>4</sup> National Target Population does not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS).

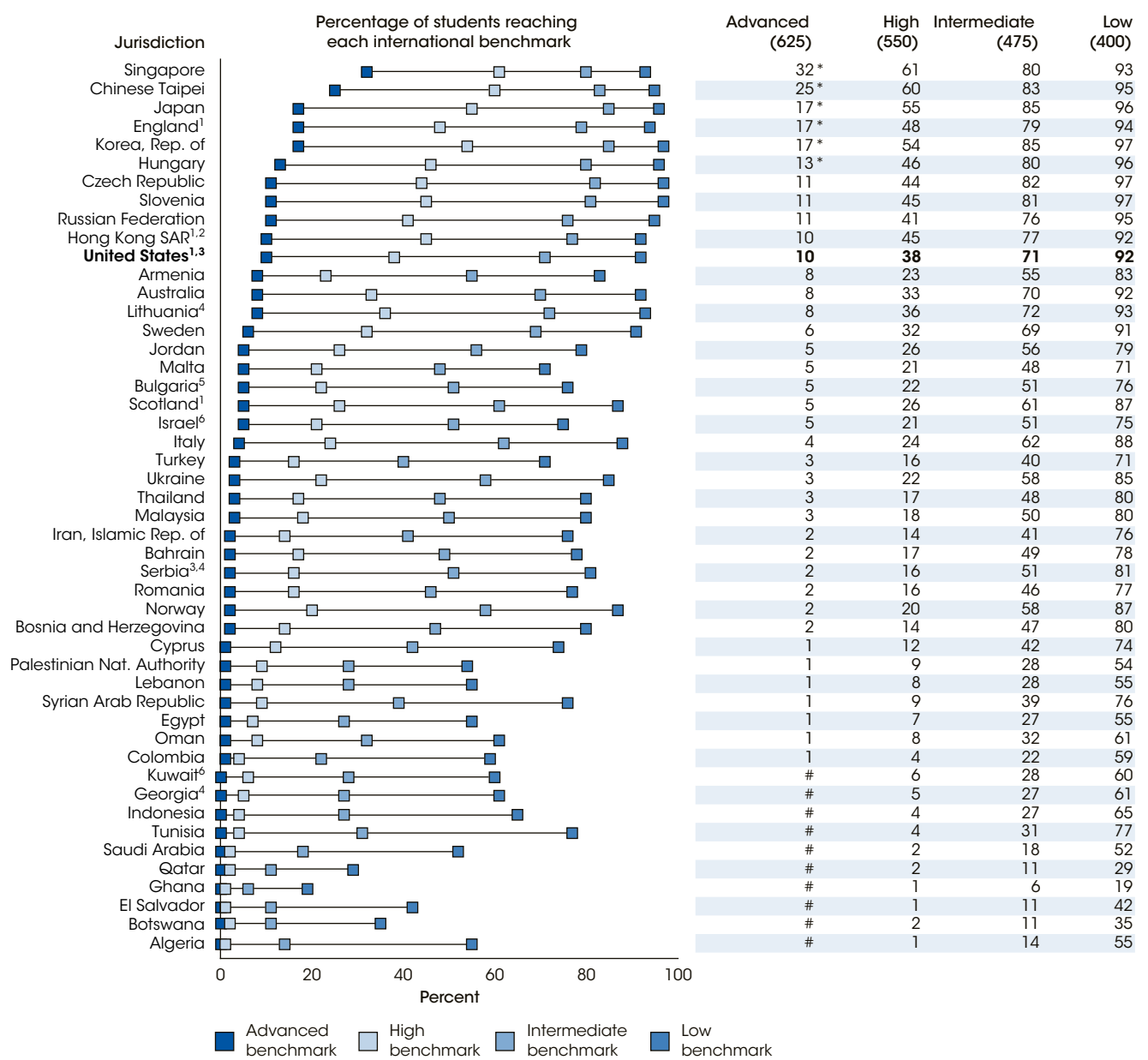
<sup>5</sup> Nearly satisfied guidelines for sample participation rates only after substitute schools were included.

<sup>6</sup> Kuwait tested the same cohort of students as other countries, but later in 2007, at the beginning of the next school year.

NOTE: Jurisdictions are ordered by the percentage of students who reached the advanced international benchmark on a scale of 0 to 1,000. The International Association for the Evaluation of Educational Achievement (IEA) set international benchmarks for TIMSS based on an analysis of score points. Students who did not reach benchmarks are not shown nor are students whose placement relative to the benchmarks was uncertain. Because IEA benchmarks are anchored by scores, it is possible to calculate the percentage of students who reach each benchmark's score threshold with accuracy but not the percentage distribution of students by benchmarks. Apparent differences may not be statistically significant.

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

**Figure 12. Percentage of eighth-grade students reaching the TIMSS international benchmarks in science, by jurisdiction: 2007**



\*  $p < .05$ . Percentage reaching advanced benchmark is measurably higher than U.S. percentage.

# Rounds to zero.

<sup>1</sup> Met guidelines for sample participation rates only after substitute schools were included.

<sup>2</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>3</sup> National Defined Population covers 90 percent to 95 percent of National Target Population.

<sup>4</sup> National Target Population does not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS).

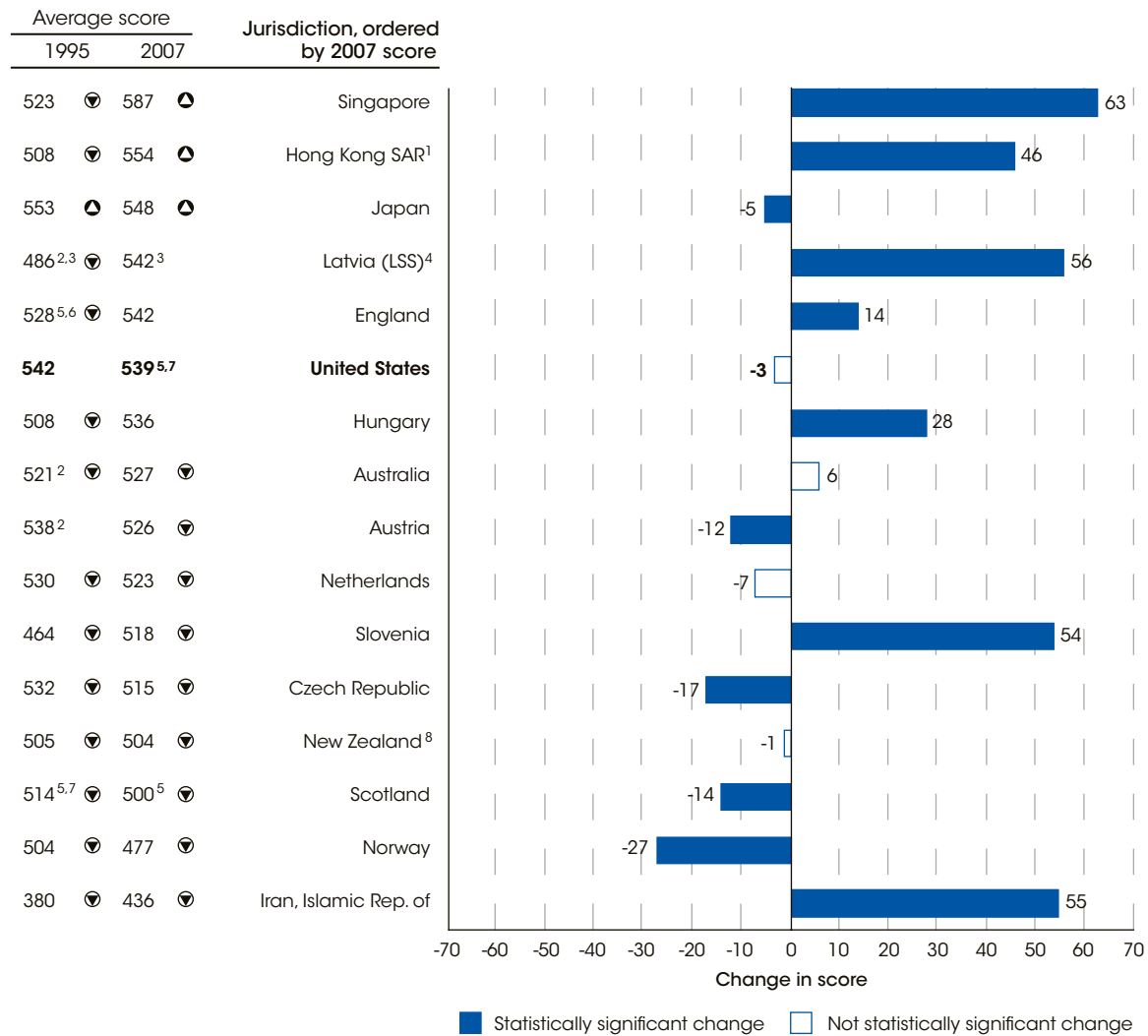
<sup>5</sup> National Defined Population covers less than 90 percent of National Target Population (but at least 77 percent).

<sup>6</sup> Kuwait tested the same cohort of students as other countries, but later in 2007, at the beginning of the next school year.

NOTE: Jurisdictions are ordered by the percentage of students who reached the advanced international benchmark on a scale of 0 to 1,000. The International Association for the Evaluation of Educational Achievement (IEA) set international benchmarks for TIMSS based on an analysis of score points. Students who did not reach benchmarks are not shown nor are students whose placement relative to the benchmarks was uncertain. Because IEA benchmarks are anchored by scores, it is possible to calculate the percentage of students who reach each benchmark's score threshold with accuracy but not the percentage distribution of students by benchmarks. Apparent differences may not be statistically significant.

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

**Figure 13. Change in average TIMSS science scores of fourth-grade students in selected jurisdictions, by jurisdiction: 1995 to 2007**



▲ Score is higher than U.S. score.

▼ Score is lower than U.S. score.

<sup>1</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>2</sup> Did not satisfy guidelines for sample participation rates.

<sup>3</sup> National Target Population did not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS).

<sup>4</sup> Only Latvian-speaking schools (LSS) are included.

<sup>5</sup> Met guidelines for sample participation rates only after substitute schools were included.

<sup>6</sup> National Defined Population covers less than 90 percent of National Target Population (but at least 77 percent).

<sup>7</sup> National Defined Population covers 90 percent to 95 percent of National Target Population.

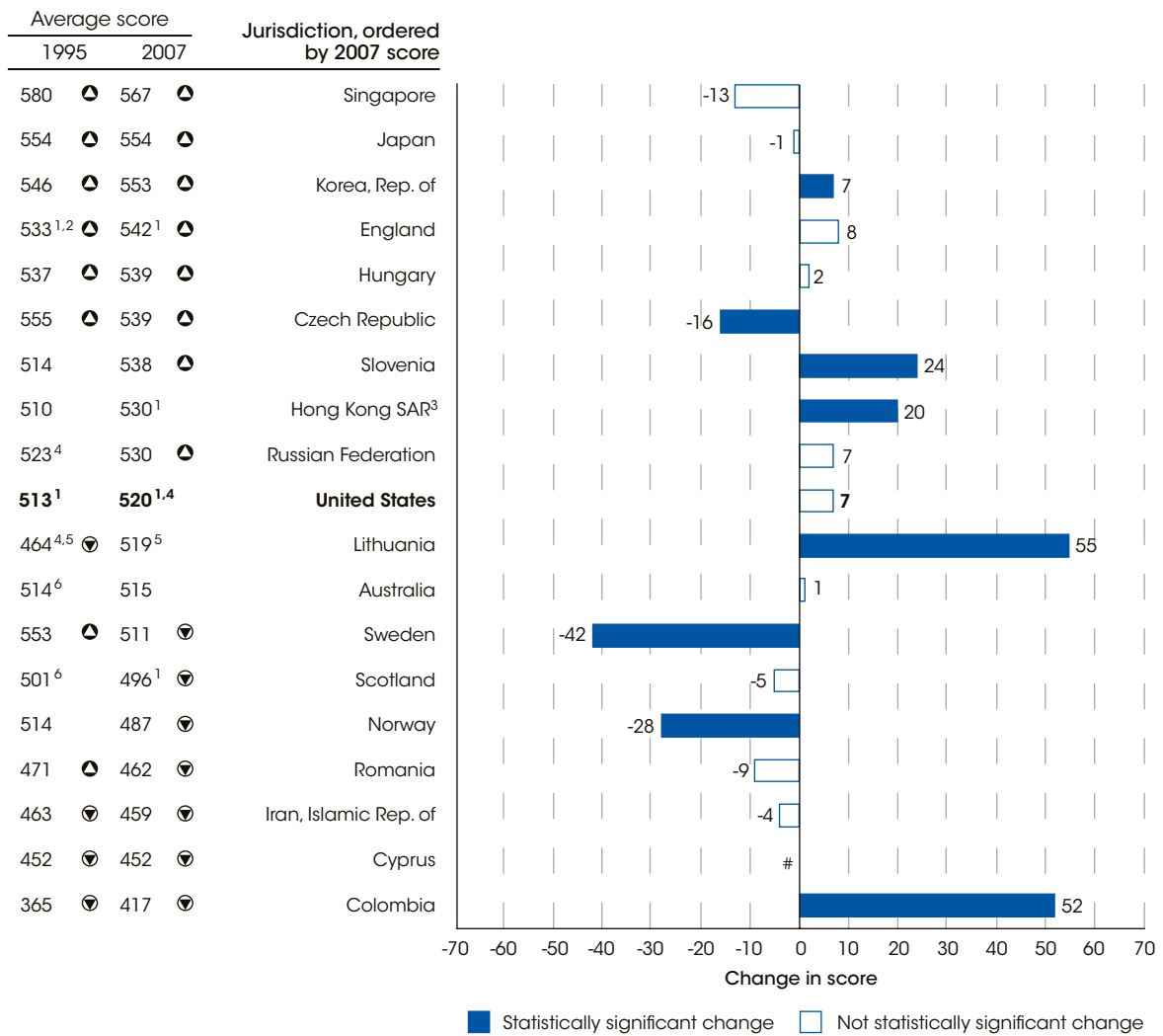
<sup>8</sup> Estimates for New Zealand have been computed for students taught in English only, which represents 98 to 99 percent of the student population.

NOTE: Selected jurisdictions include those that participated in both TIMSS 1995 and TIMSS 2007. Average science scores are reported on a scale from 0 to 1,000. Tests for statistical significance take into account the standard errors for scores in both years. Because the size of standard errors can vary, a small difference between the scores of one jurisdiction may be significant while a larger difference between the scores of another jurisdiction may not be significant. Detail may not sum to totals due to rounding.

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



**Figure 14. Change in average TIMSS science scores of eighth-grade students in selected jurisdictions, by jurisdiction: 1995 to 2007**



▲ Score is higher than U.S. score.

▼ Score is lower than U.S. score.

# Rounds to zero.

<sup>1</sup> Met guidelines for sample participation rates only after substitute schools were included.

<sup>2</sup> National Defined Population covers less than 90 percent of National Target Population (but at least 77 percent).

<sup>3</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>4</sup> National Defined Population covers 90 percent to 95 percent of National Target Population.

<sup>5</sup> National Target Population did not include all of the International Target Population defined by TIMSS.

<sup>6</sup> Nearly satisfied guidelines for sample participation rates only after substitute schools were included.

NOTE: Selected jurisdictions include those that participated in both TIMSS 1995 and TIMSS 2007. Average science scores are reported on a scale from 0 to 1,000. Tests for statistical significance take into account the standard errors for scores in both years. Because the size of standard errors can vary, a small difference between the scores of one jurisdiction may be significant while a larger difference between the scores of another jurisdiction may not be significant. Detail may not sum to totals due to rounding.

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

## How Much Does Performance Within the United States Vary by School Poverty?

As a measure of school poverty, TIMSS asked principals at public schools to report the percentage of students at the school eligible to receive free or reduced-price lunch through the National School Lunch Program. This is a federally assisted meal program that provides nutritionally balanced, low-cost or free lunches to eligible children each school day. TIMSS compares mathematics and science achievement results of students from schools with various poverty levels with the TIMSS scale average and the U.S. national average.

In both mathematics and science, the average score of U.S. 4th-graders in the highest poverty public schools (at least 75 percent of students eligible for free or reduced-price lunch) in 2007 (479 in mathematics and 477 in science) was lower than the TIMSS scale average (500); the average scores of 4th-graders in each of the other

categories of school poverty were higher than the TIMSS scale average (data not shown). The average score of U.S. 4th-graders in the lowest poverty public schools (less than 10 percent of students eligible for free or reduced-price lunch) in 2007 (583 in mathematics and 590 in science) was also higher than the U.S. national average (529 in mathematics and 539 in science).

At 8th grade for both mathematics and science, the average score of U.S. students in the highest poverty public schools in 2007 (465 in mathematics and 466 in science) was lower, on average, than the TIMSS scale average (500) (data not shown). On the other hand, U.S. 8th-graders attending public schools with fewer than 50 percent of students eligible for free or reduced-price lunch scored higher than the TIMSS scale average. The average score of U.S. 8th-graders in the lowest poverty public schools in 2007 (557 in mathematics and 572 in science) was also higher than the U.S. national average (508 in mathematics and 520 in science).

## Science results for 15-year-olds

In PISA 2006, U.S. 15-year-old students' average science literacy score of 489 was lower than the OECD average of 500 (table 9), and placed U.S. 15-year-olds in the bottom third of participating OECD nations. Fifteen-year-old students in 16 of the 29 other participating OECD-member countries outperformed their U.S. peers (as did 15-year-olds in 6 of the 27 non-OECD countries that participated) in terms of average scores. U.S. 15-year-olds in the top 10 percent scored 628 or higher, a cutpoint score below that of the top 10 percent of students in 9 OECD and 4 non-OECD countries. In these 13 countries, cutpoint scores for the top 10 percent of students ranged from 673 in Finland to 640 in Estonia. The bottom 10 percent of U.S. 15-year-olds scored 349 or lower, a cutpoint score below that of the bottom 10 percent of students in 21 OECD and 9 non-OECD countries.

PISA has developed six levels of student achievement to help analyze the range of student performance in science within each participating country.<sup>39</sup> For PISA 2006, the highest two levels of proficiency in science (above 633 score points) denote students who can identify the scientific components of many complex life situations, apply both scientific concepts and knowledge about science to these situations, and can compare, select and evaluate appropriate scientific evidence for responding to life situations. They can use well-developed inquiry abilities,

link knowledge appropriately and bring critical insights to situations. They can construct explanations based on evidence and arguments based on their critical analysis.

In 2006, nine percent of U.S. 15-year-olds performed at the highest two levels (figure 15). Thirteen countries had a higher percentage of 15-year-olds who performed at the highest two levels, with the largest percentage in Finland (21 percent). The percentage of students at the highest two levels in New Zealand, Hong Kong, Japan, Chinese Taipei, Australia, Canada, United Kingdom, The Netherlands, Slovenia, Liechtenstein, Germany, and the Czech Republic ranged from 18 to 12 percent.

### *Change over time*

Because of the revisions to the PISA science assessment for PISA 2006, direct comparisons of 2006 scores with those from 2000 and 2003 are not possible. Thus we cannot reliably say whether any country's scores increased, decreased, or were not significantly different in 2006 from the earlier administrations.

Further details on the PISA science literacy results can be found in Baldi et al. (2007; available at <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2008016>) and OECD (2007; available at <http://www.oecd.org/dataoecd/30/17/39703267.pdf>). For more information on PISA, see <http://nces.ed.gov/surveys/pisa/>.

<sup>39</sup> See figure 15 for the cut scores for all six levels of proficiency. For details about all six levels, see OECD 2007, pp. 42–44.

**Table 9. Average PISA scores of 15-year-old students on science literacy scale and cutpoint scores for bottom and top 10 percent of students in each jurisdiction, by jurisdiction: 2006**

Jurisdiction	Average score	Cutpoint score	
		Bottom 10 percent	Top 10 percent
<b>All jurisdictions</b>	<b>500</b> ▲	<b>375</b> ▲	<b>622</b>
<b>OECD</b>			
Finland	563 ▲	453 ▲	673 ▲
Canada	534 ▲	410	651 ▲
Japan	531 ▲	396 ▲	654 ▲
New Zealand	530 ▲	389 ▲	667 ▲
Australia	527 ▲	395 ▲	653 ▲
Netherlands	525 ▲	395 ▲	646 ▲
Korea, Republic of	522 ▲	403 ▲	635
Germany	516 ▲	381 ▲	642 ▲
United Kingdom	515 ▲	376 ▲	652 ▲
Czech Republic	513 ▲	385 ▲	641 ▲
Switzerland	512 ▲	378 ▲	636
Austria	511 ▲	378 ▲	633
Belgium	510 ▲	374 ▲	634
Ireland	508 ▲	385 ▲	630
Hungary	504 ▲	388 ▲	617 ▼
Sweden	503 ▲	381 ▲	622
Poland	498	381 ▲	615 ▼
Denmark	496	373 ▲	615 ▼
France	495	359	623
Iceland	491	364 ▲	614 ▼
<b>United States</b>	<b>489</b>	<b>349</b>	<b>628</b>
Slovak Republic	488	368 ▲	609 ▼
Spain	488	370 ▲	604 ▼
Norway	487	365 ▲	610 ▼
Luxembourg	486	358	609 ▼
Italy	475 ▼	351	598 ▼
Portugal	474 ▼	357	588 ▼
Greece	473 ▼	353	589 ▼
Turkey	424 ▼	325 ▼	540 ▼
Mexico	410 ▼	306 ▼	516 ▼
<b>Non-OECD</b>			
Hong Kong-China	542 ▲	418 ▲	655 ▲
Chinese Taipei	532 ▲	402 ▲	651 ▲
Estonia	531 ▲	422 ▲	640 ▲
Liechtenstein	522 ▲	393 ▲	643
Slovenia	519 ▲	391 ▲	647 ▲
Macao-China	511 ▲	409 ▲	611 ▼
Croatia	493	383 ▲	604 ▼
Latvia	490	380 ▲	597 ▼
Lithuania	488	370 ▲	604 ▼
Russian Federation	479	364	596 ▼
Israel	454 ▼	310 ▼	601
Chile	438 ▼	323 ▼	560 ▼
Serbia, Republic of	436 ▼	327 ▼	545 ▼
Bulgaria	434 ▼	300 ▼	577 ▼
Uruguay	428 ▼	306 ▼	550 ▼
Jordan	422 ▼	309 ▼	537 ▼
Thailand	421 ▼	325 ▼	524 ▼
Romania	418 ▼	314 ▼	526 ▼
Montenegro, Republic of	412 ▼	312 ▼	517 ▼
Indonesia	393 ▼	307 ▼	488 ▼
Argentina	391 ▼	259 ▼	520 ▼
Brazil	390 ▼	281 ▼	510 ▼

See notes at end of table.

**Table 9. Average PISA scores of 15-year-old students on science literacy scale and cutpoint scores for bottom and top 10 percent of students in each jurisdiction, by jurisdiction: 2006—Continued**

Jurisdiction	Average score	Cutpoint score	
		Bottom 10 percent	Top 10 percent
<b>All jurisdictions</b>	<b>500</b> ▲	<b>375</b> ▲	<b>622</b>
<b>Non-OECD—Continued</b>			
Colombia	388 ▼	280 ▼	496 ▼
Tunisia	386 ▼	283 ▼	495 ▼
Azerbaijan	382 ▼	316 ▼	456 ▼
Qatar	349 ▼	253 ▼	462 ▼
Kyrgyz Republic	322 ▼	220 ▼	428 ▼

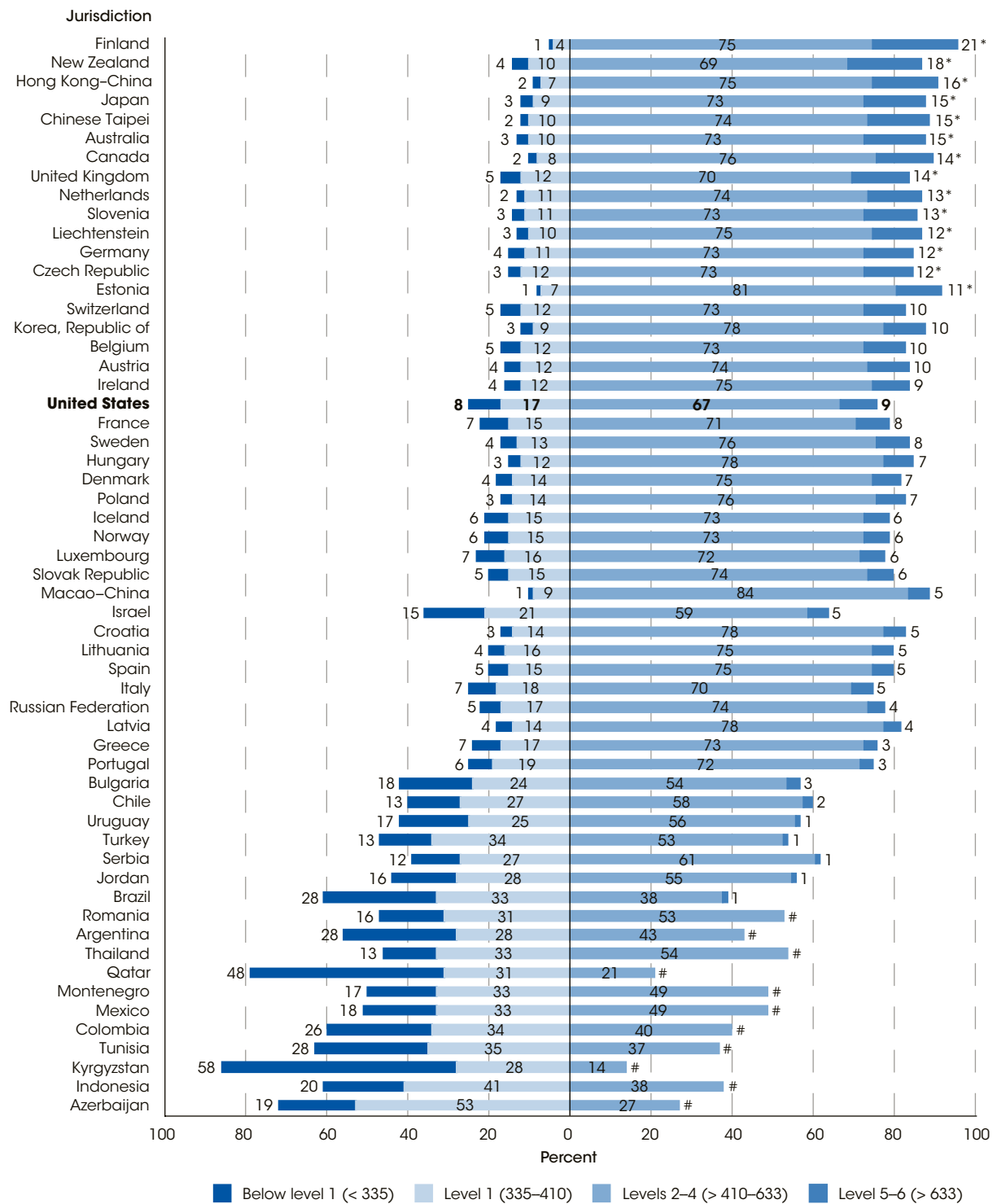
▲ Score is higher than U.S. score.

▼ Score is lower than U.S. score.

NOTE: Because the Program for International Student Assessment (PISA) is principally an Organization for Economic Cooperation and Development (OECD) study, the results for non-OECD jurisdictions are displayed separately from those of the OECD countries. Jurisdictions are ordered on the basis of average scores, from highest to lowest within the OECD countries and the non-OECD jurisdictions. Combined science literacy scores are reported on a scale from 0 to 1,000. A cutpoint score is the threshold score for an established level of performance. The cutpoint score for students in the top 10 percent is the 90th percentile score within the jurisdiction. The cutpoint score for students in the bottom 10 percent is the 10th percentile score within the jurisdiction. 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.

SOURCE: Organization for Economic Cooperation and Development (OECD). (2007). *PISA 2006: Science Competencies for Tomorrow's World, Volume 2: Data*, table 2.1c. Paris: Author.

**Figure 15. Percentage distribution of 15-year-old students on PISA science literacy scale, by proficiency level and jurisdiction: 2006**



\*  $p < .05$ . Percentage reaching advanced benchmark is measurably higher than U.S. percentage.

# Rounds to zero.

NOTE: Jurisdictions are ordered by the percentage of students scoring 633.33 or above on a scale from 0 to 1000. Students with a score of 409.54 or lower appear on the left side of the percentage distribution. These students performed at proficiency level 1 or below. At level 1, students have such a limited scientific knowledge that it can only be applied to a few familiar situations. They should be able to present scientific explanations that are obvious and follow concretely from given evidence. The Program for International Student Assessment (PISA) has defined six levels of proficiency based on specific student proficiencies. These specific student proficiencies remain the same across all PISA assessments; however, the score point threshold for students who demonstrate these specific student proficiencies may vary slightly from assessment to assessment. Because OECD proficiency levels are anchored by specific student proficiencies (i.e., by items not scores), the percentage distribution by level can be reported. Apparent differences may not be statistically significant.

SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

---

## Synthesis of science results

Both Japan and Chinese Taipei outperformed the United States in science at grade 4 each time they participated in TIMSS (see supplemental table A-6).<sup>40</sup> In addition, between 1995 and 2007, U.S. 4th-graders fell behind their peers in Hong Kong and Singapore, where science scores improved while U.S. scores did not. In 1995, U.S. 4th-graders outperformed their peers in Hong Kong and Singapore; in 2007, 4th-graders in Hong Kong and Singapore outperformed their U.S. peers.

At grade 8, seven countries consistently outperformed the United States in science each time they participated in TIMSS: Chinese Taipei, the Czech Republic, England, Hungary, Japan, Korea, and Singapore (see supplemental table A-7). In addition, U.S. 8th-graders fell behind their peers in the Russian Federation. However, U.S. 8th-graders' standing improved relative to their peers in Norway, Scotland, and Sweden. Declines in scores in Sweden moved their 8th-graders from scoring higher

than the United States in 1995 to scoring below the United States in 2007. In Norway and Scotland, 8th-graders moved from being not measurably different from their U.S. peers in 1995 to scoring below their U.S. peers in 2007.

Although direct comparisons of PISA 2006 science scores with those from 2000 and 2003 are not possible, one can compare the relative standing of the United States with other participating countries. In 2000, 15-year-olds in 8 countries (all OECD countries) outperformed their U.S. peers in terms of average scores: Australia, Austria, Canada, Finland, Japan, Korea, New Zealand, and the United Kingdom (see supplemental table A-8). In 2006 these same 8 countries again outperformed the United States along with 8 additional countries that did not measurably differ from the United States in 2000 (Belgium, the Czech Republic, Germany, Hungary, Ireland, Sweden, and Switzerland—all members of the OECD—and one non-OECD country, Liechtenstein).<sup>41</sup>

---

<sup>40</sup> Japan participated in TIMSS in all three rounds (1995, 2003, and 2007) while Chinese Taipei participated in TIMSS only in 2003 and 2007.

<sup>41</sup> The Netherlands, Hong Kong, Chinese Taipei, Estonia, Slovenia, and Macao also outperformed the United States in 2006, but they did not participate in PISA 2000.

## Summary

The performance of U.S. students neither leads nor trails the world in reading, mathematics, or science at any grade or age. In the most recent international assessments, the highest average scores

- in reading were from the Russian Federation, Hong Kong, Alberta, Singapore, and British Columbia at 4th-grade; and from Korea among 15-year-olds;
- in mathematics were from Hong Kong and Singapore at 4th-grade; from Chinese Taipei, Korea, and Singapore at 8th-grade; and from Chinese Taipei, Finland, Hong Kong, and Korea among 15-year-olds; and
- in science were from Singapore at 4th-grade; from Singapore and Chinese Taipei at 8th-grade; and from Finland among 15-year-olds.

In reading, U.S. student performance has not changed since the first administrations of PIRLS and PISA, and U.S. students in general have performed either not measurably different or higher than their peers in roughly three-quarters of the other countries that have participated in PIRLS and PISA assessments. However, between 2001 and 2006, 4th-graders in Hong Kong, Hungary, Italy, the Russian Federation, Singapore, and Ontario improved their scores and in 2006 surpassed their U.S. peers, whose scores were not measurably different than in 2001. Because of the lack of valid 2006 PISA reading scores for the United States, we do not know whether the reading achievement of U.S. 15-year-olds changed between 2003 and 2006, but in 2003, U.S. 15-year-olds performed in the middle third of OECD countries.

In mathematics, U.S. student scores have improved at both grades 4 and 8 since the first administration of TIMSS; however, the scores of U.S. 15-year-olds in PISA have not measurably changed. In 2007, U.S. 4th-graders improved their average score and surpassed their peers in four European countries that outperformed the United States in 1995, though England and Latvia improved more and outperformed the United States in 2003 and 2007. Also in 2007, U.S. 8th-graders gained ground relative to their European and Australian peers—though they have not caught up with their peers in Chinese Taipei, Hong Kong, Japan, Korea, or Singapore. The most recent PISA assessments, however, suggest that U.S. 15-year-olds are not as successful in applying mathematics knowledge and skills to real-world tasks as their peers in most other OECD countries. In the 2006 PISA assessment, U.S. 15-year-olds' average scores were not measurably different than in 2003, keeping the United States in the bottom quarter of OECD countries.

In science, results from the most recent TIMSS assessment show that U.S. 4th-graders have fallen behind their peers in several countries, even though their average scores in science have not declined since the first TIMSS assessment in 1995. At the 8th grade, U.S. scores on the most recent assessment were also not measurably different than in 1995, but they fell behind those in the Russian Federation and they continue to lag behind those in Chinese Taipei, the Czech Republic, England, Hungary, Japan, Korea, and Singapore. The most recent PISA assessment suggests that U.S. 15-year-olds are not able to apply scientific knowledge and skills to real-world tasks as well as their peers in the majority of other OECD countries: in the most recent science assessment of 15-year-olds, the United States continued to perform below the OECD average.

---

*This page intentionally left blank.*

---



# References

## References

- Baer, J., Baldi, S., Ayotte, K., and Green, P. (2007). *The Reading Literacy of U.S. Fourth-Grade Students in an International Context: Results From the 2001 and 2006 Progress in International Reading Literacy Study (PIRLS)*. (NCES 2008-017). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- Baldi, S., Jin, Y., Skemer, M., Green, P.J., and Herget, D. (2007). *Highlights From PISA 2006: Performance of U.S. 15-Year-Old Students in Science and Mathematics Literacy in an International Context* (NCES 2008-016). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. (Available at <http://nces.ed.gov/pubs2008/2008016.pdf>)
- Beaton, A.E., and Gonzalez, E. (1995). *The NAEP Primer*. Chestnut Hill, MA: Boston College.
- Education Sciences Reform Act of 2002, P.L. 107-279, 116 Stat. 1940 (2002).
- 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.
- Lemke, M., Sen, A., Pahlke, E., Partelow, L., Miller, D., Williams, T., Kastberg, D., and Jocelyn, L. (2004). *International Outcomes of Learning in Mathematics Literacy and Problem Solving: PISA 2003 Results From the U.S. Perspective* (NCES 2005-003R). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Martin, M.O., Mullis, I.V.S., and Kennedy, A.M. (2007). *PIRLS 2006 Technical Report*. Chestnut Hill, MA: Boston College.
- Mullis, I.V.S., Martin, M.O., Kennedy, A.M., and Foy, P. (2007). *IEA's Progress in International Reading Literacy Study in Primary School in 40 Countries*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I.V.S., Martin, M.O., & Foy, P. (with Olson, J.F., Preuschoff, C., Erberber, E., Arora, A., & Galia, J.). (2008a). *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I.V.S., Martin, M.O., & Foy, P. (with Olson, J.F., Preuschoff, C., Erberber, E., Arora, A., & Galia, J.). (2008b). *TIMSS 2007 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- National Research Council. (2002). *Methodological Advances in Cross-National Surveys of Educational Achievement*. Washington, DC: National Academy Press.
- National Center for Education Statistics, U.S. Department of Education. (1997). *Pursuing Excellence: A Study of U.S. Fourth-Grade Mathematics and Science Achievement in International Context* (NCES 97-255). Washington, DC: U.S. Government Printing Office.
- National Center for Education Statistics, U.S. Department of Education. (2000). *Highlights from TIMSS-R* (NCES 2001-027). Washington, DC: U.S. Government Printing Office.
- Neidorf, T.S., Binkley, M., Gattis, K., and Nohara, D. (2006). *Comparing Mathematics Content in the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 Assessments* (NCES 2006-029). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Olson, J.F., Martin, M.O., and Mullis, I.V.S. (Eds.). (2008). *TIMSS 2007 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Organization for Economic Cooperation and Development (OECD). (2004). *Learning for Tomorrow's World: First Results From PISA 2003*. Paris: Author.
- Organization for Economic Cooperation and Development (OECD). (2006). *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006*. Paris: Author.
- Organization for Economic Cooperation and Development (OECD). (2007). *PISA 2006 Science Competencies for Tomorrow's World*. Paris: Author.
- Organization for Economic Cooperation and Development (OECD). (2008). *PISA 2006 Technical Report*. Paris: Author.
- Peak, L. (1996). *Pursuing Excellence: A Study of U.S. Eighth-Grade Mathematics and Science Teaching, Learning, Curriculum, and Achievement in International Context* (NCES 97-198). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.
- Seastrom, M. (2003). *NCES Statistical Standards*. (NCES 2003-061). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- For more information, visit the following websites:  
TIMSS: <http://nces.ed.gov/timss> or <http://www.timss.org>  
PIRLS: <http://nces.ed.gov/surveys/pirls> or <http://www.pirls.org>  
PISA: <http://nces.ed.gov/surveys/pisa> or <http://www.pisa.oecd.org>  
International Comparisons: <http://nces.ed.gov/surveys/international>

# Supplemental Tables

# Supplemental Tables

**Table A-1. Average PIRLS scores of fourth-grade students on combined reading literacy scale, by jurisdiction: 2001 and 2006**

Jurisdiction	2001		2006	
	Average score	Standard error	Average score	Standard error
PIRLS scale average	500	0.0	500	0.0
Argentina	420	5.9	—	†
Austria	—	†	538	2.2
Belgium (Flemish)	—	†	547 <sup>1</sup>	2.0
Belgium (French)	—	†	500	2.6
Belize	327	4.7	—	†
Bulgaria	550	3.8	547	4.4
Canada, Alberta	—	†	560	2.4
Canada, British Columbia	—	†	558	2.6
Canada, Nova Scotia	—	†	542	2.2
Canada, Ontario	548	3.3	555	2.7
Canada, Quebec	537	3.0	533	2.8
Chinese Taipei	—	†	535	2.0
Colombia	422	4.4	—	†
Cyprus	494	3.0	—	†
Czech Republic	537	2.3	—	†
Denmark	—	†	546	2.3
England	553 <sup>1,2</sup>	3.4	539	2.6
France	525	2.4	522	2.1
Georgia	—	†	471	2.2
Germany	539	1.9	548	2.4
Greece	524 <sup>2</sup>	3.5	—	†
Hong Kong, SAR <sup>3</sup>	528	3.1	564	2.4
Hungary	543	2.2	551	3.0
Iceland	512	1.2	511	1.3
Indonesia	—	†	405	4.1
Iran, Islamic Republic of	414	4.2	421	3.1
Israel	509 <sup>4</sup>	2.8	512	3.3
Italy	541	2.4	551	2.9
Kuwait	( <sup>5</sup> )	†	330	4.2
Latvia	545	2.3	541	2.3
Lithuania	543 <sup>6</sup>	2.6	537	1.6
Luxembourg	—	†	557	1.1
Macedonia, Republic of	442	4.6	442	4.1
Moldova, Republic of	492	4.0	500	3.0
Morocco	350 <sup>7</sup>	9.6	323	5.9
Netherlands	554 <sup>1</sup>	2.5	547 <sup>1</sup>	1.5
New Zealand	529	3.6	532	2.0
Norway	499	2.9	498 <sup>8</sup>	2.6
Poland	—	†	519	2.4
Qatar	—	†	353	1.1
Romania	512	4.6	489	5.0
Russian Federation	528	4.4	565	3.4
Scotland	528 <sup>1</sup>	3.6	527 <sup>1</sup>	2.8
Singapore	528	5.2	558	2.9
Slovak Republic	518	2.8	531	2.8
Slovenia	502	2.0	522	2.1
South Africa	—	†	302	5.6
Spain	—	†	513	2.5
Sweden	561	2.2	549	2.3

See notes at end of table.

**Table A-1. Average PIRLS scores of fourth-grade students on combined reading literacy scale, by jurisdiction: 2001 and 2006—Continued**

Jurisdiction	2001		2006	
	Average score	Standard error	Average score	Standard error
PIRLS scale average	500	0.0	500	0.0
Trinidad and Tobago	—	†	436	4.9
Turkey	449	3.5	—	†
<b>United States</b>	<b>542<sup>1</sup></b>	<b>3.8</b>	<b>540<sup>1</sup></b>	<b>3.5</b>

▲ Average score is higher than U.S. average score.

▼ Average score is lower than U.S. average score.

— Not available. (Data were not collected or not reported.)

† Not applicable.

<sup>1</sup> Met guidelines for sample participation rates only after substitute schools were included.

<sup>2</sup> National Defined Population covers less than 95 percent of National Target Population.

<sup>3</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>4</sup> National Defined Population covers less than 80 percent of National Target Population.

<sup>5</sup> Although Kuwait participated in 2001 and 2006, the International Association for the Evaluation of Educational Achievement (IEA) elected not to report the 2001 estimates for the country because of concerns about the quality of Kuwait's data.

<sup>6</sup> National Target Population does not cover all of International Target Population because coverage falls below 65 percent.

<sup>7</sup> Nearly satisfied guidelines for sample participation rates after substitute schools were included.

<sup>8</sup> Did not meet guidelines for sample participation rates after substitute schools were included.

NOTE: Jurisdictions are ordered alphabetically. Combined reading literacy scores are reported on a scale from 0 to 1,000. 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.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), 2001 and 2006. Baer, J., Baldi, S., Ayotte, K., and Green, P. (2007). *The Reading Literacy of U.S. Fourth-Grade Students in an International Context: Results From the 2001 and 2006 Progress in International Reading Literacy Study (PIRLS)* (NCES 2008-017), figure 3 and table R-1. National Center for Education Statistics, Institute for Education Sciences, U.S. Department of Education, Washington, DC; Ogle, L., Sen, A., Pahlke, E., Jocelyn, L., Kastberg, D., Roey, S., and Williams, T. (2003). *International Comparisons in Fourth-Grade Reading Literacy: Findings from the Progress in International Reading Literacy Study (PIRLS) of 2001* (NCES 2003-073), figure 3 and table A1.1. National Center for Education Statistics, Institute for Education Sciences, U.S. Department of Education, Washington, DC.

**Table A-2. Average PISA scores of 15-year-old students on combined reading literacy scale, by jurisdiction: 2000, 2003, and 2006**

Jurisdiction	2000 <sup>1</sup>		2003		2006	
	Average score	Standard error	Average score	Standard error	Average score	Standard error
OECD average	500	0.6	494	0.7	492	0.6
<b>OECD</b>						
Australia	▲ 528	3.5	▲ 525	2.1	513	2.1
Austria	507	2.4	491	3.8	490	4.1
Belgium	507	3.6	▲ 507	2.6	501	3.0
Canada	▲ 534	1.6	▲ 528	1.7	527	2.4
Czech Republic	492	2.4	489	3.5	483	4.2
Denmark	497	2.4	492	2.8	494	3.2
Finland	▲ 546	2.6	▲ 543	1.6	547	2.1
France	505	2.7	496	2.7	488	4.1
Germany	▼ 484	2.5	491	3.4	495	4.4
Greece	▼ 474	5.0	▼ 472	4.1	460	4.0
Hungary	▼ 480	4.0	▼ 482	2.5	482	3.3
Iceland	507	1.5	492	1.6	484	1.9
Ireland	▲ 527	3.2	▲ 515	2.6	517	3.5
Italy	▼ 487	2.9	▼ 476	3.0	469	2.4
Japan	▲ 522	5.2	498	3.9	498	3.6
Korea, Republic of	▲ 525	2.4	▲ 534	3.1	556	3.8
Luxembourg	▼ 441	1.6	▼ 479	1.5	479	1.3
Mexico	▼ 422	3.3	▼ 400	4.1	410	3.1
Netherlands	( <sup>2</sup> )	†	▲ 513	2.9	507	2.9
New Zealand	▲ 529	2.8	▲ 522	2.5	521	3.0
Norway	505	2.8	500	2.8	484	3.2
Poland	▼ 479	4.5	497	2.9	508	2.8
Portugal	▼ 470	4.5	▼ 478	3.7	472	3.6
Slovak Republic	—	†	▼ 469	3.1	466	3.1
Spain	493	2.7	▼ 481	2.6	461	2.2
Sweden	516	2.2	▲ 514	2.4	507	3.4
Switzerland	494	4.2	499	3.3	499	3.1
Turkey	—	†	▼ 441	5.8	447	4.2
United Kingdom	▲ 523	2.6	( <sup>3</sup> )	†	495	2.3
<b>United States</b>	<b>504</b>	<b>7.0</b>	<b>495</b>	<b>3.2</b>	<b>(<sup>4</sup>)</b>	<b>†</b>
<b>Non-OECD</b>						
Albania	▼ 349 <sup>5</sup>	3.3	—	†	—	†
Argentina	▼ 418 <sup>5</sup>	9.9	—	†	374	7.2
Azerbaijan	—	†	—	†	353	3.1
Brazil	▼ 396	3.1	▼ 403	4.6	393	3.7
Bulgaria	▼ 430 <sup>5</sup>	4.9	—	†	402	6.9
Chile	▼ 410 <sup>5</sup>	3.6	—	†	442	5.0
Chinese Taipei	—	†	—	†	496	3.4
Colombia	—	†	—	†	385	5.1
Croatia	—	†	—	†	477	2.8
Estonia	—	†	—	†	501	2.9
Hong Kong–China	▲ 525 <sup>5</sup>	2.9	▲ 510	3.7	536	2.4
Indonesia	▼ 371 <sup>5</sup>	4.0	▼ 382	3.4	393	5.9
Israel	▼ 452 <sup>5</sup>	8.5	—	†	439	4.6
Jordan	—	†	—	†	401	3.3
Kyrgyz Republic	—	†	—	†	285	3.5
Latvia	▼ 458	5.3	491	3.7	479	3.7
Liechtenstein	▼ 483	4.1	▲ 525	3.6	510	3.9
Lithuania	—	†	—	†	470	3.0
Macao–China	—	†	498	2.2	492	1.1
Macedonia, FYR	▼ 373 <sup>5</sup>	1.9	—	†	—	†
Montenegro, Republic of <sup>6</sup>	—	†	▼ 412	3.6	392	1.2
Peru	▼ 327 <sup>5</sup>	4.4	—	†	—	†

See notes at end of table.

**Table A-2. Average PISA scores of 15-year-old students on combined reading literacy scale, by jurisdiction: 2000, 2003, and 2006—Continued**

Jurisdiction	2000 <sup>1</sup>		2003		2006	
	Average score	Standard error	Average score	Standard error	Average score	Standard error
OECD average	500	0.6	494	0.7	492	0.6
<b>Non-OECD—Continued</b>						
Qatar	—	†	—	†	312	1.2
Romania	Ⓣ 428	3.5	—	†	396	4.7
Russian Federation	Ⓣ 462	4.2	Ⓣ 442	3.9	440	4.3
Serbia, Republic of <sup>6</sup>	—	†	Ⓣ 412	3.6	401	3.5
Slovenia	—	†	—	†	494	1.0
Thailand	Ⓣ 431 <sup>5</sup>	3.2	Ⓣ 420	2.8	417	2.6
Tunisia	—	†	Ⓣ 375	2.8	380	4.0
Uruguay	—	†	Ⓣ 434	3.4	413	3.4

Ⓢ Average score is higher than U.S. average score.

Ⓣ Average score is lower than U.S. average score.

— Not available. (Data were not collected or not reported.)

† Not applicable.

<sup>1</sup> Tests for significance have been updated from what was presented in the Program for International Student Assessment 2000 U.S. report (Lemke et al. 2001). Some country differences reported here were not reported as statistically significant in the PISA 2000 U.S. report. In that report, a Bonferroni adjustment was used in all multiple comparisons of countries. This was not the case when PISA 2003 and 2006 data were analyzed and reported, thus PISA 2000 tests have been updated in this report to keep consistency across years. For more details, see appendix A.

<sup>2</sup> Because of technical problems with its sample, 2000 data for The Netherlands have been omitted from the Program for International Student Assessment (PISA) reports.

<sup>3</sup> Because of low response rates, 2003 data for the United Kingdom have been omitted from PISA reports.

<sup>4</sup> PISA 2006 reading literacy results for the United States were compromised because of an error in printing the test booklets.

<sup>5</sup> Did not participate with other jurisdictions in PISA in 2000; administered PISA in 2001.

<sup>6</sup> The Republics of Montenegro and Serbia were a united jurisdiction for the PISA 2003 assessment.

NOTE: The OECD average is the average of the national averages of the OECD member jurisdictions. Because PISA is principally an OECD study, the results for non-OECD jurisdictions are displayed separately from those of the OECD jurisdictions and are not included in the OECD average. Jurisdictions are ordered alphabetically. Combined reading literacy scores are reported on a scale from 0 to 1,000. 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.

SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2000, 2003, and 2006.

**Table A-3. Average TIMSS mathematics scores of fourth-grade students on combined mathematics scale, by jurisdiction: 1995, 2003, and 2007**

Jurisdiction	1995		2003		2007	
	Average score	Standard error	Average score	Standard error	Average score	Standard error
TIMSS scale average	500	0.0	500	0.0	500	0.0
Algeria	—	†	—	†	378	5.2
Armenia	—	†	456	3.5	500	4.3
Australia	495 <sup>1</sup>	3.4	499 <sup>2</sup>	3.9	516	3.5
Austria	531 <sup>1</sup>	2.9	—	†	505	2.0
Belgium (Flemish)	—	†	551	1.8	—	†
Canada	506	3.4	—	†	—	†
Chinese Taipei	—	†	564	1.8	576	1.7
Colombia	—	†	—	†	355	5.0
Cyprus	475	3.2	510	2.4	—	†
Czech Republic	541	3.1	—	†	486	2.8
Denmark	—	†	—	†	523 <sup>2</sup>	2.4
El Salvador	—	†	—	†	330	4.1
England	484 <sup>2,3</sup>	3.3	531 <sup>2</sup>	3.7	541	2.9
Georgia	—	†	—	†	438 <sup>4</sup>	4.2
Germany	—	†	—	†	525	2.3
Greece	463	5.6	—	†	—	†
Hong Kong SAR <sup>5</sup>	557	4.0	575 <sup>2</sup>	3.2	607	3.6
Hungary	521	3.6	529 <sup>6</sup>	3.1	510	3.5
Iceland	453	3.0	—	†	—	†
Indonesia	(?)	†	—	†	—	†
Iran, Islamic Republic of	387	5.0	389 <sup>6</sup>	4.2	402	4.1
Ireland	523	3.5	—	†	—	†
Israel	( <sup>8</sup> )	4.5	—	†	—	†
Italy	(?)	†	503	3.7	507	3.1
Japan	567	1.9	565	1.6	568	2.1
Kazakhstan	—	†	—	†	549 <sup>4</sup>	7.1
Korea	581	1.8	—	†	—	†
Kuwait	( <sup>8</sup> )	†	—	†	316 <sup>9</sup>	3.6
Latvia (LSS) <sup>10</sup>	499 <sup>1,4</sup>	4.6	533 <sup>4</sup>	3.1	537 <sup>4</sup>	2.3
Lithuania	—	†	534 <sup>4</sup>	2.8	530 <sup>4</sup>	2.4
Mexico	( <sup>11</sup> )	†	—	†	—	†
Moldova, Republic of	—	†	504	4.9	—	†
Morocco	—	†	347	5.1	341	4.7
Netherlands	549 <sup>1</sup>	3.0	540 <sup>2</sup>	2.1	535 <sup>12</sup>	2.1
New Zealand <sup>13</sup>	469	4.4	496	2.1	492	2.3
Norway	476	3.0	451	2.3	473	2.5
Philippines	—	†	358	7.9	—	†
Portugal	442	3.9	—	†	—	†
Qatar	—	†	—	†	296	1.0
Russian Federation	—	†	532 <sup>6</sup>	4.7	544	4.9
Scotland	493 <sup>2,6</sup>	4.2	490 <sup>2</sup>	3.3	494 <sup>2</sup>	2.2
Singapore	590	4.5	594	5.6	599	3.7
Slovak Republic	—	†	—	†	496	4.5
Slovenia	462	3.1	479	2.6	502	1.8
Sweden	—	†	—	†	503	2.5
Thailand	( <sup>8</sup> )	4.8	—	†	—	†
Tunisia	—	†	339	4.7	327	4.5

See notes at end of table.



**Table A-3. Average TIMSS mathematics scores of fourth-grade students on combined mathematics scale, by jurisdiction: 1995, 2003, and 2007—Continued**

Jurisdiction	1995		2003		2007	
	Average score	Standard error	Average score	Standard error	Average score	Standard error
TIMSS scale average	📍 500	0.0	📍 500	0.0	📍 500	0.0
Ukraine	—	†	—	†	📍 469	2.9
<b>United States</b>	<b>518</b>	<b>3.0</b>	<b>518<sup>2</sup></b>	<b>2.4</b>	<b>529<sup>2,6</sup></b>	<b>2.4</b>
Yemen	—	†	—	†	📍 224	6.0

📍 Average score is higher than U.S. average score.

📍 Average score is lower than U.S. average score.

— Not available. (Data were not collected or not reported.)

† Not applicable.

<sup>1</sup> Did not satisfy sampling participation rates.

<sup>2</sup> Met guidelines for sample participation rates only after substitute schools were included.

<sup>3</sup> National Defined Population covers less than 90 percent of National Target Population (but at least 77 percent).

<sup>4</sup> National Target Population did not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS).

<sup>5</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>6</sup> National Defined Population covers 90 percent to 95 percent of National Target Population.

<sup>7</sup> Indonesia and Italy were unable to complete the steps necessary for their data to appear in TIMSS 1995 report.

<sup>8</sup> Data not shown because of unapproved sampling procedures at classroom level and failure to meet other guidelines.

<sup>9</sup> Kuwait tested the same cohort of students as other countries, but later in 2007, at the beginning of the next school year.

<sup>10</sup> Only Latvian-speaking schools (LSS) are included.

<sup>11</sup> Mexico participated in the testing portion of TIMSS, but chose not to release its results at grades 3 and 4 in the international report.

<sup>12</sup> Nearly satisfied guidelines for sample participation rates only after substitute schools were included.

<sup>13</sup> Estimates for New Zealand have been computed for students taught in English only, which represents 98 to 99 percent of the student population.

NOTE: Jurisdictions are ordered alphabetically. Combined mathematics literacy scores are reported on a scale from 0 to 1,000. The TIMSS 1995 average scores were rescaled to allow for trend analysis. For more details, see appendix A. 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.

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

**Table A-4. Average TIMSS mathematics scores of eighth-grade students on combined mathematics scale, by jurisdiction: 1995, 1999, 2003, and 2007**

Jurisdiction	1995		1999		2003		2007	
	Average score	Standard error	Average score	Standard error	Average score	Standard error	Average score	Standard error
TIMSS scale average	500	0.0	500	0.0	500	0.0	500	0.0
Algeria	—	†	—	†	—	†	387	2.1
Argentina	( <sup>1</sup> )	†	—	†	—	†	—	†
Armenia	—	†	—	†	478	3.0	499	3.5
Australia	▲ 509 <sup>2</sup>	3.7	▲ 525 <sup>3</sup>	4.8	505	4.6	496	3.9
Austria	▲ 529 <sup>4</sup>	3.1	—	†	—	†	—	†
Bahrain	—	†	—	†	401	1.7	398	1.6
Belgium (Flemish)	▲ 550 <sup>5</sup>	5.9	▲ 558 <sup>5</sup>	3.3	▲ 537	2.8	—	†
Belgium (French)	▲ 518 <sup>4</sup>	3.8	—	†	—	†	—	†
Bosnia and Herzegovina	—	†	—	†	—	†	456	2.7
Botswana	—	†	—	†	366	2.6	364	2.3
Bulgaria	▲ 527 <sup>6</sup>	5.8	511	5.9	476	4.3	464	5.0
Canada	▲ 521	2.2	▲ 531	2.5	—	†	—	†
Chile	—	†	392	4.4	387	3.3	—	†
Chinese Taipei	—	†	▲ 585	4.0	▲ 585	4.6	▲ 598	4.5
Colombia	▼ 332	5.6	—	†	—	†	380	3.6
Cyprus	▼ 468	2.2	▼ 476	1.8	▼ 459	1.7	▼ 465	1.6
Czech Republic	▲ 546	4.5	▲ 520	4.2	—	†	504	2.4
Denmark	( <sup>7</sup> )	3.1	—	†	—	†	—	†
Egypt	—	†	—	†	406	3.5	391	3.6
El Salvador	—	†	—	†	—	†	340	2.8
England	498 <sup>5,8</sup>	3.0	496 <sup>4</sup>	4.2	( <sup>9</sup> )	†	513 <sup>5</sup>	4.8
Estonia	—	†	—	†	▲ 531	3.0	—	†
Finland	—	†	▲ 520	2.7	—	†	—	†
France	▲ 530	2.8	—	†	—	†	—	†
Georgia	—	†	—	†	—	†	410 <sup>10</sup>	6.0
Germany	502 <sup>4,5,10</sup>	4.5	—	†	—	†	—	†
Ghana	—	†	—	†	▼ 276	4.7	309	4.4
Greece	( <sup>7</sup> )	3.4	—	†	—	†	—	†
Hong Kong SAR <sup>11</sup>	▲ 569	6.1	▲ 582 <sup>5</sup>	4.3	▲ 586 <sup>5</sup>	3.3	▲ 572 <sup>5</sup>	5.8
Hungary	▲ 527	3.2	▲ 532	3.7	▲ 529 <sup>12</sup>	3.2	517	3.5
Iceland	484	4.9	—	†	—	†	—	†
Indonesia	( <sup>1</sup> )	†	▼ 403	4.9	▼ 411 <sup>10</sup>	4.8	▼ 397	3.8
Iran, Islamic Republic of	▼ 418	3.9	▼ 422	3.4	▼ 411 <sup>12</sup>	2.4	▼ 403	4.1
Ireland	▲ 519	4.8	—	†	—	†	—	†
Israel	( <sup>13</sup> )	†	▼ 466 <sup>8</sup>	3.9	496 <sup>8</sup>	3.4	▼ 463 <sup>8</sup>	3.9
Italy	( <sup>1</sup> )	†	▼ 479 <sup>12</sup>	3.8	▼ 484	3.2	▼ 480	3.0
Japan	▲ 581	1.6	▲ 579	1.7	▲ 570	2.1	▲ 570	2.4
Jordan	—	†	▼ 428	3.6	▼ 424	4.1	▼ 427	4.1
Korea, Republic of	▲ 581	2.0	▲ 587	2.0	▲ 589 <sup>14</sup>	2.2	▲ 597	2.7
Kuwait	( <sup>13</sup> )	†	—	†	—	†	▼ 354 <sup>14</sup>	2.3
Latvia (LSS) <sup>15</sup>	488 <sup>10</sup>	3.6	505 <sup>10</sup>	3.4	505	3.8	—	†
Lebanon	—	†	—	†	▼ 433	3.1	▼ 449	4.0
Lithuania	▼ 472 <sup>10,12</sup>	4.1	▼ 482 <sup>10,14</sup>	4.3	502 <sup>10</sup>	2.5	506 <sup>10</sup>	2.3
Macedonia, Republic of	—	†	▼ 447	4.2	▼ 435 <sup>8</sup>	3.5	—	†
Malaysia	—	†	▲ 519	4.4	508	4.1	▼ 474	5.0
Malta	—	†	—	†	—	†	▼ 488	1.2
Mexico	( <sup>16</sup> )	†	—	†	—	†	—	†
Moldova, Republic of	—	†	▼ 469	3.9	▼ 460	4.0	—	†
Morocco	—	†	▼ 337	2.6	▼ 387 <sup>2,10,17</sup>	2.5	( <sup>18</sup> )	†
Netherlands	▲ 529 <sup>4</sup>	6.1	▲ 540 <sup>5</sup>	7.1	▲ 536 <sup>5</sup>	3.8	—	†
New Zealand	501	4.7	491	5.2	494	5.3	—	†
Norway	498	2.2	—	†	▼ 461	2.5	▼ 469	2.0

See notes at end of table.

**Table A-4. Average TIMSS mathematics scores of eighth-grade students on combined mathematics scale, by jurisdiction: 1995, 1999, 2003, and 2007—Continued**

Jurisdiction	1995		1999		2003		2007	
	Average score	Standard error	Average score	Standard error	Average score	Standard error	Average score	Standard error
TIMSS scale average	500	0.0	500	0.0	500	0.0	▼ 500	0.0
Oman	—	†	—	†	—	†	▼ 372	3.4
Palestinian National Authority	—	†	—	†	▼ 390	3.1	▼ 367	3.5
Philippines	( <sup>19</sup> )	†	▼ 345	6.0	▼ 378	5.2	—	†
Portugal	▼ 451	3.0	—	†	—	†	—	†
Qatar	—	†	—	†	—	†	▼ 307	1.4
Romania	▼ 474	4.6	▼ 472	5.8	▼ 475	4.8	▼ 461	4.1
Russian Federation	▲ 524 <sup>12</sup>	5.3	▲ 526	5.9	508	3.7	512	4.1
Saudi Arabia	—	†	—	†	▼ 332	4.6	▼ 329 <sup>17</sup>	2.9
Scotland	493 <sup>2</sup>	5.7	—	†	498 <sup>5</sup>	3.7	▼ 487 <sup>5</sup>	3.7
Serbia	—	†	—	†	▼ 477 <sup>10</sup>	2.6	▼ 486 <sup>10,12</sup>	3.3
Singapore	▲ 609	4.0	▲ 604	6.3	▲ 605	3.6	▲ 593	3.8
Slovak Republic	▲ 534	3.1	▲ 534	4.0	508	3.3	—	†
Slovenia	494 <sup>4</sup>	2.9	530 <sup>3</sup>	2.8	▼ 493	2.2	▼ 501	2.1
South Africa	( <sup>13</sup> )	†	▼ 275	6.8	▼ 264	5.5	—	†
Spain	483	2.3	—	†	—	†	—	†
Sweden	▲ 540	4.3	—	†	499	2.6	▼ 491	2.3
Switzerland	▲ 534 <sup>10</sup>	2.7	—	†	—	†	—	†
Syrian Arab Republic	—	†	—	†	—	†	▼ 395	3.8
Thailand	( <sup>7</sup> )	†	▼ 467	5.1	—	†	▼ 441	5.0
Tunisia	—	†	▼ 448	2.4	▼ 410	2.2	▼ 420	2.4
Turkey	—	†	▼ 429	4.3	—	†	▼ 432 <sup>17</sup>	4.8
Ukraine	—	†	—	†	—	†	▼ 462	3.6
<b>United States</b>	<b>492<sup>5</sup></b>	<b>4.7</b>	<b>502</b>	<b>4.0</b>	<b>504<sup>2</sup></b>	<b>3.3</b>	<b>508<sup>5,12</sup></b>	<b>2.8</b>

▲ Average score is higher than U.S. average score.

▼ Average score is lower than U.S. average score.

— Not available. (Data were not collected or not reported.)

† Not applicable.

<sup>1</sup> Argentina, Italy, and Indonesia were unable to complete the steps necessary for their data to appear in TIMSS 1995 report.

<sup>2</sup> Nearly satisfied guidelines for sample participation rates only after substitute schools were included.

<sup>3</sup> Because of national-level changes in the starting age/date for school, data cannot be compared with other years.

<sup>4</sup> Sampling issues identified by the Trends in International Mathematics and Science Study (TIMSS). See TIMSS 1995 report for details.

<sup>5</sup> Met guidelines for sample participation rates only after substitute schools were included.

<sup>6</sup> Did not satisfy guidelines for sample participation rates.

<sup>7</sup> Data not shown because of unapproved sampling procedures at classroom level.

<sup>8</sup> National Defined Population covers less than 90 percent of National Target Population (but at least 77 percent).

<sup>9</sup> England collected data in 2003, but due to problems with meeting the minimum sampling requirements its eighth-grade data are not reported.

<sup>10</sup> National Target Population did not include all of the International Target Population defined by TIMSS.

<sup>11</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>12</sup> National Defined Population covers 90 percent to 95 percent of National Target Population.

<sup>13</sup> Data not shown because of unapproved sampling procedures at classroom level and failure to meet other guidelines.

<sup>14</sup> The same cohort of students was tested as other countries, but later in the same year, at the beginning of the next school year.

<sup>15</sup> Only Latvian-speaking schools (LSS) are included.

<sup>16</sup> Mexico participated in the testing portion of TIMSS, but chose not to release its results at grades 7 and 8 in the international report.

<sup>17</sup> Data not comparable with previous cycles.

<sup>18</sup> Morocco failed to meet the required school participation rates in 2007 because of a procedural difficulty with some schools.

<sup>19</sup> Data for the Philippines are not shown because characteristics of its school sample are not completely known.

NOTE: Jurisdictions are ordered alphabetically. Combined mathematics literacy scores are reported on a scale from 0 to 1,000. The TIMSS 1995 average scores were rescaled to allow for trend analysis. For more details, see appendix A. 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.

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

**Table A-5. Average PISA scores of 15-year-old students on combined mathematics literacy scale, by jurisdiction: 2000, 2003, and 2006**

Jurisdiction	2000 <sup>1,2</sup>		2003		2006	
	Average score	Standard error	Average score	Standard error	Average score	Standard error
OECD average	500	0.7	▲ 500	0.6	▲ 498	0.5
<b>OECD</b>						
Australia	▲ 533	3.5	▲ 524	2.1	▲ 520	2.2
Austria	▲ 515	2.5	▲ 506	3.3	▲ 505	3.7
Belgium	▲ 520	3.9	▲ 529	2.3	▲ 520	3.0
Canada	▲ 533	1.4	▲ 532	1.8	▲ 527	2.0
Czech Republic	498	2.8	▲ 516	3.5	▲ 510	3.6
Denmark	▲ 514	2.4	▲ 514	2.7	▲ 513	2.6
Finland	▲ 536	2.2	▲ 544	1.9	▲ 548	2.3
France	▲ 517	2.7	▲ 511	2.5	▲ 496	3.2
Germany	490	2.5	▲ 503	3.3	▲ 504	3.9
Greece	▼ 447	5.6	▼ 445	3.9	▼ 459	3.0
Hungary	488	4.0	490	2.8	▲ 491	2.9
Iceland	▲ 514	2.3	▲ 515	1.4	▲ 506	1.8
Ireland	503	2.7	▲ 503	2.4	▲ 501	2.8
Italy	▼ 457	2.9	▼ 466	3.1	▼ 462	2.3
Japan	▲ 557	5.5	▲ 534	4.0	▲ 523	3.3
Korea, Republic of	▲ 547	2.8	▲ 542	3.2	▲ 547	3.8
Luxembourg	▼ 446	2.0	▲ 493	1.0	▲ 490	1.1
Mexico	▼ 387	3.4	▼ 385	3.6	▼ 406	2.9
Netherlands	( <sup>3</sup> )	†	▲ 538	3.1	▲ 531	2.6
New Zealand	▲ 537	3.1	▲ 523	2.3	▲ 522	2.4
Norway	499	2.8	▲ 495	2.4	▲ 490	2.6
Poland	▼ 470	5.5	490	2.5	▲ 495	2.4
Portugal	▼ 454	4.1	▼ 466	3.4	466	3.1
Slovak Republic	—	†	▲ 498	3.3	▲ 492	2.8
Spain	▼ 476	3.1	485	2.4	480	2.3
Sweden	▲ 510	2.5	▲ 509	2.6	▲ 502	2.4
Switzerland	▲ 529	4.4	▲ 527	3.4	▲ 530	3.2
Turkey	—	†	▼ 423	6.7	▼ 424	4.9
United Kingdom	▲ 529	2.5	( <sup>4</sup> )	†	▲ 495	2.1
<b>United States</b>	<b>493</b>	<b>7.6</b>	<b>483</b>	<b>2.9</b>	<b>474</b>	<b>4.0</b>
<b>Non-OECD</b>						
Albania	▼ 381 <sup>5</sup>	3.1	—	†	—	†
Argentina	▼ 388 <sup>5</sup>	9.4	—	†	▼ 381	6.2
Azerbaijan	—	†	—	†	476	2.3
Brazil	▼ 334	3.7	▼ 356	4.8	▼ 370	2.9
Bulgaria	▼ 430 <sup>5</sup>	5.7	—	†	▼ 413	6.1
Chile	▼ 384 <sup>5</sup>	3.7	—	†	▼ 411	4.6
Chinese Taipei	—	†	—	†	▲ 549	4.1
Colombia	—	†	—	†	▼ 370	3.8
Croatia	—	†	—	†	467	2.4
Estonia	—	†	—	†	▲ 515	2.7
Hong Kong–China	▲ 560 <sup>5</sup>	3.3	▲ 550	4.5	▲ 547	2.7
Indonesia	▼ 367 <sup>5</sup>	4.5	▼ 360	3.9	▼ 391	5.6
Israel	▼ 433 <sup>5</sup>	9.3	—	†	▼ 442	4.3
Jordan	—	†	—	†	▼ 384	3.3
Kyrgyz Republic	—	†	—	†	▼ 311	3.4
Latvia	▼ 463	4.5	483	3.7	▲ 486	3.0
Liechtenstein	▲ 514	7.0	▲ 536	4.1	▲ 525	4.2
Lithuania	—	†	—	†	▲ 486	2.9
Macao–China	—	†	▲ 527	2.9	▲ 525	1.3
Macedonia, FYR	▼ 381	2.7	—	†	—	†

See notes at end of table.

**Table A-5. Average PISA scores of 15-year-old students on combined mathematics literacy scale, by jurisdiction: 2000, 2003, and 2006—Continued**

Jurisdiction	2000 <sup>1,2</sup>		2003		2006	
	Average score	Standard error	Average score	Standard error	Average score	Standard error
OECD average	500	0.7	▲ 500	0.6	▲ 498	0.5
<b>Non-OECD—Continued</b>						
Montenegro, Republic of <sup>6</sup>	—	†	—	†	▼ 399	1.4
Peru	▼ 292	4.4	—	†	—	†
Qatar	—	†	—	†	▼ 318	1.0
Romania	▼ 426	4.3	—	†	▼ 415	4.2
Russian Federation	478	5.5	▼ 468	4.2	476	3.9
Serbia, Republic of <sup>6</sup>	—	†	▼ 437	3.8	▼ 435	3.5
Slovenia	—	†	—	†	▲ 504	1.0
Thailand	▼ 432 <sup>5</sup>	3.6	▼ 417	3.0	▼ 417	2.3
Tunisia	—	†	▼ 359	2.5	▼ 365	4.0
Uruguay	—	†	▼ 422	3.3	▼ 427	2.6

▲ Average score is higher than U.S. average score.

▼ Average score is lower than U.S. average score.

— Not available. (Data were not collected or not reported.)

† Not applicable.

<sup>1</sup> Tests for significance have been updated from what was presented in the Program for International Student Assessment 2000 U.S. report (Lemke et al. 2001). Some country differences reported here were not reported as statistically significant in the PISA 2000 U.S. report. In that report, a Bonferroni adjustment was used in all multiple comparisons of countries. This was not the case when PISA 2003 and 2006 data were analyzed and reported, thus PISA 2000 tests have been updated in this report to keep consistency across years. For more details, see appendix A.

<sup>2</sup> The PISA mathematics framework was revised in 2003. Because of changes in the framework, it is not possible to compare mathematics learning outcomes from PISA 2000 with those from PISA 2003 and PISA 2006. The line dividing PISA 2000 results from those of PISA 2003 and PISA 2006 indicates that these data are not comparable. For more details, see OECD 2007 (available at <http://www.oecd.org/dataoecd/30/17/39703267.pdf>).

<sup>3</sup> Because of technical problems with its sample, 2000 data for The Netherlands have been omitted from PISA reports.

<sup>4</sup> Because of low response rates, 2003 data for the United Kingdom have been omitted from PISA reports.

<sup>5</sup> Did not participate with other jurisdictions in PISA in 2000; administered PISA in 2001.

<sup>6</sup> The Republics of Montenegro and Serbia were a united jurisdiction for the PISA 2003 assessment.

NOTE: The OECD average is the average of the national averages of the OECD member jurisdictions. Because PISA is principally an OECD study, the results for non-OECD jurisdictions are displayed separately from those of the OECD jurisdictions and are not included in the OECD average. Jurisdictions are ordered alphabetically. Combined mathematics literacy scores are reported on a scale from 0 to 1,000. 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.

SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2000, 2003, and 2006.

**Table A-6. Average TIMSS science scores of fourth-grade students on combined science scale, by jurisdiction: 1995, 2003, and 2007**

Jurisdiction	1995		2003		2007	
	Average score	Standard error	Average score	Standard error	Average score	Standard error
TIMSS scale average	500	0.0	500	0.0	500	0.0
Algeria	—	†	—	†	354	6.0
Armenia	—	†	437	4.3	484	5.7
Australia	541 <sup>1</sup>	3.6	521 <sup>2</sup>	4.2	527	3.3
Austria	538 <sup>1</sup>	3.6	—	†	526	2.5
Belgium (Flemish)	—	†	518	1.8	—	†
Canada	525	3.1	—	†	—	†
Chinese Taipei	—	†	551	1.7	557	2.0
Colombia	—	†	—	†	400	5.4
Cyprus	450	3.2	480	2.4	—	†
Czech Republic	532	3.0	—	†	515	3.1
Denmark	—	†	—	†	517 <sup>2</sup>	2.9
El Salvador	—	†	—	†	390	3.4
England	528 <sup>2,3</sup>	3.1	540 <sup>2</sup>	3.6	542	2.9
Georgia	—	†	—	†	418 <sup>4</sup>	4.6
Germany	—	†	—	†	528	2.4
Greece	473	5.3	—	†	—	†
Hong Kong SAR <sup>5</sup>	508	3.3	542 <sup>2</sup>	3.1	554	3.5
Hungary	508	3.4	530 <sup>6</sup>	3.0	536	3.3
Iceland	479	4.3	—	†	—	†
Indonesia	( <sup>7</sup> )	†	—	†	—	†
Iran, Islamic Republic of	380	4.6	414 <sup>6</sup>	4.1	436	4.3
Ireland	515	3.5	—	†	—	†
Israel	( <sup>8</sup> )	4.9	—	†	—	†
Italy	( <sup>7</sup> )	†	516	3.8	535	3.2
Japan	553	1.8	543	1.5	548	2.1
Kazakhstan	—	†	—	†	533 <sup>4</sup>	5.6
Korea	576	2.1	—	†	—	†
Kuwait	( <sup>8</sup> )	4.5	—	†	348 <sup>9</sup>	4.4
Latvia (LSS) <sup>10</sup>	486 <sup>1,4</sup>	4.9	530 <sup>4</sup>	2.8	542 <sup>4</sup>	2.3
Lithuania	—	†	512 <sup>4</sup>	2.6	514 <sup>4</sup>	2.4
Mexico	( <sup>11</sup> )	†	—	†	—	†
Moldova, Republic of	—	†	496	4.6	—	†
Morocco	—	†	304	6.7	297	5.9
Netherlands	530 <sup>1</sup>	3.2	525 <sup>2</sup>	2.0	523 <sup>12</sup>	2.6
New Zealand <sup>13</sup>	505	5.3	523	2.3	504	2.6
Norway	504	3.7	466	2.6	477	3.5
Philippines	—	†	332	9.4	—	†
Portugal	452	4.1	—	†	—	†
Qatar	—	†	—	†	294	2.6
Russian Federation	—	†	526 <sup>6</sup>	5.2	546	4.8
Scotland	514 <sup>2,6</sup>	4.5	502 <sup>2</sup>	2.9	500 <sup>2</sup>	2.3
Singapore	523	4.8	565	5.5	587	4.1
Slovak Republic	—	†	—	†	526	4.8
Slovenia	464	3.1	490	2.5	518	1.9
Sweden	—	†	—	†	525	2.9
Thailand	( <sup>8</sup> )	5.6	—	†	—	†
Tunisia	—	†	314	5.7	318	5.9

See notes at end of table.

**Table A-6. Average TIMSS science scores of fourth-grade students on combined science scale, by jurisdiction: 1995, 2003, and 2007—Continued**

Jurisdiction	1995		2003		2007	
	Average score	Standard error	Average score	Standard error	Average score	Standard error
TIMSS scale average	👉 500	0.0	👉 500	0.0	👉 500	0.0
Ukraine	—	†	—	†	👉 474	3.1
<b>United States</b>	<b>542</b>	<b>3.3</b>	<b>536<sup>2</sup></b>	<b>2.5</b>	<b>539<sup>2,6</sup></b>	<b>2.7</b>
Yemen	—	†	—	†	👉 197	7.2

👈 Average score is higher than U.S. average score.

👉 Average score is lower than U.S. average score.

— Not available. (Data were not collected or not reported.)

† Not applicable.

<sup>1</sup> Did not satisfy guidelines for sample participation rates.

<sup>2</sup> Met guidelines for sample participation rates only after substitute schools were included.

<sup>3</sup> National Defined Population covers less than 90 percent of National Target Population (but at least 77 percent).

<sup>4</sup> National Target Population did not include all of the International Target Population defined by the Trends in International Mathematics and Science Study (TIMSS).

<sup>5</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>6</sup> National Defined Population covers 90 percent to 95 percent of National Target Population.

<sup>7</sup> Indonesia and Italy were unable to complete the steps necessary for their data to appear in TIMSS 1995 report.

<sup>8</sup> Data not shown because of unapproved sampling procedures at classroom level and failure to meet other guidelines.

<sup>9</sup> Kuwait tested the same cohort of students as other countries, but later in 2007, at the beginning of the next school year.

<sup>10</sup> Only Latvian-speaking schools (LSS) were included.

<sup>11</sup> Mexico participated in the testing portion of TIMSS, but chose not to release its results at grades 3 and 4 in the international report.

<sup>12</sup> Nearly satisfied guidelines for sample participation rates only after substitute schools were included.

<sup>13</sup> Estimates for New Zealand have been computed for students taught in English only, which represents 98 to 99 percent of the student population.

NOTE: Jurisdictions are ordered alphabetically. Combined mathematics literacy scores are reported on a scale from 0 to 1,000. The TIMSS 1995 average scores were rescaled to allow for trend analysis. For more details, see appendix A. 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.

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

**Table A-7. Average TIMSS science scores of eighth-grade students on combined science scale, by jurisdiction: 1995, 1999, 2003, and 2007**

Jurisdiction	1995		1999		2003		2007	
	Average score	Standard error	Average score	Standard error	Average score	Standard error	Average score	Standard error
TIMSS scale average	500	0.0	500	0.0	500	0.0	500	0.0
Algeria	—	†	—	†	—	†	408	1.7
Argentina	( <sup>1</sup> )	†	—	†	—	†	—	†
Armenia	—	†	—	†	461	3.5	488	5.8
Australia	514 <sup>2</sup>	3.9	540 <sup>3</sup>	4.4	527	3.8	515	3.6
Austria	539 <sup>4</sup>	3.8	—	†	—	†	—	†
Bahrain	—	†	—	†	438	1.8	467	1.7
Belgium (Flemish)	533 <sup>5</sup>	6.4	535 <sup>5</sup>	3.1	516	2.5	—	†
Belgium (French)	466 <sup>4</sup>	3.8	—	†	—	†	—	†
Bosnia and Herzegovina	—	†	—	†	—	†	466	2.8
Botswana	—	†	—	†	365	2.8	355	3.1
Bulgaria	545 <sup>4</sup>	5.2	518	5.4	479	5.2	470 <sup>6</sup>	5.9
Canada	514	2.6	533	2.1	—	†	—	†
Chile	—	†	420	3.7	413	2.9	—	†
Chinese Taipei	—	†	569	4.4	571	3.5	561	3.7
Colombia	365	6.2	—	†	—	†	417	3.5
Cyprus	452	2.1	460	2.4	441	2.0	452	2.0
Czech Republic	555	4.5	539	4.2	—	†	539	1.9
Denmark	( <sup>7</sup> )	†	—	†	—	†	—	†
Egypt	—	†	—	†	421	3.9	408	3.6
El Salvador	—	†	—	†	—	†	387	2.9
England	533 <sup>5,8</sup>	†	538 <sup>5</sup>	4.8	( <sup>9</sup> )	†	542 <sup>5</sup>	4.5
Estonia	—	†	—	†	552	2.5	—	†
Finland	—	†	535	3.5	—	†	—	†
France	488	3.2	—	†	—	†	—	†
Georgia	—	†	—	†	—	†	421 <sup>10</sup>	4.8
Germany	518 <sup>4,5,10</sup>	5.5	—	†	—	†	—	†
Ghana	—	†	—	†	255	5.9	303	5.4
Greece	( <sup>7</sup> )	†	—	†	—	†	—	†
Hong Kong SAR <sup>11</sup>	510	5.8	530 <sup>5</sup>	3.7	556 <sup>5,12</sup>	3.0	530 <sup>5</sup>	4.9
Hungary	537	3.1	552	3.7	543	2.8	539	2.9
Iceland	484	5.8	—	†	—	†	—	†
Indonesia	( <sup>1</sup> )	†	435	4.5	420 <sup>10</sup>	4.1	427	3.4
Iran, Islamic Republic of	463	3.6	448	3.8	453 <sup>12</sup>	2.3	459	3.6
Ireland	518	5.1	—	†	—	†	—	†
Israel	( <sup>13</sup> )	†	468 <sup>8</sup>	4.9	488 <sup>8</sup>	3.1	468 <sup>8</sup>	4.3
Italy	( <sup>1</sup> )	†	493 <sup>12</sup>	3.9	491	3.1	495	2.8
Japan	554	1.8	550	2.2	552	1.7	554	1.9
Jordan	—	†	450	3.8	475	3.8	482	4.0
Korea, Republic of	546	2.0	549	2.6	558 <sup>14</sup>	1.6	553	2.0
Kuwait	( <sup>13</sup> )	†	—	†	—	†	418 <sup>14</sup>	2.8
Latvia (LSS) <sup>15</sup>	476 <sup>10</sup>	3.3	503 <sup>10</sup>	4.8	513	2.9	—	†
Lebanon	—	†	—	†	393	4.3	414	5.9
Lithuania	464 <sup>10,12</sup>	4.0	488 <sup>10,14</sup>	4.1	519 <sup>10</sup>	2.1	519 <sup>10</sup>	2.6
Macedonia, Republic of	—	†	458	5.2	449 <sup>8</sup>	3.6	—	†
Malaysia	—	†	492	4.4	510	3.7	471	6.0
Malta	—	†	—	†	—	†	457	1.4
Mexico	( <sup>16</sup> )	†	—	†	—	†	—	†
Moldova, Republic of	—	†	459	4.0	472	3.4	—	†
Morocco	—	†	323	4.3	396 <sup>2,6,10</sup>	2.5	( <sup>17</sup> )	†
Netherlands	541 <sup>9</sup>	6.0	545 <sup>5</sup>	6.9	536 <sup>5</sup>	3.1	—	†
New Zealand	511	4.9	510	4.9	520	5.0	—	†
Norway	514	2.4	—	†	494	2.2	487	2.2

See notes at end of table.



**Table A-7. Average TIMSS science scores of eighth-grade students on combined science scale, by jurisdiction: 1995, 1999, 2003, and 2007—Continued**

Jurisdiction	1995		1999		2003		2007	
	Average score	Standard error	Average score	Standard error	Average score	Standard error	Average score	Standard error
TIMSS scale average	500	0.0	500	0.0	500	0.0	500	0.0
Oman	—	†	—	†	—	†	423	3.0
Palestinian National Authority	—	†	—	†	435	3.2	404	3.5
Philippines	( <sup>18</sup> )	†	345	7.5	377	5.8	—	†
Portugal	473	3.1	—	†	—	†	—	†
Qatar	—	†	—	†	—	†	319	1.7
Romania	471	5.1	472	5.8	470	4.9	462	3.9
Russian Federation	523 <sup>12</sup>	4.5	529	6.4	514	3.7	530	3.9
Saudi Arabia	—	†	—	†	398	4.0	403 <sup>6</sup>	2.4
Scotland	501 <sup>2</sup>	5.6	—	†	512 <sup>5</sup>	3.4	496 <sup>5</sup>	3.4
Serbia	—	†	—	†	468 <sup>10</sup>	2.5	470 <sup>10,12</sup>	3.2
Singapore	580	5.5	568	8.0	578	4.3	567	4.4
Slovak Republic	532	3.3	535	3.3	517	3.2	—	†
Slovenia	514	2.7	533 <sup>3</sup>	3.2	520	1.8	538	2.2
South Africa	( <sup>13</sup> )	†	243	7.9	244	6.7	—	†
Spain	504	2.3	—	†	—	†	—	†
Sweden	553	4.4	—	†	524	2.7	511	2.6
Switzerland	509 <sup>10</sup>	2.8	—	†	—	†	—	†
Syrian Arab Republic	—	†	—	†	—	†	452	2.9
Thailand	( <sup>7</sup> )	†	482	4.0	—	†	471	4.3
Tunisia	—	†	430	3.4	404	2.1	445	2.1
Turkey	—	†	433	4.3	—	†	454 <sup>6</sup>	3.7
Ukraine	—	†	—	†	—	†	485	3.5
<b>United States</b>	<b>513<sup>5</sup></b>	<b>5.6</b>	<b>515</b>	<b>4.6</b>	<b>527<sup>2</sup></b>	<b>3.1</b>	<b>520<sup>5,12</sup></b>	<b>2.9</b>

▲ Average score is higher than U.S. average score.

▼ Average score is lower than U.S. average score.

— Not available. (Data were not collected or not reported.)

† Not applicable.

<sup>1</sup> Argentina, Italy, and Indonesia were unable to complete the steps necessary for their data to appear in TIMSS 1995 report.

<sup>2</sup> Nearly satisfied guidelines for sample participation rates only after substitute schools were included.

<sup>3</sup> Because of national-level changes in the starting age/date for school, data cannot be compared with other years.

<sup>4</sup> Sampling issue identified by the Trends in International Mathematics and Science Study (TIMSS). See TIMSS 1995 report for more details.

<sup>5</sup> Met guidelines for sample participation rates only after substitute schools were included.

<sup>6</sup> Data not comparable with previous cycles.

<sup>7</sup> Data not shown because of unapproved sampling procedures at classroom level.

<sup>8</sup> National Defined Population covers less than 90 percent of National Target Population (but at least 77 percent).

<sup>9</sup> England collected data in 2003, but due to problems with meeting the minimum sampling requirements its eighth-grade data are not reported.

<sup>10</sup> National Target Population did not include all of the International Target Population defined by TIMSS.

<sup>11</sup> Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

<sup>12</sup> National Defined Population covers 90 percent to 95 percent of National Target Population.

<sup>13</sup> Data not shown because of unapproved sampling procedures at classroom level and failure to meet other guidelines.

<sup>14</sup> Tested the same cohort of students, but later, at the beginning of the next school year.

<sup>15</sup> Only Latvian-speaking schools (LSS) were included.

<sup>16</sup> Mexico participated in the testing portion of TIMSS, but chose not to release its results at grades 7 and 8 in the international report.

<sup>17</sup> Morocco failed to meet the required school participation rates in 2007 because of a procedural difficulty with some schools.

<sup>18</sup> Data for the Philippines are not shown because characteristics of its school sample are not completely known.

NOTE: Jurisdictions are ordered alphabetically. Combined mathematics literacy scores are reported on a scale from 0 to 1,000. The TIMSS 1995 average scores were rescaled to allow for trend analysis. For more details, see appendix A. 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.

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

**Table A-8. Average PISA scores of 15-year-old students on combined science literacy scale, by jurisdiction: 2000, 2003, and 2006**

Jurisdiction	2000 <sup>1</sup>		2003		2006 <sup>2</sup>	
	Average score	Standard error	Average score	Standard error	Average score	Standard error
OECD average	500	0.7	▲ 500	0.6	▲ 500	0.5
<b>OECD</b>						
Australia	▲ 528	3.5	▲ 525	2.1	▲ 527	2.3
Austria	▲ 519	2.6	▲ 491	3.4	▲ 511	3.9
Belgium	496	4.3	▲ 509	2.5	▲ 510	2.5
Canada	▲ 529	1.6	▲ 519	2.0	▲ 534	2.0
Czech Republic	511	2.4	▲ 523	3.4	▲ 513	3.5
Denmark	▼ 481	2.8	▲ 475	3.0	496	3.1
Finland	▲ 538	2.5	▲ 548	1.9	▲ 563	2.0
France	500	3.2	▲ 511	3.0	495	3.4
Germany	487	2.4	▲ 502	3.6	▲ 516	3.8
Greece	▼ 461	4.9	▼ 481	3.8	▼ 473	3.2
Hungary	496	4.2	503	2.8	▲ 504	2.7
Iceland	496	2.2	▲ 495	1.5	491	1.6
Ireland	513	3.2	▲ 505	2.7	▲ 508	3.2
Italy	▼ 478	3.1	▼ 486	3.1	▼ 475	2.0
Japan	▲ 550	5.5	▲ 548	4.1	▲ 531	3.4
Korea, Republic of	▲ 552	2.7	▲ 538	3.5	▲ 522	3.4
Luxembourg	▼ 443	2.3	▲ 483	1.5	486	1.1
Mexico	▼ 422	3.2	▼ 405	3.5	▼ 410	2.7
Netherlands	( <sup>3</sup> )	†	▲ 524	3.1	▲ 525	2.7
New Zealand	▲ 528	2.4	▲ 521	2.4	▲ 530	2.7
Norway	500	2.8	▲ 484	2.9	487	3.1
Poland	483	5.1	498	2.9	498	2.3
Portugal	▼ 459	4.0	▼ 468	3.5	▼ 474	3.0
Slovak Republic	—	†	▲ 495	3.7	488	2.6
Spain	491	3.0	487	2.6	488	2.6
Sweden	512	2.5	▲ 506	2.7	▲ 503	2.4
Switzerland	496	4.4	▲ 513	3.7	▲ 512	3.2
Turkey	—	†	▼ 434	5.9	▼ 424	3.8
United Kingdom	▲ 532	2.7	( <sup>4</sup> )	†	▲ 515	2.3
<b>United States</b>	<b>499</b>	<b>7.3</b>	<b>491</b>	<b>3.1</b>	<b>489</b>	<b>4.2</b>
<b>Non-OECD</b>						
Albania	▼ 376 <sup>5</sup>	2.9	—	†	—	†
Argentina	▼ 396 <sup>5</sup>	8.6	—	†	▼ 391	6.1
Azerbaijan	—	†	—	†	▼ 382	2.8
Brazil	▼ 375	3.3	390	4.3	▼ 390	2.8
Bulgaria	▼ 448 <sup>5</sup>	4.6	—	†	▼ 434	6.1
Chile	▼ 415 <sup>5</sup>	3.4	—	†	▼ 438	4.3
Chinese Taipei	—	†	—	†	▲ 532	3.6
Colombia	—	†	—	†	▼ 388	3.4
Croatia	—	†	—	†	493	2.4
Estonia	—	†	—	†	▲ 531	2.5
Hong Kong–China	▲ 541 <sup>5</sup>	3.0	▲ 539	4.3	▲ 542	2.5
Indonesia	▼ 393 <sup>5</sup>	3.9	▼ 395	3.2	▼ 393	5.7
Israel	▼ 434 <sup>5</sup>	9.0	—	†	▼ 454	3.7
Jordan	—	†	—	†	▼ 422	2.8
Kyrgyz Republic	—	†	—	†	▼ 322	2.9
Latvia	▼ 460	5.6	489	3.9	490	3.0
Liechtenstein	▼ 476	7.1	▲ 525	4.3	▲ 522	4.1
Lithuania	—	†	—	†	488	2.8
Macao–China	—	†	▲ 525	3.0	▲ 511	1.1
Macedonia, FYR	▼ 401	2.1	—	†	—	†

See notes at end of table.

**Table A-8. Average PISA scores of 15-year-old students on combined science literacy scale, by jurisdiction: 2000, 2003, and 2006—Continued**

Jurisdiction	2000 <sup>1</sup>		2003		2006 <sup>2</sup>	
	Average score	Standard error	Average score	Standard error	Average score	Standard error
OECD average	500	0.7	▲ 500	0.6	▲ 500	0.5
<b>Non-OECD—Continued</b>						
Montenegro, Republic of <sup>6</sup>	—	†	▼ 436	3.5	▼ 412	1.1
Peru	▼ 333	4.0	—	†	—	†
Qatar	—	†	—	†	▼ 349	0.9
Romania	▼ 441	3.4	—	†	▼ 418	4.2
Russian Federation	▼ 460	4.7	▼ 489	4.1	479	3.7
Serbia, Republic of <sup>6</sup>	—	†	▼ 436	3.5	▼ 436	3.0
Slovenia	—	†	—	†	▲ 519	1.1
Thailand	▼ 436 <sup>5</sup>	3.1	▼ 429	2.7	▼ 421	2.1
Tunisia	—	†	▼ 385	2.6	▼ 386	3.0
Uruguay	—	†	▼ 438	2.9	▼ 428	2.7

▲ Average score is higher than U.S. average score.

▼ Average score is lower than U.S. average score.

— Not available. (Data were not collected or not reported.)

† Not applicable.

<sup>1</sup> Tests for significance have been updated from what was presented in the Program for International Student Assessment 2000 U.S. report (Lemke et al. 2001). Some country differences reported here were not reported as statistically significant in the PISA 2000 U.S. report. In that report, a Bonferroni adjustment was used in all multiple comparisons of countries. This was not the case when PISA 2003 and 2006 data were analyzed and reported, thus PISA 2000 tests have been updated in this report to keep consistency across years. For more details, see appendix A.

<sup>2</sup> The science framework was revised for PISA 2006. Because of changes in the framework, it is not possible to compare science learning outcomes from PISA 2006 with those of earlier PISA assessments. The line dividing PISA 2006 results from those of PISA 2000 and PISA 2003 indicates that these data are not comparable. For more details, see OECD 2007 (available at <http://www.oecd.org/dataoecd/30/17/39703267.pdf>).

<sup>3</sup> Because of technical problems with its sample, 2000 data for The Netherlands have been omitted from PISA reports.

<sup>4</sup> Because of low response rates, 2003 data for the United Kingdom have been omitted from PISA reports.

<sup>5</sup> Did not participate with other jurisdictions in PISA in 2000; administered PISA in 2001.

<sup>6</sup> The Republics of Montenegro and Serbia were a united jurisdiction for the PISA 2003 assessment.

NOTE: The OECD average is the average of the national averages of the OECD member jurisdictions. Because PISA is principally an OECD study, the results for non-OECD jurisdictions are displayed separately from those of the OECD jurisdictions and are not included in the OECD average. Jurisdictions are ordered alphabetically. Combined science literacy scores are reported on a scale from 0 to 1,000. 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. SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2000, 2003, and 2006.

---

*This page intentionally left blank.*

---

# Appendix A: Technical Notes

# Appendix A: Technical Notes

These technical notes describe the various issues that are important to keep in mind when interpreting sampling data and the various procedures governing the collection and analysis of assessment data. For detailed information about NCES datasets, see <http://nces.ed.gov/surveys/>.

## A.1 Limitations of sampled data

Estimating the achievement of the total population or subpopulations from an assessment based on a sample of the entire population requires consideration of several factors before the results become meaningful. However conscientious an organization may be in collecting assessment data from a sample of a population, there will always be the possibility of *nonsampling errors* (errors made in the collection and processing of data) and some *sampling errors* (the margin of error in estimating the achievement of the actual total population or subpopulation because the data are available from only a portion of the total population).

### Nonsampling errors

“Nonsampling error” is a term used to describe variations in the estimates that may be caused by population coverage limitations, nonresponse bias, and measurement error, as well as data collection, processing, and reporting procedures. The sources of nonsampling errors are typically problems such as unit and item nonresponse, the differences in respondents’ interpretations of the meaning of questions, response differences related to the particular time the assessment was conducted, and mistakes in data preparation. Sections A.2 through A.5 describe the international policies and procedures put in place to minimize nonsampling errors. Section A.11 describes NCES’s policy of nonresponse bias analysis.

### Sampling errors

Sampling errors occur when a discrepancy between a population characteristic and the sample estimate arises because not all members of the target population are sampled for the survey. The margin of error or the magnitude of sampling error depends on several factors, such as the amount of variation in the responses, the size and representativeness of the sample, and the size of the subgroup for which the estimate is computed. The magnitude of this margin of error is measured by what statisticians call the *standard error* of an estimate.

The standard error for each estimate in this special analysis was calculated in order to determine the “margin of error” for each estimate. An estimate with a smaller standard error provides a more reliable estimate of the true value than an estimate with a higher standard error. The standard errors for all the estimated average scores, cutpoint scores, and percentages reported in the figures and tables of the special analysis can be found on *The Condition of Education* website at <http://nces.ed.gov/programs/coe/>.

## Analysis and interpretation

Due to standard errors, caution is warranted when drawing conclusions about the achievement results estimated for one population in comparison to another or whether a time series of achievement results is increasing, decreasing, or staying about the same. Although one estimate of the achievement results may be larger than another, a statistical test may reveal that there is no measurable difference between the two estimates due to their uncertainty. Whether differences in averages (means) or cutpoint scores are statistically significant can be determined by using the standard errors of the estimates. When differences are statistically significant, the probability that the difference occurred by chance is usually small; about 5 times out of 100. For this special analysis, differences between means or cutpoint scores (including increases or decreases) are stated only when they are statistically significant. To determine whether differences reported are statistically significant, two-tailed *t* tests, at the .05 level of significance, were used. In addition, the *t* test formula for determining statistical significance was adjusted when a linking error term needed to be accounted for (see below for more on linking errors, under A.8). No multiple comparisons (Bonferroni adjustments) were used in this special analysis (see below for more on past significance tests, under A.8 and A.9).

## A.2 International requirements for sampling, data collection, and response rates

To provide valid estimates of student achievement and characteristics, the sample of students for each assessment is selected in a way that represents the full target population in each jurisdiction. The international desired population or full target population in each jurisdiction is carefully defined for each study. This international desired population includes *all* students in the target grade or age range.

All three assessments require a minimum of 150 schools to participate and have minimum student sample sizes: PIRLS and PISA require a minimum of 4,500 students, and TIMSS requires a minimum of 4,000 students per grade. To realize these target sample sizes for a nationally representative sample of students, samples are drawn with two substitute schools identified for each sampled school. Substitute schools can replace a sampled school if the sampled school refuses to participate. The specific procedures for drawing the sample and for using substitute schools differ by study and even by administration: see the appropriate assessment’s technical manual for specific details.

Each jurisdiction collects its own data, following international guidelines and specifications. International guidelines and specifications require that testing for each assessment occur within a set time period. They define

response rate targets for schools (typically 85 percent) and students (typically 85 or 80 percent) as well as minimum response rates for results to be reported. International guidelines and specifications also define maximum rates of exclusion.

Schools and students excluded from the national defined target population are referred to as the excluded population. Exclusions can occur at the school level, with entire schools being excluded (e.g., schools that are extremely small or remote), or within schools, with specific students (e.g., functionally or intellectually disabled students) or entire classrooms excluded (e.g., classrooms for non-native language speakers). See the appropriate technical manual for each study's specific policy on exclusion and student accommodations.

To ensure that testing procedures are carried out in a consistent manner, international quality monitors visit a sample of schools in every jurisdiction.

### A.3 Test development

The development of the assessment instruments is an interactive multi-step process that involves representatives of the participating jurisdictions, various expert committees, and an international consortium of contractors. Generally, the first step is to develop or revise a framework to guide the construction of the assessment. Items that fit within the framework are submitted by participating jurisdictions as well as developed by each study's international consortium of test developers. Potential items are reviewed by representatives of each jurisdiction for possible bias and relevance to the study's framework. All items are meant to reflect the national, cultural, and linguistic variety among participating jurisdictions. Approved items are field-tested, and items that perform well are identified for inclusion in an "item pool" that is used to create the actual or "main" assessment instruments.

### Design of instruments

The main assessment *instruments* consist of (1) test booklets made up of instructions and blocks or clusters of items (small sets of items from the final pool of items) and (2) questionnaires for students, schools, teachers (in PIRLS and TIMSS), and parents (in PIRLS and PISA, although the United States has not administered the parent questionnaire). The test booklets for PIRLS, PISA, and TIMSS are constructed such that no student responds to all of the items. This construction is consistent with other large-scale assessments, such as the U.S. National Assessment of Educational Progress (NAEP). To keep the testing burden to a minimum, and to ensure broad subject-matter coverage, the assessments use a "rotated block design" that includes items from different content domains and, in the case of PISA and TIMSS, different subject areas. The number of booklets

for each assessment varies as does the amount of time students have to complete each booklet they receive.

In each assessment, a portion of the items are identical to items used in the prior administration of the assessment. These items allow a study to "maintain trend" (i.e., be able to make reliable comparisons among the results of different administrations of an assessment over time) and to provide for corrections through equating, if necessary.

### Translation

Source versions of all instruments are prepared in English and, in the case of PISA, in French as well. These are translated into the primary language or languages of instruction in each participating jurisdiction. In addition, it is sometimes necessary to adapt the instrument for cultural purposes, even in nations such as the United States that use English as the primary language of instruction. (For example, British terms such as "lift" might be replaced with "elevator" in the U.S. version.) The national translation and adaptation of all instruments are reviewed and approved by the sponsoring organization of each international study to ensure that the translations and adaptations did not change the substance or intent of the question or answer choices.

### A.4 Scoring

PIRLS, PISA, and TIMSS assessment items include both multiple-choice and "open-ended" or "constructed-response" items (i.e., items that require students to write their answer in the space provided, usually in the form of a list or short sentences). To score each item, a scoring guide (or *rubric*) is created to train test-scorers and anchor the scoring in each jurisdiction. The scoring guides are carefully written and reviewed by the national research coordinators and other experts as part of the field test of items, and revised accordingly. For each test item, the scoring guide describes the intent of the question and how to score students' responses—full credit, partial credit, or no credit—for the range of possible responses. In addition, the scoring guides include real examples of students' responses accompanied by a rationale for their classification for purposes of clarity and illustration.

### A.5 Data entry and cleaning

Once all items have been scored, the results are entered into data files using a common international software and format. The software facilitates the checking and correction of data by providing various data consistency checks. The data are then generally sent to a central international data processing center (DPC) for cleaning. The DPC checks that the international data structure is followed; checks the identification system within and between files; corrects single case problems manually; and applies standard cleaning procedures to questionnaire files. Results of the data cleaning process are documented

by the DPC. This documentation is shared with the national research coordinator so that specific questions can be addressed. The national research coordinator then provides the DPC with revisions to coding or solutions for anomalies. The DPC will then compile background univariate statistics and preliminary test scores based on classical and Rasch item analyses. For more detailed information on the data entry, cleaning process, and scoring of each assessment, see the appropriate assessment's technical manual.

## A.6 Weighting and scaling

Before the data are analyzed, responses from students are assigned sampling weights to ensure that the proportion or representation of different subgroups of assessed students (e.g., public/private, census region, urban/suburban/rural, race/ethnicity) match the actual percentage of that subgroup among the school population of the target grade or age. The use of sampling weights is necessary for the computation of sound, nationally representative estimates. The basic weight assigned to a student's responses is the inverse of the probability that the student would be selected for the sample. Adjustments to weights are also made by the international consortium for various situations (such as school and student nonresponse) because data cannot be assumed to be randomly missing. (NCES may conduct a nonresponse bias analysis after these adjustments are made to see how much bias still exists, compared with the original sample frame. For more details, see A.11.)

Once these sampling weights are in place, item response theory (IRT) procedures are used to deduce the difficulty of each item, using information about how likely it is for students to get some items correct versus other items. Once the difficulty of each item is determined, the items are assigned a value on a standardized logit *scale* of item difficulty. Scaling items in this way makes it possible for the ability of groups of students to be estimated or *scored*, even though not all students were administered the same items.

### Scale scores

In order to make the estimated scores more meaningful and to facilitate their interpretation, the scores are transformed to a new scale with a mean of 500 and a standard deviation of 100. These *scale scores* are what are reported in PIRLS, PISA, and TIMSS reports and throughout this special analysis. Strictly speaking, scale scores are specific to a given assessment and cannot be compared across assessments even within the same study. However, statistical equating procedures are commonly employed to allow comparisons over time between assessments within a study.

For example, the scales from TIMSS 1999 (the scales established for each subject and grade in 1999) were statistically equated with the scales from TIMSS 1995 (the scales established for each subject and grade in 1995) so that the TIMSS 1999 results could be placed on the TIMSS 1995 scales. The scales of each subsequent TIMSS assessment, in turn, have been statistically equated with the 1995 scale for the respective subject and grade. Thus, a TIMSS 8th-grade mathematics score of 500 in 1995, for instance, is equivalent to a TIMSS 8th-grade mathematics score of 500 in 2007.

In PISA, the three subject matter scales were developed successively in the year that each subject was first assessed in depth as the major subject matter domain (i.e., reading in 2000, mathematics in 2003, and science in 2006), and all subsequent assessment scales have been statistically equated with those scales. This is to say, PISA established a reading scale in 2000 and placed PISA 2003 and PISA 2006 reading results on the same scale; PISA established a mathematics scale in 2003 and placed PISA 2006 mathematics results on the same scale; and PISA established a science scale in 2006 and will place future PISA science results on the same scale. Thus, a PISA reading score of 500 in 2000, for instance, is equivalent to a PISA reading score of 500 in 2006, but a PISA mathematics score from 2000 cannot be equated with a PISA mathematics score from 2003 or 2006.

It is also important to keep in mind that the procedures used to determine scale scores were developed to produce accurate assessment results for *groups* of students while limiting the testing burden on individual students. They are not intended to produce assessment results for individual students. However, the procedures to determine scale scores provide data that can be readily used in secondary analyses that is done at the student level.

Specifically, during the scaling process, *plausible values* are estimated to characterize the performance of individual students participating in the assessment. Plausible values are imputed values and not test scores for individuals in the usual sense. In fact, they are biased estimates of the proficiencies of individual students. Plausible values do, however, provide unbiased estimates of population characteristics. Plausible values represent what the true performance of an individual might have been, had it been observed. They are estimated as random draws (usually five) from an empirically derived distribution of score values based on the student's observed responses to assessment items and on background variables. Each random draw from the distribution is considered a representative value from the distribution of potential scale scores for all students in the sample who have similar characteristics and identical patterns of item responses. Differences between the plausible values quantify the



degree of precision (the width of the spread) in the underlying distribution of possible scale scores that could have caused the observed performances.

An accessible treatment of the derivation and use of plausible values can be found in Beaton and González (1995). A more technical treatment can be found in the TIMSS 2007 Technical Report (Olson, Martin, and Mullis 2008).

## A.7 Cutpoint scores and achievement levels

### International benchmarks

The IEA has developed international benchmarks for achievement on TIMSS and PIRLS based on *cutpoint* scores that describe what students—who have reached each benchmark’s threshold or “cutpoint” score—know and can do in regard to the subject assessed. For example, 4th-grade students who have reached the TIMSS Intermediate benchmark in mathematics (scored 475 or better)

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. (Gonzales et. al. 2008, p. 13)

The IEA describes student achievement in this manner at four points on its assessment scales: Advanced International Benchmark (cutpoint score of 625), High International Benchmark (550), Intermediate International Benchmark (475), and Low International Benchmark (400). With these four equally spaced benchmarks serving as touchstones for reference, it is possible to interpret what the scores on the PIRLS and TIMSS achievement scales mean more concretely (i.e., understand what knowledge and skills may be demonstrated with a scale score of 513 versus 426).

To describe student performance at the selected points or benchmarks along the TIMSS and PIRLS achievement scales, the IEA uses scale anchoring. Scale anchoring involves selecting a cutpoint score that will “anchor” a benchmark and then identifying items that students scoring within plus or minus 5 scale score points of these anchor points are likely to answer correctly. (The range of plus and minus 5 points around a benchmark’s anchor point is intended to provide a sample that is adequate to analyze the items defining student performance at each benchmark, yet one that is small enough so that performance at each benchmark anchor point is clearly distinguishable from the next.) Subsequently, these items are grouped by content area within benchmarks and

reviewed by subject matter experts. These experts focus on the content of each item and describe the kind of knowledge demonstrated by students answering the item correctly. The experts then provide a summary description of performance at each anchor point leading to a content-referenced interpretation of the achievement results. (Detailed information on the creation of the benchmarks is provided in Mullis, Martin, and Foy 2008a and 2008b and Martin et al. 2007.)

### Levels of proficiency

The OECD has identified levels of proficiency for each of the subject areas of PISA to describe concretely what particular ranges of scores mean. Unlike benchmarks, which are anchored by scale scores, levels of proficiency are anchored by items, which reflect particular proficiencies. Specifically, the knowledge and skills that students are asked to demonstrate in the assessment are classified into one of five or six levels, and the items associated with those specific knowledge and skills become the basis both for classifying students into one of these levels of proficiency and for determining the cutpoint scores for each level.

In PISA, all students within a level are expected to answer at least half of the items from that level correctly. Students at the bottom of a level are able to provide the correct answers to about 52 percent of all items from that level, have a 62 percent chance of success on the easiest items from that level, and have a 42 percent chance of success on the hardest items from that level. Students in the middle of a level have a 62 percent chance of correctly answering items of average difficulty for that level (an overall response probability of 62 percent). Students at the top of a level are able to provide the correct answers to about 70 percent of all items from that level, have a 78 percent chance of success on the easiest items from that level, and have a 62 percent chance of success on the hardest items from that level. Students just below the top of a level would score less than 50 percent on an assessment at the next higher level.

Students at a particular level demonstrate not only the knowledge and skills associated with that level but also the proficiencies classified at lower levels. Thus, all students proficient at level 3 are also proficient at levels 1 and 2. Patterns of responses for students below level 1 suggest that these students are unable to answer at least half of the items from level 1 correctly.

Given that items are the basis for classifying students into the levels of proficiency, the cutpoint scores for particular levels vary from assessment to assessment. For more details about the PISA levels of proficiency, see the *PISA 2006 Technical Report* (OECD 2008).

## A.8 Comparing results from PISA 2000, 2003, and 2006

The PISA 2000, 2003, and 2006 assessments of reading, mathematics and science are linked assessments. That is, the sets of items used to assess reading, mathematics, and science in PISA 2000, 2003, and 2006 include a subset of common items. For example, there were 20 common mathematics items that were used in PISA in 2000 and 2003. To establish common reporting metrics for PISA, the difficulty of each link item is measured on different occasions and compared. Using procedures that are detailed in the *PISA 2006 Technical Report* (OECD 2008), the change in the difficulty of each of the individual link items is used in determining a score transformation that allows the reporting of the data on a common scale. As each item provides slightly different information about the link transformation, it follows that the chosen sample of link items will influence the estimated transformation. Thus, if an alternative set of link items had been chosen, the resulting transformation would be slightly different. The consequence is an uncertainty in the transformation due to the sampling of the link items, just as there is uncertainty in values such as country means due to the sampling of students.

### Linking error

Such uncertainty that results from the link-item sampling is referred to as linking error and this error must be taken into account when making certain comparisons between PISA 2000, 2003, and 2006 results.<sup>1</sup> Just as with the error that is introduced through the process of sampling students, the exact magnitude of this linking error can only be estimated. As with sampling errors, the likely range of magnitude for the errors is represented as a standard error. The standard errors of linking are as follows:

PISA survey cycle	Assessment domain	Linking error
2000–2003	Science	3.11
2000–2003	Reading	5.31
2000–2006	Reading	4.98
2003–2006	Reading	4.47
2003–2006	Mathematics	1.38

When comparing two country means from PISA taken at different times (e.g., 2000 and 2003), the calculation of the standard error of the difference includes the standard errors of the two individual scores in addition to the linking error, making the resulting statement of statistical significance more conservative than if there were no linking error. For example, to calculate the standard error on the difference between scores obtained for a country in 2000 and 2003, the following formula is applied when

$\sigma^2_{(\hat{\mu}_{2000})}$  and  $\sigma^2_{(\hat{\mu}_{2003})}$  represent the standard errors for the results of PISA 2000 and PISA 2003, respectively, and  $\sigma^2_{(linking\ error)}$  represents the linking error between PISA 2000 and PISA 2003:

$$SE = \sigma^2_{(\hat{\mu}_{2000})} + \sigma^2_{(\hat{\mu}_{2003})} + \sigma^2_{(linking\ error)}$$

Because linking error should be taken into account when comparing means from different PISA assessment cycles, the results of simple *t*-tests that do not include the linking error will differ from the results published in the official PISA reports and this special analysis. For example, without adjusting for linking error, significance tests comparing reading literacy scores between PISA 2000 and PISA 2003 indicate that 15 jurisdictions measurably changed. However, after adjusting for linking error, only 9 jurisdictions are shown to have measurably changed at the .05 level of significance.

### PISA tests of significance in 2000

Results from PISA 2000 summarized in this special analysis have been updated from what was presented in the PISA 2000 U.S. report (Lemke et al. 2001). Some country differences discussed in this report were not reported as statistically significant in the PISA 2000 U.S. report. In that report, a Bonferroni adjustment was used in all multiple comparisons of countries. This was not the case when PISA 2003 and 2006 data were analyzed and reported, which makes it difficult to compare results from the PISA 2000 U.S. report with results from the PISA 2003 and 2006 U.S. reports. The use of the Bonferroni adjustment for multiple comparisons was discontinued in order to avoid the possibility that comparisons of achievement between countries could be interpreted differently depending on the numbers of countries compared.

## A.9 Comparing results from TIMSS 1995 and 1999

### TIMSS 1995 scale scores

TIMSS 1995 utilized a one-parameter item response theory (IRT) model to produce score scales that summarized the achievement results in the original reports. The TIMSS 1995 data were rescaled using a three-parameter IRT model to match the procedures used to scale the 1999, 2003, and 2007 TIMSS data. The three-parameter model was preferred to the one-parameter model because it can more accurately account for the differences among items in their ability to discriminate between students of high and low ability. After careful study of the rescaling process, the International Study Center concluded that the fit between the original TIMSS data and the rescaled TIMSS data met acceptable standards. However, as a result of rescaling, the average achievement scores of some countries changed from those initially reported in 1996

<sup>1</sup> Because PIRLS and TIMSS are designed differently, there is no need to account for linking error.

and 1997 (Peak 1996; NCES 1997). The rescaled TIMSS scores are included in this special analysis.

## TIMSS tests of significance in 1995 and 1999

Tests of significance used in this special analysis to make multiple country comparisons for TIMSS 1995 and TIMSS 1999 differ from those presented in their respective U.S. reports (NCES 1997, NCES 2000). Some country differences discussed in this special analysis were not reported as statistically significant in the TIMSS 1995 and 1999 U.S. reports. The reason for this is that a Bonferroni adjustment was used in all multiple comparisons of countries in those reports. However, the TIMSS 2003 and 2007 U.S. reports discontinued use of the Bonferroni adjustment. To maintain the comparability of results across all four TIMSS assessments, none of the tests of significance presented in this report used the Bonferroni adjustment for multiple comparisons.

## A.10 Confidentiality and disclosure limitations

In accord with NCES statistical standard 4-2-6 (Seastrom 2003), confidentiality analyses for the United States are implemented to provide reasonable assurance that public-use data files issued by the IEA and OECD do not allow the identification of individual U.S. schools or students when compared against publicly available data collections. Disclosure limitations include the identification and masking of potential disclosure risks for schools and adding an additional measure of uncertainty of school, teacher, and student identification by randomly swapping a small number of data elements within the student, teacher, and school files.

## A.11 Nonresponse bias analysis

In accord with NCES statistical standard 4-4-1 (Seastrom 2003), nonresponse bias analyses are conducted for any survey stage of data collection with a unit or item response rate less than 85 percent. Estimates of survey characteristics for nonrespondents and respondents are compared on sampling frame variables, for which data on respondents and nonrespondents are available, so as to assess the potential nonresponse bias. Comparisons are made before and after weight adjustments for nonresponse bias, and these comparisons examine both bias and relative bias. Bias is calculated as the difference between the value of a survey characteristic for all schools (or students) that were sampled and the value of that characteristic for just those schools (or students) that actually responded. Relative bias reports this difference as a percentage of the value of the survey characteristic for all schools (or students).

For example, in PIRLS 2006, of the 214 U.S. schools in the original sample eligible to participate in PIRLS, 120 agreed to participate (for a response rate of 57 percent, weighted). An additional 63 substitute schools agreed to participate, bringing the total up to 183 schools for a weighted response rate, using final adjusted weights, of 86 percent of eligible schools (NCES 2009-050). With these substitute schools, the United States met the international guidelines for school response rates. However, a nonresponse bias analysis was conducted to determine what amount of bias and relative bias existed in the original and final sample. Such an analysis is considered to have found potential bias in a survey characteristic if either (a) a test of statistical significance indicates a significant difference between the responding and nonresponding sample percentages for that survey characteristic or (b) the relative bias for that survey characteristic is greater than 10 percent.

The PIRLS 2006 nonresponse bias analysis identified four variables as either (a) significant in bivariate or multivariate bias analysis or (b) having a relative bias of greater than 10 percent. The following list summarizes these variables and the difference found, after weight adjustments, between the value of the variable in the final sample of responding schools (which includes substitute schools) and value in the full sample from which the schools were drawn:

- private schools (19.4 percent of the schools in the final sample vs. 17.0 percent of the schools in the full sample)
- schools in the Northeast Census region (20.7 percent of schools in the final sample vs. 21.7 percent of schools in the full sample)
- percentage of Asian or Pacific Islander students (2.6 percent of the students in schools in the final sample vs. 2.8 percent of the students in schools in the full sample)
- percentage of White, non-Hispanic students (53.9 percent of the students in schools in the final sample vs. 53.5 percent of the students in schools in the full sample)

In TIMSS 2007, of the 290 U.S. schools in the original sample eligible to participate in TIMSS at 4th grade (TIMSS-4) and 287 at 8th grade (TIMSS-8), 202 schools at grade 4 and 197 schools at grade 8 agreed to participate (or, respectively, 70 and 68 percent, weighted). An additional 55 substitute schools for TIMSS-4 and 42 substitute schools for TIMSS-8 agreed to participate, bringing the total up to 257 schools at grade 4 and 239 schools at grade 8 for weighted response rates, using final

adjusted weights, of 89 and 83 percent, respectively, of eligible schools (NCES 2009-012). With these substitute schools, the United States met the international guidelines for school response rates. However, a nonresponse bias analysis was conducted to determine what amount of bias and relative bias existed in the original and final sample.

The TIMSS 2007 nonresponse bias analysis identified several variables as either (a) significant in bivariate or multivariate bias analysis or (b) having a relative bias of greater than 10 percent. The following list summarizes these variables (in the 4th and 8th grade sample) and the difference found, after weight adjustments, between the value of the variable in the final sample of responding schools (which includes substitute schools) and value in the full sample from which the schools were drawn:

#### TIMSS-4

- central city schools (29.6 percent of the schools in the final sample vs. 32.1 percent of the schools in the full sample)
- rural schools (28.4 percent of the schools in the final sample vs. 26.2 percent of the schools in the full sample)
- percentage of American Indian/Alaska Native students (1.6 percent of the students in the final sample vs. 1.4 percent of the students in schools in the full sample)
- percentage of White, non-Hispanic students (60.9 percent of the students in schools in the final sample vs. 59.2 percent of the students in schools in the full sample)

#### TIMSS-8

- central city schools (33.5 percent of the schools in the final sample vs. 30.0 percent of the schools in the full sample)
- percentage of American Indian/Alaska Native students (1.5 percent of the students in the final sample vs. 1.3 percent of the students in schools in the full sample)
- percentage of Black, non-Hispanic students (15.1 percent of the students in the final sample vs. 15.7 percent of the students in schools in the full sample)

In PISA 2006, of the 209 U.S. schools in the original sample eligible to participate in PISA, 145 agreed to participate (69 percent, weighted). An additional 21

substitute schools agreed to participate, bringing the total up to 166 schools for a 79 percent overall response rate (NCES 2009-055). With these substitute schools, the United States met the PISA standards for a minimum participation rate (65 percent of original sample schools). However, since the U.S. response rate level did not meet the NCES standard of 85 percent, the nonresponse bias analysis was conducted to shed light on the quality of the data.

The PISA 2006 nonresponse bias analysis identified eight variables as either (a) significant in bivariate or multivariate bias analysis or (b) having a relative bias of greater than 10 percent. The following list summarizes these variables and the difference found, after weight adjustments, between the value of the variable in the final sample of responding schools (which includes substitute schools) and the value in the full sample from which the schools were drawn:

- central city schools (25.8 percent of the schools in the final sample vs. 22.8 percent of the schools in the full sample)
- urban fringe or large town schools (24.0 percent of the schools in the final sample vs. 24.7 percent of the schools in the full sample)
- schools in the South Census region (35.9 percent of schools in the final sample vs. 34.0 percent of schools in the full sample)
- schools in the West Census region (15.6 percent of schools in the final sample vs. 20.4 percent of schools in the full sample)
- percentage of American Indian/Alaska Native students (1.4 percent of the students in the final sample vs. 2.2 percent of the students in schools in the full sample)
- percentage of Hispanic students (12.0 percent of the students in schools in the final sample vs. 10.8 percent of the students in schools in the full sample)
- percentage of White, non-Hispanic students (61.5 percent of the students in schools in the final sample vs. 63.5 percent of the students in schools in the full sample)
- percentage of students of Other race/ethnicity (2.9 percent of the students in the final sample vs. 2.5 percent of the students in schools in the full sample)

---

## A.12 State participation in international assessments

U.S. states have not participated in PIRLS or PISA to date, but various states have participated in different TIMSS administrations as benchmarking participants. States that have been benchmarking participants in TIMSS are able to compare the performance of their students with that of all the other participating countries; however, their students do not contribute to the U.S. national sample. The list that follows identifies the states that participated in each of the administrations of TIMSS as benchmarking participants.

### TIMSS 1995

- Colorado
- Illinois
- Minnesota
- Missouri
- Oregon

### TIMSS 1999 (grade 8 only)

- Connecticut
- Idaho

- Illinois
- Indiana
- Maryland
- Massachusetts
- Michigan
- Missouri
- North Carolina
- Oregon
- Pennsylvania
- South Carolina
- Texas

### TIMSS 2003

- Indiana

### TIMSS 2007

- Massachusetts
- Minnesota