

Reading Achievement of U.S. Fourth-Grade Students in an International Context

First Look at the Progress in International Reading Literacy
Study (PIRLS) and ePIRLS 2016



This page intentionally left blank

Reading Achievement of U.S. Fourth-Grade Students in an International Context

First Look at the Progress in International Reading Literacy
Study (PIRLS) and ePIRLS 2016

DECEMBER 2017

Catharine Warner-Griffin
Huili Liu
Chrystine Tadler
Insight Policy Research

Debbie Herget
Ben Dalton
RTI International

Sheila Thompson
Project Officer
National Center for Education Statistics

NCES 2018-017
U.S. DEPARTMENT OF EDUCATION

 **NATIONAL CENTER FOR
EDUCATION STATISTICS**
Institute of Education Sciences

U.S. Department of Education

Betsy DeVos
Secretary

Institute of Education Sciences

Thomas W. Brock
Commissioner, National Center for Education Research
Delegated Duties of the Director

National Center for Education Statistics

Peggy G. Carr
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:

NCES, IES, U.S. Department of Education
Potomac Center Plaza
550 12th Street SW
Washington, DC 20202

December 2017

The NCES Home Page address is <https://nces.ed.gov>.

The NCES Publications and Products address is <https://nces.ed.gov/pubsearch>.

This publication is only available online. To download, view, and print the report as a PDF file, go to the NCES Publications and Products address shown above.

This report was prepared for the National Center for Education Statistics under Contract No. ED-IES-14-C-0057 with RTI International. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Suggested Citation

Warner-Griffin, C., Liu, H., Tadler, C., Herget, D., and Dalton, B. (2017). *Reading Achievement of U.S. Fourth-Grade Students in an International Context: First Look at the Progress in International Reading Literacy Study (PIRLS) 2016 and ePIRLS 2016* (NCES 2018-017). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from <https://nces.ed.gov/pubsearch>.

Content Contact

Sheila Thompson
(202) 245-8330
sheila.thompson@ed.gov

Acknowledgments

The authors wish to thank the students, teachers, and school officials who participated in PIRLS 2016. Without their assistance and cooperation, this study would not be possible.

Contents

Acknowledgments	iii
List of Tables	v
List of Figures	v
List of Exhibits	vi
Introduction	1
What Are PIRLS and ePIRLS?	1
What PIRLS and ePIRLS Measure	1
Reporting PIRLS and ePIRLS 2016 Results	3
Selected Findings	6
U.S. PIRLS Performance	6
U.S. ePIRLS Performance	7
PIRLS Tables and Figures	8
ePIRLS Tables and Figures	17
References	20
Appendix A. Technical Notes	21
Sampling and Response Rates	21
Assessment and Questionnaires	25
Reporting Results	26
Sampling Weights and Standard Errors	26
Statistical Comparisons	27

List of Tables

Table 1.	PIRLS overall reading average scale scores of fourth-grade students, by education system: 2016	8
Table 2.	PIRLS purposes for reading and processes of comprehension average subscale scores of fourth-grade students, by education system: 2016.....	11
Table 3.	ePIRLS online informational reading average scale scores of fourth-grade students, by education system: 2016.....	17
Table A-1.	Sampling or other issues for PIRLS and ePIRLS, by education system: 2016	23
Table A-2.	Sampling or other issues for PIRLS, by education system: 2001, 2006, and 2011	24

List of Figures

Figure 1.	Percentage of fourth-grade students reaching the PIRLS international benchmarks in reading, by education system: 2016.....	9
Figure 2.	Change in PIRLS overall reading average scale scores of fourth-grade students, by education system: 2011 to 2016 and 2001 to 2016	13
Figure 3.	Trends in U.S. fourth-grade students' PIRLS overall reading average scale scores and cut scores at the 10th, 25th, 75th, and 90th percentiles: 2001, 2006, 2011, and 2016.....	15
Figure 4.	PIRLS overall reading average scale scores of U.S. fourth-grade students, by sex, race/ethnicity, and percentage of public school students eligible for free or reduced-price lunch: 2016.....	16
Figure 5.	Percentage of fourth-grade students reaching the ePIRLS international benchmarks in online informational reading, by education system: 2016	18
Figure 6.	ePIRLS online informational reading average scale scores of U.S. fourth-grade students, by sex, race/ethnicity, and percentage of public school students eligible for free or reduced-price lunch: 2016	19

List of Exhibits

Exhibit 1.	Percentages of the PIRLS and ePIRLS assessment items devoted to each reading purpose and devoted to each comprehension process in the PIRLS framework: 2016	2
Exhibit 2.	Description of PIRLS international reading benchmarks: 2016.....	4
Exhibit 3.	Description of ePIRLS international reading benchmarks: 2016.....	5

Introduction

What Are PIRLS and ePIRLS?

The Progress in International Reading Literacy Study (PIRLS) is an international assessment of student performance in reading literacy at the fourth grade. PIRLS measures students in the fourth year of formal schooling¹ because this is typically when students' learning transitions from a focus on *learning to read* to a focus on *reading to learn*. PIRLS is a collaborative effort between participating countries and the International Association for the Evaluation of Educational Achievement (IEA).² Administered every 5 years since 2001, PIRLS 2016 marks the survey's fourth administration. The current administration also provides the first results for ePIRLS: a computer-based extension of PIRLS designed to assess students' comprehension of online information. In 2016, some 58 education systems³ participated in the PIRLS assessment at the fourth year of formal schooling and 16 of these systems also participated in ePIRLS.

What PIRLS and ePIRLS Measure

PIRLS provides a comprehensive picture of students' reading literacy achievement. Reading literacy is defined as follows:

Reading literacy is the ability to understand and use those written language forms required by society and/or valued by the individual. Readers can construct meaning from texts in a variety of forms. They read to learn, to participate in communities of readers in school and everyday life, and for enjoyment (Mullis and Martin 2015, p. 12).

The PIRLS framework focuses on two key content domains: purposes for reading and processes of comprehension. There are two key purposes for reading: reading for literary experience and reading to acquire and use information. Reading for literary experience involves reading for interest or pleasure using texts that emphasize characters, plot events, settings, and themes. Reading to acquire and use information includes reading to learn, typically using informational texts that tell students about themselves and the world around them. The assessment also integrates four comprehension

¹For ease of presentation, throughout this report, student participants are referred to as fourth-grade students.

²PIRLS 2016, and its partner assessment ePIRLS, were sponsored by the IEA and implemented by the TIMSS & PIRLS International Study Center at Boston College.

³The term "education system" refers to IEA member countries and benchmarking participants. IEA member "countries" may be complete, independent political entities or nonnational entities that represent a portion of a country (e.g., England, Hong Kong, the Flemish community of Belgium). Nonnational entities are indicated by italics in the tables and figures, with the three-letter international abbreviation for their country following their name. Nonnational entities that are represented by their larger country in the main results (e.g., Abu Dhabi in the United Arab Emirates, Ontario in Canada), or whose countries are not IEA members (Buenos Aires) are designated as "benchmarking participants." Benchmarking participants are included in a separate section in the tables and figures. For convenience, this report uses the generic term "education systems" when summarizing results. PIRLS was administered in 61 education systems; this report does not include the results of three education systems that did not administer PIRLS at the target grade.

processes within both of the purposes for reading. The four processes of comprehension consist of the ability to (1) focus on, and retrieve, explicitly stated information, (2) make straightforward inferences, (3) interpret and integrate ideas and information, and (4) evaluate, and critique, content and textual elements. These four comprehension processes are presented through two combined measures: retrieving and straightforward inferencing and interpreting, integrating, and evaluating.

The new ePIRLS is an innovative, computer-based assessment of online reading. It is designed to measure students' approaches to informational reading in an online environment. As webpages become a more common source for acquiring information, ePIRLS provides measures of students' online reading skills and competencies. Example tasks include identifying a specific webpage, filtering content on the webpage for the most relevant information, relating information across websites, and judging the credibility of information on the website. Exhibit 1 illustrates the framework for reading literacy used for PIRLS and ePIRLS.

Exhibit 1. Percentages of the PIRLS and ePIRLS assessment items devoted to each reading purpose and comprehension process in the PIRLS framework: 2016

Domain skills	Percentages devoted to each domain	
	PIRLS	ePIRLS
Purposes for Reading		
Literary experience		
 Literary passages with questions addressing theme, plot, characters, and setting	50	0
Acquire and use information		
 Informative or instructional passages with questions addressing definitions or facts	50	100
Processes of comprehension		
Retrieving and straightforward inferencing		
 Focus on and retrieve explicitly stated information: Locate and understand text related to a question posed	20	20
 Make straightforward inferences: Move beyond stated material and connect pieces of information	30	30
Interpreting, integrating, and evaluating		
 Interpret and integrate ideas and information: Integrate personal knowledge with explicit contextual information	30	30
 Evaluate and critique content and textual elements: Critically assess the text from a personal or objective view	20	20

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), 2016.

The PIRLS assessment provides students with short reading passages and accompanying items to assess understanding. These items are standardized measures that use multiple-choice and open-ended formats. The ePIRLS assessment provides students with passages or a series of web pages on a desktop or laptop computer and requires students to use a mouse to navigate the assessment. The ePIRLS simulated web pages consisted of two tasks. In the tasks, a teacher avatar guides students through the ePIRLS assignments, prompting the students with questions about the online information. For both PIRLS and ePIRLS 2016, multiple-choice items were worth one point, whereas the value of open-ended items (in which students wrote or typed responses) depended on the depth of understanding required: short-answer items were worth 1 or 2 points, while extended response items were worth up to 3 points.

Reporting PIRLS and ePIRLS 2016 Results

This report summarizes performance on PIRLS and ePIRLS 2016 from a U.S. perspective. PIRLS results are based on nationally representative samples of fourth-graders. The international data reported for PIRLS 2016 in this report cover 58 countries or other education systems, including the United States. Education systems that did not administer PIRLS at the target grade are not included in this report; see the international report for their results (Mullis et al. 2017a, 2017b).

Scale scores and percentiles

Average scores on PIRLS and ePIRLS are reported on an overall reading scale ranging from 0 to 1,000, with a fixed scale centerpoint of 500 and a standard deviation of 100. The PIRLS reading achievement scale was established in PIRLS 2001 based on the achievement distribution across all countries or education systems that participated in 2001. The results of each successive administration of PIRLS have been placed on the same scale so that scores are comparable across years. The ePIRLS online informational reading scale summarizes fourth-grade students' performance in a simulated online environment. The ePIRLS results are reported on the same scale as PIRLS. However, ePIRLS assesses how well students comprehend material and navigate content in an online environment, using content tabs, links, and icons. The PIRLS scale provides a more general measure of reading comprehension and purposes for reading.

For PIRLS, average subscale scores are also reported for the individual domain skills in the two content domains: purposes for reading and processes of comprehension (exhibit 1). However, direct comparisons should not be made between subscale scores because the scaling of PIRLS data is conducted separately for the combined overall reading scale and each of the subscales.

The distribution of student achievement is reported using percentiles. Percentile scores identify the threshold (or cut) scores along the scale that correspond to the lowest 10 percent, lowest quarter, highest quarter, and top 10 percent of students (i.e., the 10th, 25th, 75th, and 90th percentiles).

International benchmarks

PIRLS international benchmarks provide a way to interpret the scale scores and to understand how student proficiency varies at different points on the scale. Each successive point, or benchmark, is associated with the knowledge and skills that students successfully demonstrate at each level. PIRLS describes four levels: *Advanced*, *High*, *Intermediate*, and *Low*. Exhibit 2 describes the skills and score cut-point for each PIRLS benchmark.

Exhibit 2. Description of PIRLS international reading benchmarks: 2016

Benchmark (score cut-point)	Description of benchmark skills
Advanced (625)	<ul style="list-style-type: none">  Students can interpret story events and character actions to describe reasons, motivations, feelings, and character development with full text-based support in relatively complex literary texts. They can begin to evaluate the effect on the reader of the author's language and style choices.  Students can distinguish and interpret complex information from different parts of text and provide full text-based support when reading relatively complex informational texts. They can integrate information across a complex informational text and explain relationships and sequence activities. They can begin to evaluate visual and textual elements to consider the author's point of view.
High (550)	<ul style="list-style-type: none">  Students can locate and distinguish significant actions and details embedded across relatively complex literary texts. They can make inferences to explain relationships between intentions, actions, events, and feelings, and give text-based support. They can interpret and integrate story events and character actions, traits, and feelings as they develop across the text. They can recognize the use of some language features (e.g., metaphor, tone, imagery).  Students can locate and distinguish relevant information within relatively complex tables or informational texts. They can make inferences about logical connections to provide explanations and reasons. They can integrate textual and visual information to interpret the relationship between ideas and evaluate and make generalizations about content and textual elements.
Intermediate (475)	<ul style="list-style-type: none">  Students can independently locate, recognize, and reproduce explicitly stated actions, events, and feelings when reading a mix of simpler and relatively complex literary texts. They can make straightforward inferences about the attitudes, feelings, and motivations of main characters. They can begin to recognize language choices and interpret obvious reasons and causes, recognize evidence, and give examples.  Students can locate and reproduce two or three pieces of information from a mix of simpler and relatively complex informational texts. They can make straightforward inferences to provide factual explanations and begin to interpret and integrate information to order events.
Low (400)	<ul style="list-style-type: none">  Students can locate and retrieve explicitly stated information, actions, or ideas when reading relatively simple literary texts. They can make straightforward inferences about events, and begin to interpret story events and central ideas.  Students can locate and reproduce explicitly stated information from relatively simple informational texts and other formats and begin to make straightforward inferences about explanations, actions, and descriptions.

NOTE: Score cut-points for the international benchmarks are determined through scale anchoring. Scale anchoring involves selecting benchmarks (scale points) on the achievement scales in terms of student performance and then identifying items that students scoring at the anchor points answer correctly. The score cut-points are set at equal intervals along the achievement scales. The score cut-points were selected to be as close as possible to the standard percentile cut-points (i.e., 90th, 75th, 50th, and 25th percentiles). More information on how the score cut-points were set can be found in the PIRLS technical documentation at <https://timssandpirls.bc.edu/publications/pirls/2016-methods.html>.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), 2016.

Exhibit 3 describes the skills and score cut-point for each ePIRLS benchmark.

Exhibit 3. Description of ePIRLS international reading benchmarks: 2016

Benchmark (score cut-point)	Description of ePIRLS benchmark skills
Advanced (625)	 Students can make inferences from relatively complex online informational texts to support an explanation. They can interpret and integrate information within and across web pages with interactive features to explain relationships and show thorough understanding. Students can evaluate the effects of textual, visual, and interactive elements and begin to consider the writer's point of view.
High (550)	 Students can make inferences to distinguish relevant information and provide comparisons when reading and viewing relatively complex online informational texts. They can interpret and integrate information within and across web pages with interactive features to provide examples and make contrasts. They can evaluate how graphic elements and language choices support content.
Intermediate (475)	 Students can locate and reproduce information presented in various forms, including independent use of navigation features, when reading and viewing relatively complex online informational texts. They can make straightforward inferences to recognize reasons and actions. They can interpret and integrate information across a web page to recognize causes, comparisons, and explanations. Students can begin to evaluate the use of interactive features to convey information.
Low (400)	 Students can locate and reproduce explicitly stated information from web pages in relatively complex online informational texts that contain text and a variety of dynamic, navigable features (e.g., timelines, pop-up boxes). They can begin to make straightforward inferences about descriptions.

NOTE: Score cut-points for the international benchmarks are determined through scale anchoring. Scale anchoring involves selecting benchmarks (scale points) on the achievement scales in terms of student performance and then identifying items that students scoring at the anchor points answer correctly. The score cut-points are set at equal intervals along the achievement scales. The score cut-points were selected to be as close as possible to the standard percentile cut-points (i.e., 90th, 75th, 50th, and 25th percentiles). More information on how the score cut-points were set can be found in the PIRLS and ePIRLS technical documentation at <https://timssandpirls.bc.edu/publications/pirls/2016-methods.html>.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), 2016.

For more PIRLS 2016 results, visit the NCES PIRLS website (<https://nces.ed.gov/surveys/pirls/pirls2016/>), the PIRLS website (<https://www.pirls.org>), or the TIMSS & PIRLS International Study Center website (<https://timssandpirls.bc.edu/pirls2016/>).

Selected Findings

U.S. PIRLS Performance

- The U.S. overall average reading score was 549 (table 1). This score was higher than the PIRLS scale centerpoint, which is set at 500 points. The U.S. overall average reading score was lower than the averages for 12 education systems, higher than the averages for 30 education systems, and not significantly different from the averages for 15 education systems.
- In 2016, some 16 percent of U.S. fourth-graders performed at or above the *Advanced* (625) benchmark, and 53 percent of fourth-graders performed at or above the *High* (550) benchmark (figure 1). The percentages of U.S. fourth-graders performing at or above the *Advanced* and *High* benchmarks were higher than the international median. Seven education systems (Singapore, the Russian Federation, *Northern Ireland-GBR*, Ireland, Poland, *England-GBR*, and *Moscow City-RUS*) had a higher percentage of fourth-graders performing at or above the *Advanced* benchmark than the United States. For the *Low* (400) benchmark, 96 percent of U.S. fourth-graders performed at or above the benchmark, and 24 education systems had a smaller percentage of students performing at or above the benchmark.
- U.S. fourth-graders scored higher, on average, than the PIRLS scale centerpoint across all four reading subscales in 2016 (table 2). The U.S. average score for each subscale ranged from 543 to 557. Seven education systems scored higher on average than the United States on every subscale: the Russian Federation, Singapore, Ireland, *Northern Ireland-GBR*, Poland, Finland and *Moscow City-RUS*.
- In 2016, among U.S. fourth-graders, females scored higher on average on the overall reading scale than males (553 vs. 545) (figure 4). Compared to the U.S. overall average reading score, White and Asian fourth-graders scored higher on average, while Black and Hispanic fourth-graders scored lower on average.
- U.S. fourth-graders in public schools with less than 50 percent of students eligible for free or reduced-price lunch scored higher on average than the U.S. overall average reading score, while fourth-graders in public schools with more than 75 percent of students eligible for free or reduced-price lunch scored lower on average than the U.S. overall average reading score (figure 4).

Performance over time

- The overall average reading scores increased in 11 education systems between the first administration of PIRLS (in 2001) and 2016 (figure 2). Since the last administration of PIRLS (in 2011) and 2016, overall average reading scores increased in 10 education systems. Overall average reading scale scores increased during both of these intervals for six education systems: the Russian Federation, Hungary, Italy, Slovenia, Norway, and *Quebec-CAN*. In five education systems, the overall average reading score decreased since the last administration.
- There was no measurable change in the U.S. overall average reading scale score between 2001 (542) and 2016 (549) (figures 2 and 3). Between 2011 and 2016, the overall average reading score for U.S. fourth-graders declined from 556 to 549. However, the U.S. overall average reading score in 2016 remained higher than the U.S. overall average reading score in 2006 (540) (figure 3).
- At the 25th percentile, U.S. fourth-graders scored lower in 2016 (501) than in 2011 (510) (figure 3). The average scores for U.S. fourth-graders at the other percentiles were not measurably different between 2011 and 2016.

U.S. ePIRLS Performance

- In the new 2016 ePIRLS online assessment, the United States scored above the ePIRLS scale centerpoint, set at 500 (table 3). U.S. fourth-graders' ePIRLS online informational reading average score was 557, which was lower than the averages for 3 education systems, higher than the averages for 10 education systems, and not measurably different from the averages for 2 education systems.
- In 2016, some 18 percent of U.S. fourth-graders scored at or above the *Advanced* benchmark in online informational reading, placing the United States above the international median, along with five other education systems (figure 5). A larger percentage of fourth-graders in one country, Singapore (34 percent), met or exceeded the *Advanced* benchmark compared to the United States (18 percent). Additionally, 56 percent of U.S. fourth-graders scored at or above the *High* benchmark in online informational reading.
- In 2016, among U.S. fourth-graders, females scored higher on average on the online informational reading scale than males (560 vs. 554) (figure 6). Compared to the U.S. average online informational reading scale score (557), White and Asian fourth-graders scored higher on average, while Black and Hispanic fourth graders scored lower on average.
- U.S. fourth-graders in public schools with less than 50 percent of students eligible for free or reduced-price lunch scored higher on average than the U.S. online informational reading average scale score, while fourth-graders in public schools with more than 75 percent of students eligible for free or reduced-price lunch scored lower on average than the U.S. online informational reading average scale score (figure 6).

PIRLS Tables and Figures

Table 1. PIRLS overall reading average scale scores of fourth-grade students, by education system: 2016

Education system	Overall average scale score	Education system	Overall average scale score
PIRLS scale centerpoint ¹	500	PIRLS scale centerpoint ¹	500
Russian Federation	581 ▲	Spain	528 ▼
Singapore ²	576 ▲	<i>Belgium (Flemish)-BEL</i>	525 ▼
<i>Hong Kong-CHN</i> ²	569 ▲	New Zealand	523 ▼
Ireland	567 ▲	France	511 ▼
Finland	566 ▲	<i>Belgium (French)-BEL</i> ²	497 ▼
Poland	565 ▲	Chile	494 ▼
<i>Northern Ireland-GBR</i>	565 ▲	Georgia ²	488 ▼
Norway	559 ▲	Trinidad and Tobago	479 ▼
<i>Chinese Taipei-CHN</i>	559 ▲	Azerbaijan	472 ▼
<i>England-GBR</i>	559 ▲	Malta ²	452 ▼
Latvia ²	558 ▲	United Arab Emirates	450 ▼
Sweden	555	Bahrain	446 ▼
Hungary	554	Qatar	442 ▼
Bulgaria	552	Saudi Arabia	430 ▼
United States²	549	Iran, Islamic Republic of	428 ▼
Lithuania	548	Oman	418 ▼
Italy	548	Kuwait	393 ▼
Denmark ²	547	Morocco	358 ▼
<i>Macao-CHN</i>	546	Egypt	330 ▼
Netherlands ²	545	South Africa	320 ▼
Australia	544		
Czech Republic	543	Benchmarking education systems	
Canada ²	543	<i>Moscow City-RUS</i>	612 ▲
Slovenia	542	<i>Madrid-ESP</i> ²	549
Austria ²	541 ▼	<i>Quebec-CAN</i> ²	547
Germany	537 ▼	<i>Ontario-CAN</i>	544
Kazakhstan	536 ▼	<i>Andalusia-ESP</i>	525 ▼
Slovak Republic	535 ▼	<i>Dubai-UAE</i>	515 ▼
Israel ²	530 ▼	<i>Buenos Aires-ARG</i>	480 ▼
Portugal ²	528 ▼	<i>Abu Dhabi-UAE</i>	414 ▼

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

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

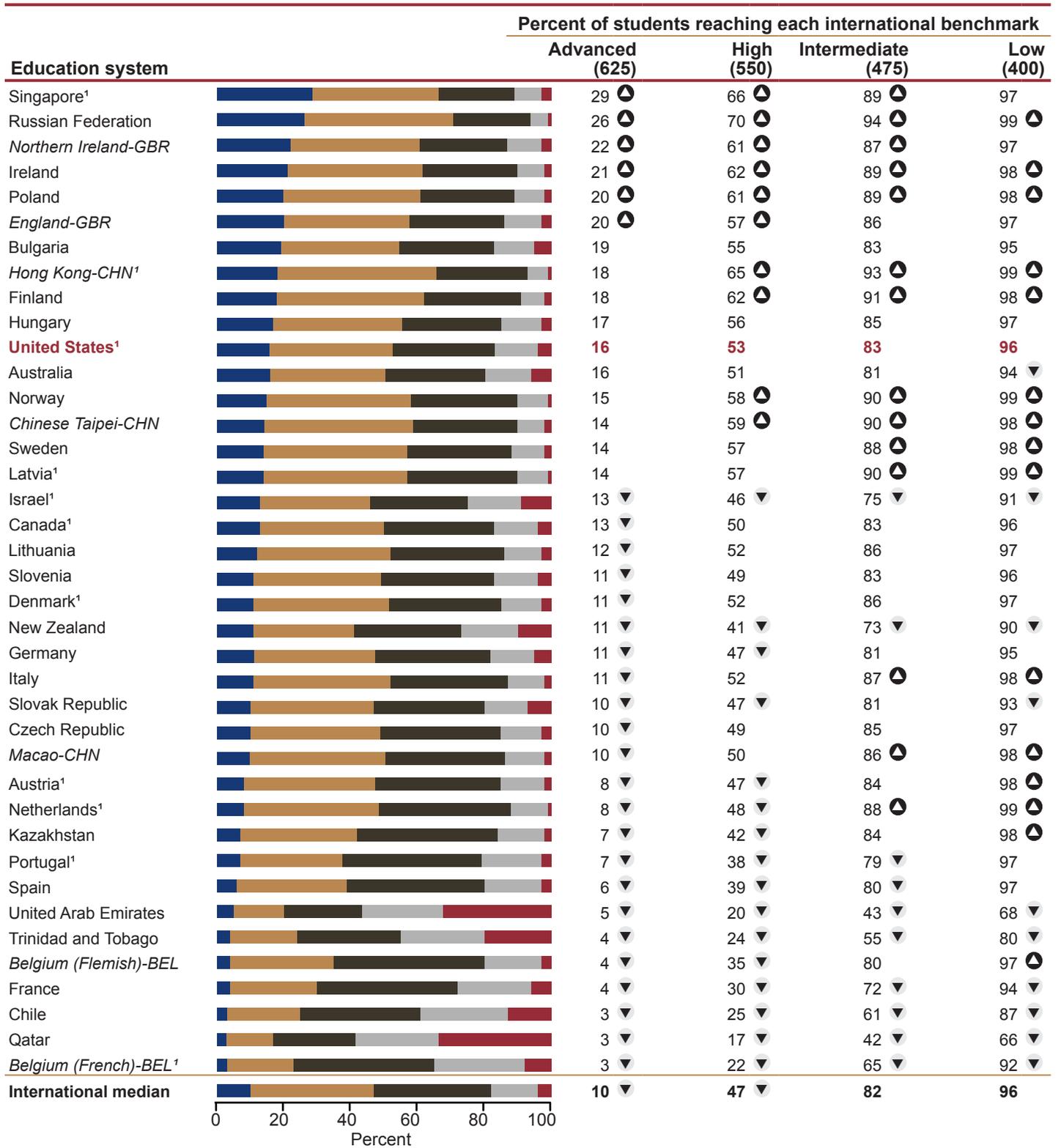
¹The PIRLS scale centerpoint is set at 500 points and represents the mean of the overall achievement distribution in 2001. The PIRLS scale is the same in each administration; thus a value of 500 in 2016 equals 500 in 2001.

²See appendix table A-1 for details on sampling issues in these education systems.

NOTE: Education systems are ordered by *overall* average scale score. Italics indicate participants identified and counted in this report as a nonnational entity that represents a portion of a country. All average scores reported as higher or lower than the U.S. average score are different at the .05 level of statistical significance. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one education system may be significant, while a large apparent difference between the United States and another education system may not be significant. The standard errors of the estimates are shown in table 1 available at https://nces.ed.gov/surveys/pirls/pirls2016/tables/pirls2016_table01.asp. Education systems that did not administer PIRLS at the target grade are not shown; see the international report for their results. Five education systems participated in PIRLS Literacy (Egypt, Iran, Kuwait, Morocco, and South Africa); two of these education systems completed both PIRLS and PIRLS Literacy (Iran and Morocco). More detail on PIRLS Literacy is available in the technical notes at <https://nces.ed.gov/surveys/pirls/pirls2016/technotes.asp>.

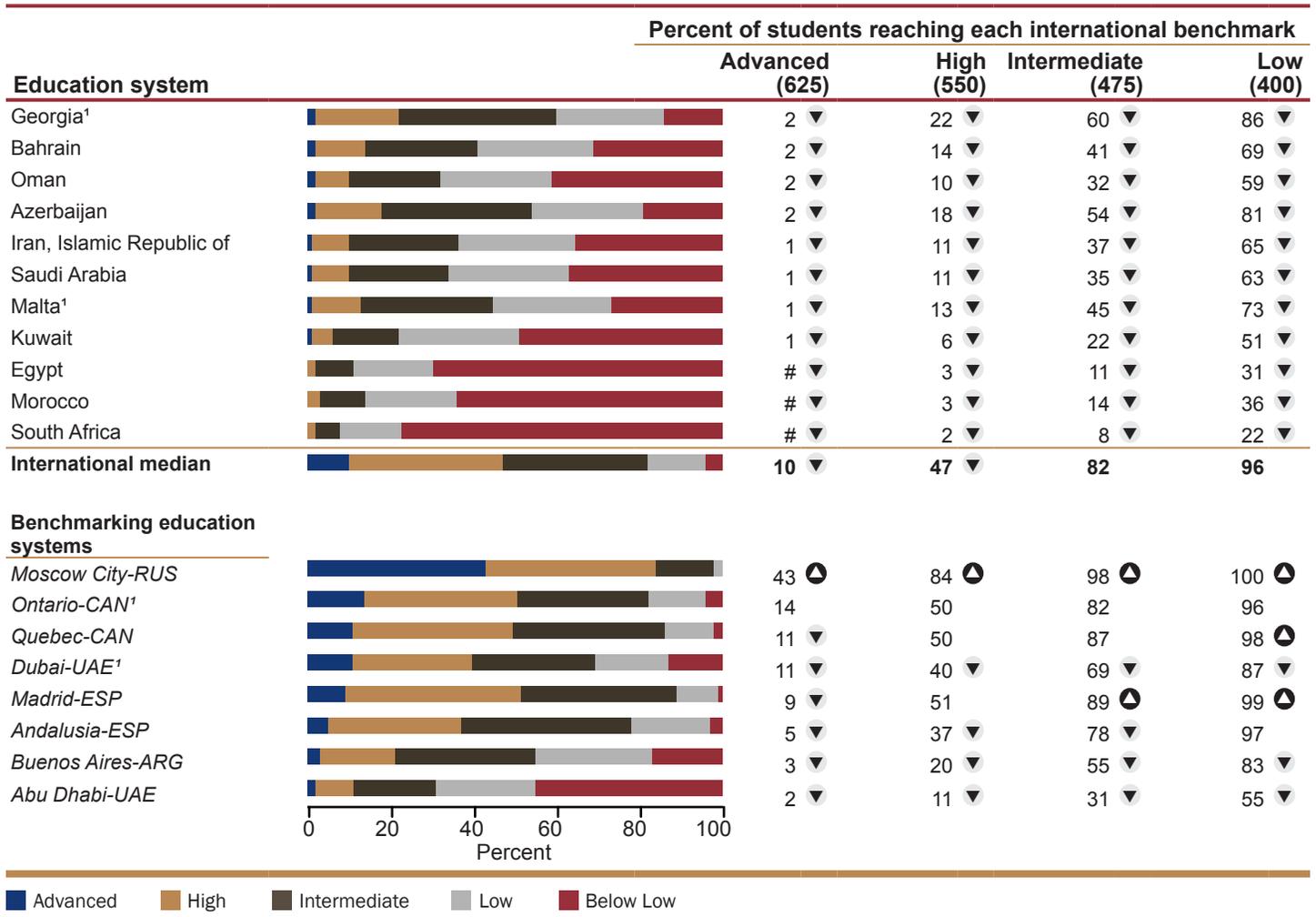
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), 2016.

Figure 1. Percentage of fourth-grade students reaching the PIRLS international benchmarks in reading, by education system: 2016



See notes at the end of the table.

Figure 1. Percentage of fourth-grade students reaching the PIRLS international benchmarks in reading, by education system: 2016—Continued



▲ Percentage is higher than the U.S. percentage at the same benchmark.
 ▼ Percentage is lower than the U.S. percentage at the same benchmark.
 # Rounds to zero.

¹ See appendix table A-1 for details on sampling issues in these education systems.

NOTE: Education systems are ordered by the percentage of students reaching the *Advanced* international benchmark. Italics indicate participants identified as a nonnational entity that represents a portion of a country. All percentages reported as higher or lower than the U.S. percentage are different at the .05 level of statistical significance. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one education system may be significant, while a large apparent difference between the United States and another education system may not be significant. The international median represents all participating PIRLS education systems, including the United States. The international median represents the percentage at which half of the education systems have that percentage of students at or above the median and half have that percentage of students below the median; benchmarking participants are not included in the international median. The standard errors of the estimates are shown in figure 1 available at https://nces.ed.gov/surveys/pirls/pirls2016/tables/pirls2016_figure01.asp. Education systems that did not administer PIRLS at the target grade are not shown; see the international report for their results. Five education systems participated in PIRLS Literacy (Egypt, Iran, Kuwait, Morocco, and South Africa); two of these education systems completed both PIRLS and PIRLS Literacy (Iran and Morocco). More detail on PIRLS Literacy is available in the technical notes at <https://nces.ed.gov/surveys/pirls/pirls2016/technotes.asp>.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), 2016.

Table 2. PIRLS purposes for reading and processes of comprehension average subscale scores of fourth-grade students, by education system: 2016

Education system	Purposes for reading		Processes of comprehension	
	Literary experience	Acquire and use information	Retrieving and straightforward inferencing	Interpreting, integrating, and evaluating
PIRLS scale centerpoint ¹	500	500	500	500
Russian Federation	579 ▲	584 ▲	581 ▲	582 ▲
Singapore ²	575 ▲	579 ▲	573 ▲	579 ▲
Ireland	571 ▲	565 ▲	566 ▲	569 ▲
<i>Northern Ireland-GBR</i>	570 ▲	561 ▲	562 ▲	567 ▲
Poland	567 ▲	564 ▲	560 ▲	570 ▲
Finland	565 ▲	569 ▲	572 ▲	562 ▲
<i>England-GBR</i>	563	556 ▲	556 ▲	561
<i>Hong Kong-CHN</i> ²	562	576 ▲	568 ▲	568 ▲
Norway	560	559 ▲	561 ▲	558
Hungary	558	551	552	557
United States²	557	543	543	555
Sweden	556	555 ▲	560 ▲	553
Latvia ²	555	561 ▲	554 ▲	562
Bulgaria	551	554 ▲	550	552
Denmark ²	551	543	550	546 ▼
Italy	549 ▼	549	547	550
<i>Chinese Taipei-CHN</i>	548 ▼	569 ▲	560 ▲	558
Lithuania	547 ▼	551	549	548
Canada ²	547 ▼	540	541	545 ▼
Australia	547 ▼	543	541	549
Netherlands ²	546 ▼	545	546	544 ▼
Czech Republic	545 ▼	541	551 ▲	538 ▼
Austria ²	544 ▼	539	550	534 ▼
Germany	542 ▼	533 ▼	546	530 ▼
Slovenia	541 ▼	544	547	539 ▼
Slovak Republic	539 ▼	531 ▼	538	531 ▼
<i>Macao-CHN</i>	536 ▼	556 ▲	549	543 ▼
Israel ²	532 ▼	529 ▼	530 ▼	530 ▼
Spain	530 ▼	527 ▼	526 ▼	529 ▼
Portugal ²	528 ▼	528 ▼	528 ▼	526 ▼
Kazakhstan	527 ▼	544	529 ▼	542 ▼
New Zealand	525 ▼	520 ▼	521 ▼	525 ▼
<i>Belgium (Flemish)-BEL</i>	524 ▼	526 ▼	526 ▼	524 ▼
France	513 ▼	510 ▼	521 ▼	501 ▼
<i>Belgium (French)-BEL</i> ²	504 ▼	490 ▼	501 ▼	494 ▼
Chile	500 ▼	485 ▼	496 ▼	491 ▼
Georgia ²	490 ▼	486 ▼	486 ▼	490 ▼
Trinidad and Tobago	478 ▼	480 ▼	483 ▼	472 ▼
Azerbaijan	466 ▼	477 ▼	477 ▼	465 ▼

See notes at the end of the table.

Table 2. PIRLS purposes for reading and processes of comprehension average subscale scores of fourth-grade students, by education system: 2016—Continued

Education system	Purposes for reading		Processes of comprehension	
	Literary experience	Acquire and use information	Retrieving and straightforward inferencing	Interpreting, integrating, and evaluating
PIRLS scale centerpoint¹	500	500	500	500
Malta ²	452 ▼	451 ▼	452 ▼	451 ▼
United Arab Emirates	440 ▼	460 ▼	448 ▼	453 ▼
Bahrain	437 ▼	453 ▼	444 ▼	446 ▼
Qatar	434 ▼	450 ▼	442 ▼	441 ▼
Iran, Islamic Republic of	430 ▼	425 ▼	429 ▼	425 ▼
Saudi Arabia	430 ▼	429 ▼	425 ▼	439 ▼
Oman	411 ▼	425 ▼	419 ▼	415 ▼
Kuwait	388 ▼	398 ▼	394 ▼	388 ▼
Morocco	353 ▼	359 ▼	364 ▼	336 ▼
Egypt	328 ▼	332 ▼	329 ▼	340 ▼
South Africa	323 ▼	314 ▼	321 ▼	308 ▼
Benchmarking education systems				
<i>Moscow City-RUS</i>	613 ▲	613 ▲	611 ▲	614 ▲
<i>Madrid-ESP²</i>	550	548	547	550
<i>Quebec-CAN²</i>	550	547	551	545 ▼
<i>Ontario-CAN</i>	549	539	539	548
<i>Andalusia-ESP</i>	526 ▼	524 ▼	522 ▼	527 ▼
<i>Dubai-UAE</i>	508 ▼	523 ▼	512 ▼	519 ▼
<i>Buenos Aires-ARG</i>	484 ▼	475 ▼	483 ▼	473 ▼
<i>Abu Dhabi-UAE</i>	406 ▼	422 ▼	413 ▼	417 ▼

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

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

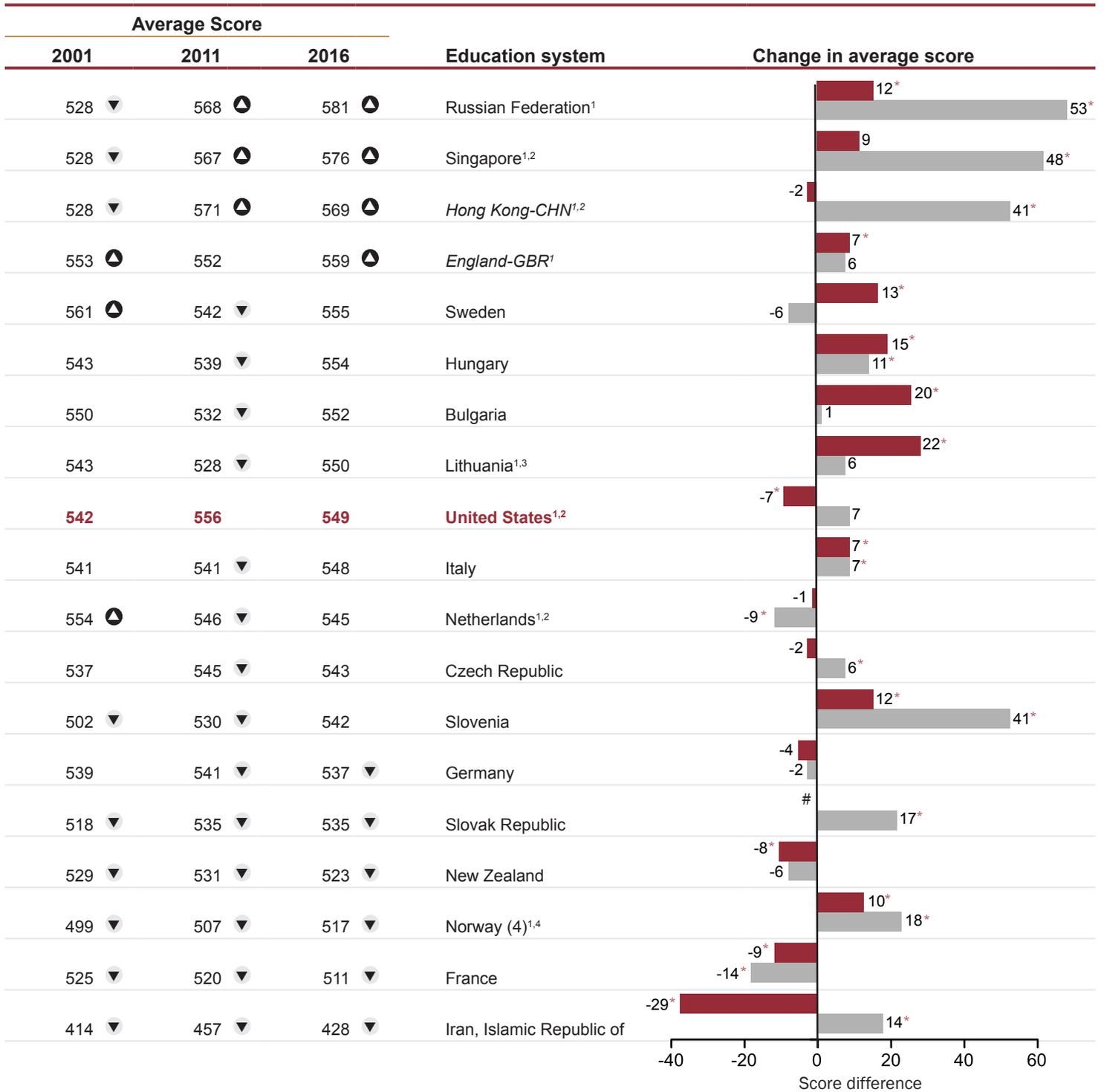
¹The PIRLS scale centerpoint is set at 500 points and represents the mean of the overall achievement distribution in 2001. The PIRLS scale is the same in each administration; thus a value of 500 in 2016 equals 500 in 2001.

² See appendix table A-1 for details on sampling issues in these education systems.

NOTE: Education systems are ordered by *literary experience* average subscale score. Italics indicate participants identified and counted in this report as a nonnational entity that represents a portion of a country. All average scores reported as higher or lower than the U.S. average score are different at the .05 level of statistical significance. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one education system may be significant, while a large apparent difference between the United States and another education system may not be significant. The standard errors of the estimates are shown in table 2 available at https://nces.ed.gov/surveys/pirls/pirls2016/tables/pirls2016_table02.asp. Education systems that did not administer PIRLS at the target grade are not shown; see the international report for their results. Five education systems participated in PIRLS Literacy (Egypt, Iran, Kuwait, Morocco, and South Africa); two of these education systems completed both PIRLS and PIRLS Literacy (Iran and Morocco). More detail on PIRLS Literacy is available in the technical notes at <https://nces.ed.gov/surveys/pirls/pirls2016/technotes.asp>.

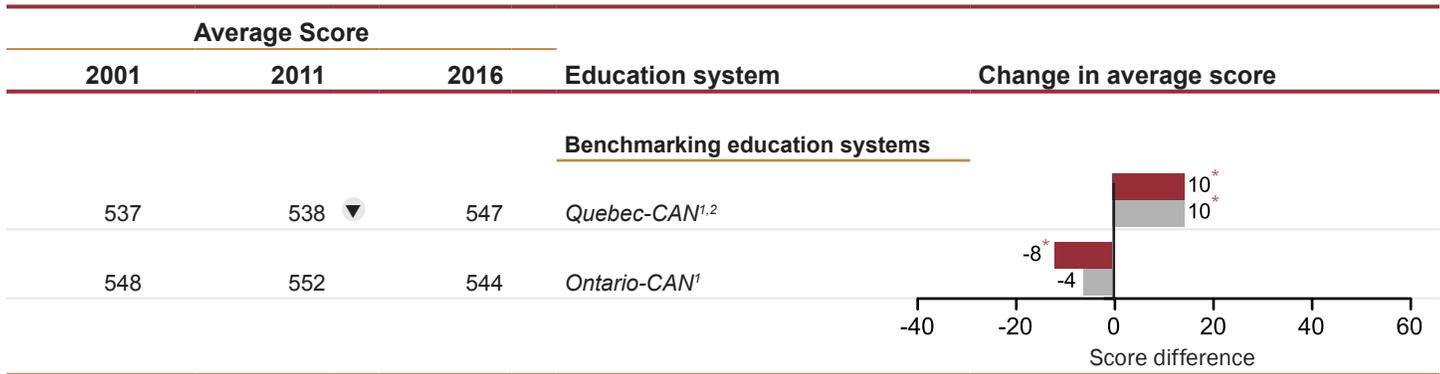
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), 2016.

Figure 2. Change in PIRLS overall reading average scale scores of fourth-grade students, by education system: 2011 to 2016 and 2001 to 2016



See notes at the end of the figure.

Figure 2. Change in PIRLS overall reading average scale scores of fourth-grade students, by education system: 2011 to 2016 and 2001 to 2016—Continued



■ Change from 2011 to 2016 ■ Change from 2001 to 2016

▲ Score is higher than U.S. average score in the same year.

▼ Score is lower than U.S. average score in the same year.

Rounds to zero.

* $p < .05$. Change in average scores is statistically significant.

¹ See appendix table A-2 for details on sampling issues in these education systems in 2001 and 2011.

² See appendix table A-1 for details on sampling issues in these education systems in 2016.

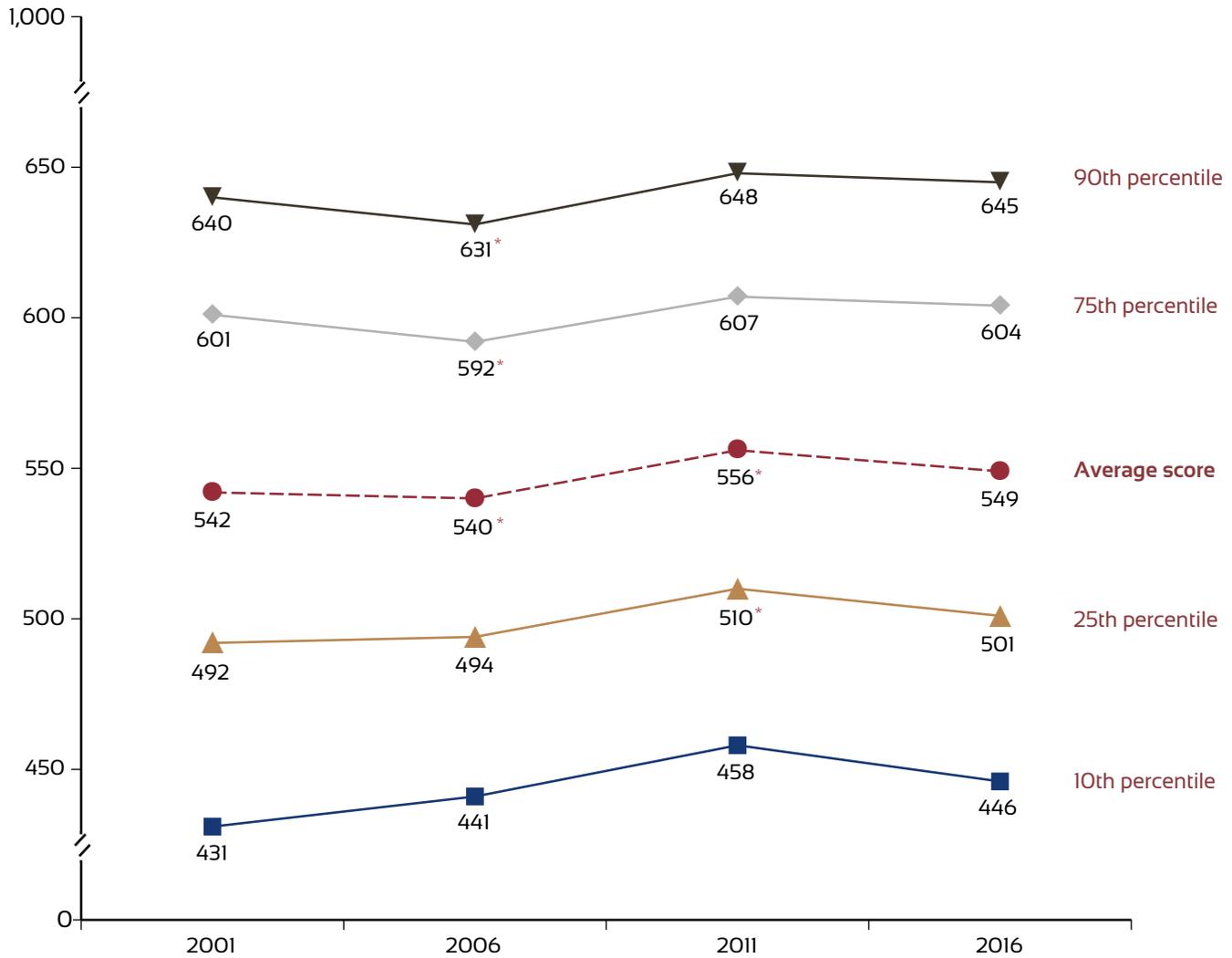
³ Trend results for Lithuania do not include students taught in Polish or Russian.

⁴ The number in parentheses indicates the grade level. For PIRLS 2016, Norway revised its assessed population to students in the fifth grade to obtain better comparisons with Sweden and Finland. However, in previous PIRLS cycles Norway assessed students in the fourth grade, which is similar to third grade in many other education systems because grade 1 in Norway is considered the equivalent of a year of kindergarten. To maintain trend with previous PIRLS cycles, in 2016 Norway also collected data from fourth-grade students, which is used in this trend table.

NOTE: Education systems are ordered by 2016 *overall* average scale scores. Italics indicate participants identified and counted in this report as a nonnational entity that represents a portion of a country. Data are not shown for some education systems because comparable data from previous cycles are not available. All average scores reported as higher or lower than the U.S. average score are different at the .05 level of statistical significance. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one education system may be significant, while a large difference between the United States and another education system may not be significant. Detail may not sum to totals because of rounding. The standard errors of the estimates are shown in figure 2 available at https://nces.ed.gov/surveys/pirls/pirls2016/tables/pirls2016_figure02.asp. Education systems that did not administer PIRLS at the target grade are not shown; see the international report for their results. In 2016, Iran participated in both PIRLS and PIRLS Literacy. More detail on PIRLS Literacy is available in the technical notes at <https://nces.ed.gov/surveys/pirls/pirls2016/technotes.asp>.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), 2001, 2011, and 2016.

Figure 3. Trends in U.S. fourth-grade students' PIRLS overall reading average scale scores and cut scores at the 10th, 25th, 75th, and 90th percentiles: 2001, 2006, 2011, and 2016

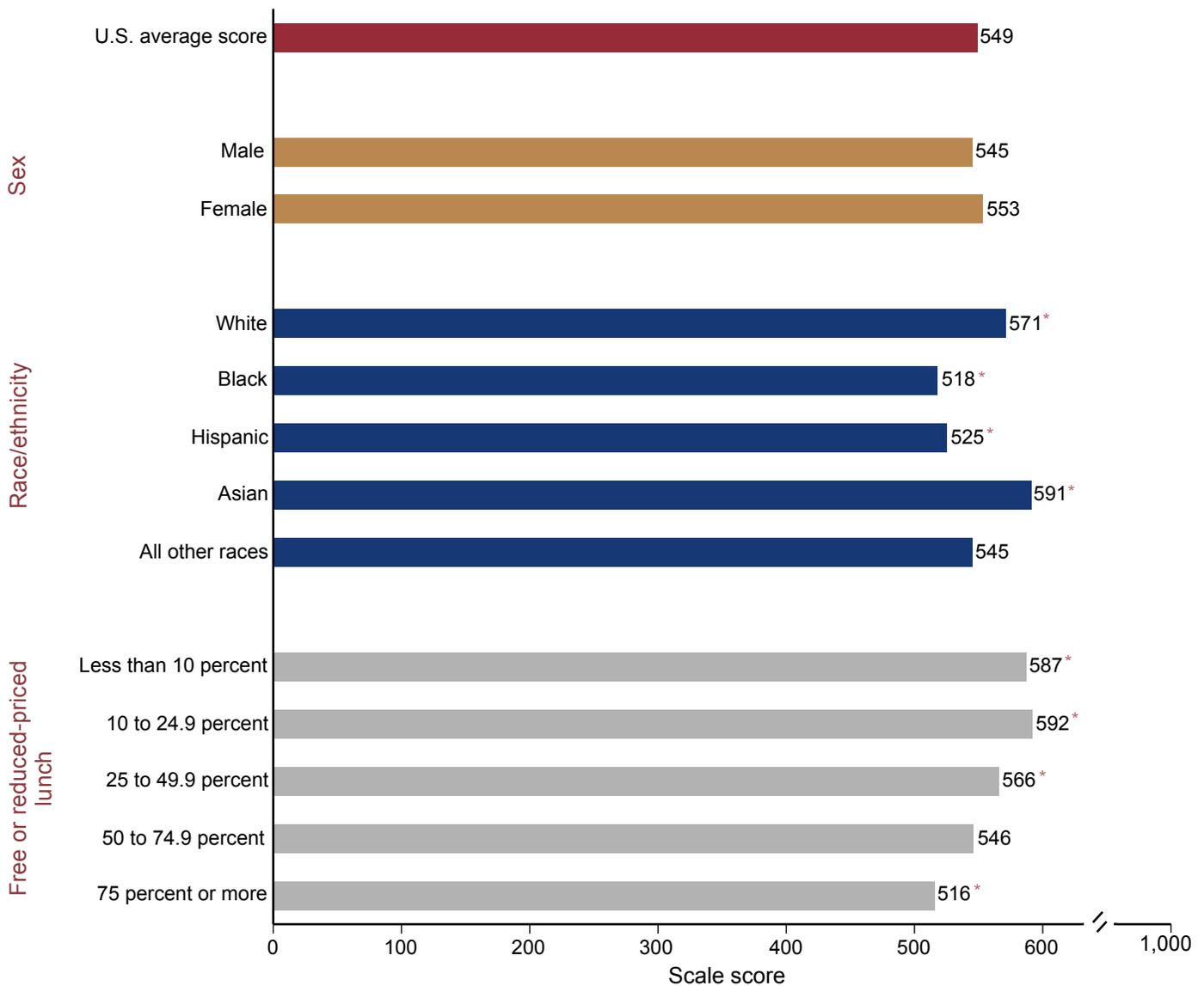


* $p < .05$. Score is significantly different from 2016 score.

NOTE: See appendix tables A-1 and A-2 for details on coverage and sampling issues in the United States for 2016 and earlier years, respectively. The standard errors of the estimates are shown in table 8 available at https://nces.ed.gov/surveys/pirls/pirls2016/tables/pirls2016_table08.asp.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), 2001, 2006, 2011, and 2016.

Figure 4. PIRLS overall reading average scale scores of U.S. fourth-grade students, by sex, race/ethnicity, and percentage of public school students eligible for free or reduced-price lunch: 2016



* $p < .05$. Significantly different from the U.S. average score.

NOTE: The U.S. met guidelines for sample participation rates only after replacement schools were included. Black includes African American, and Hispanic includes Latino. All other races includes American Indian or Alaskan Native; Native Hawaiian or other Pacific Islander; or Two or more races. Racial categories exclude Hispanic origin. Students who identified themselves as being of Hispanic origin were classified as Hispanic, regardless of their race. Analyses for free or reduced-price lunch are limited to public schools only, based on school reports of the percentage of students in public school eligible for the federal free or reduced-price lunch program. The standard errors of the estimates are shown in table 10 available at https://nces.ed.gov/surveys/pirls/pirls2016/tables/pirls2016_table10.asp.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), 2016.

ePIRLS Tables and Figures

Table 3. ePIRLS online informational reading average scale scores of fourth-grade students, by education system: 2016

Education system	Online average scale score	Education system	Online average scale score
ePIRLS scale centerpoint ¹	500	ePIRLS scale centerpoint ¹	500
Singapore ²	588 ▲	Italy	532 ▼
Norway	568 ▲	Slovenia	525 ▼
Ireland	567 ▲	Portugal ²	522 ▼
Sweden	559	Georgia ²	477 ▼
Denmark ²	558	United Arab Emirates	468 ▼
United States²	557		
<i>Chinese Taipei-CHN</i>	546 ▼	Benchmarking education systems	
Canada ²	543 ▼	<i>Dubai-UAE</i>	528 ▼
Israel ²	536 ▼	<i>Abu Dhabi-UAE</i>	431 ▼

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

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

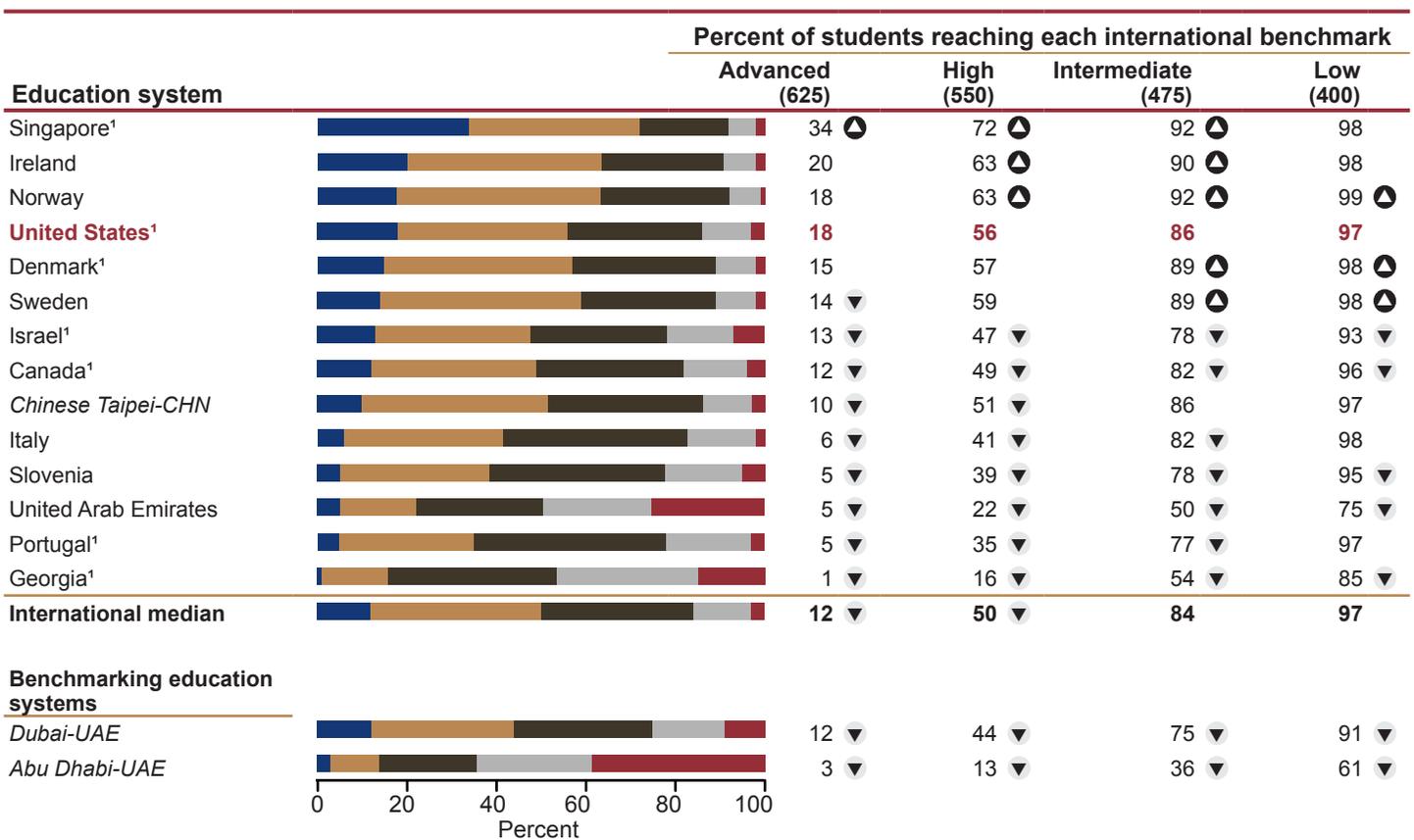
¹ The ePIRLS scale centerpoint is set at 500 points and represents the mean of the overall PIRLS achievement distribution in 2001. ePIRLS results are reported on the PIRLS reading achievement scale.

² See appendix table A-1 for details on sampling issues in these education systems.

NOTE: Education systems are ordered by *online* average scale score. Italics indicate participants identified and counted in this report as a nonnational entity that represents a portion of a country. All average scores reported as higher or lower than the U.S. average score are different at the .05 level of statistical significance. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one education system may be significant while a large apparent difference between the United States and another education system may not be significant. The standard errors of the estimates are shown in table 11 available at https://nces.ed.gov/surveys/pirls/pirls2016/tables/pirls2016_table11.asp.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), 2016.

Figure 5. Percentage of fourth-grade students reaching the ePIRLS international benchmarks in online informational reading, by education system: 2016



■ Advanced
 ■ High
 ■ Intermediate
 ■ Low
 ■ Below Low

▲ Percentage is higher than the U.S. percentage at the same benchmark.

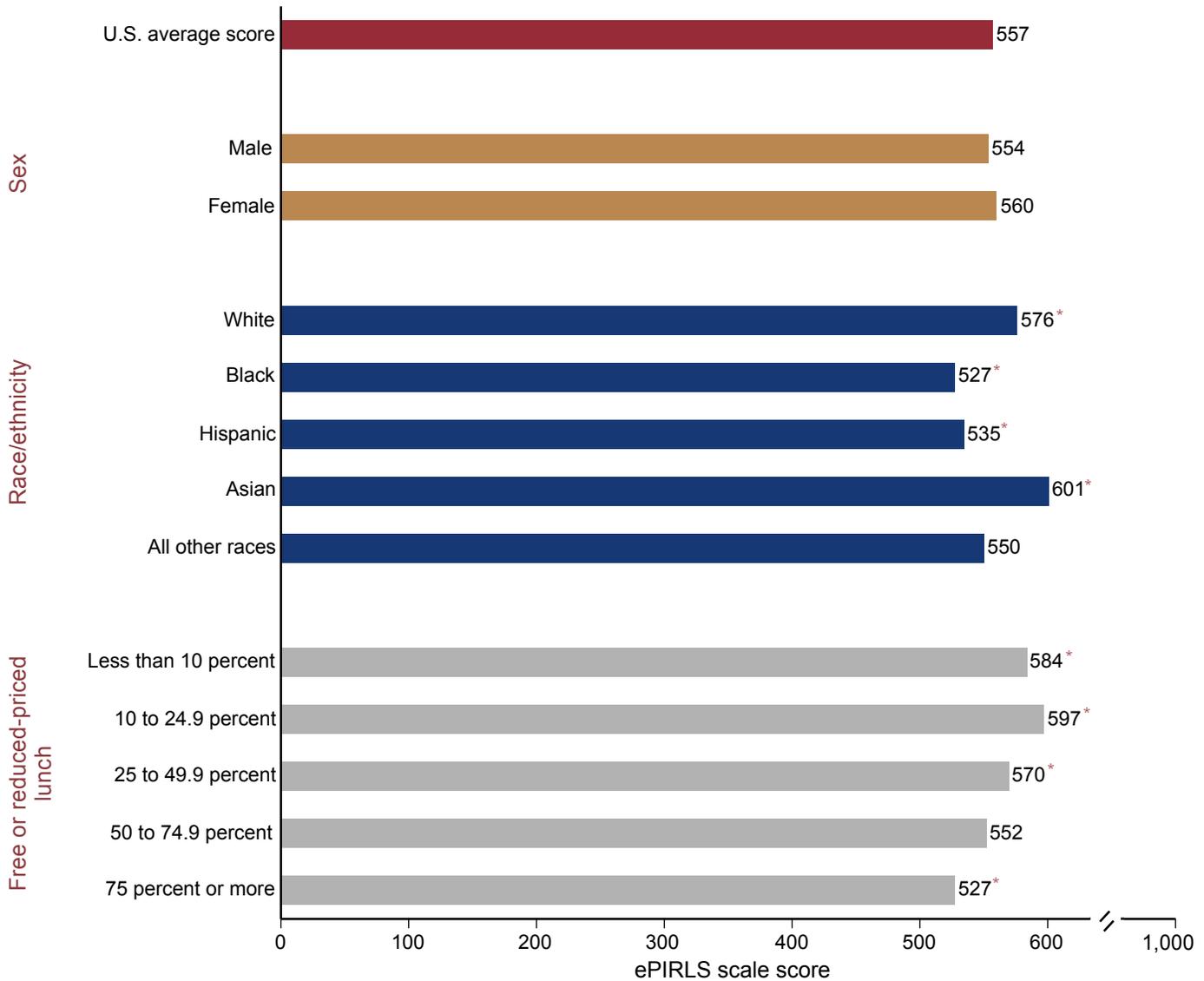
▼ Percentage is lower than the U.S. percentage at the same benchmark.

¹ See appendix table A-1 for details on sampling issues in these education systems.

NOTE: Education systems are ordered by the percentage of students reaching the *Advanced* international benchmark. Italics indicate participants identified and counted in this report as a nonnational entity that represents a portion of a country. All percentages reported as higher or lower than the U.S. percentage are different at the .05 level of statistical significance. The tests for significance take into account the standard error for the reported difference. Thus, a small difference between the United States and one education system may be significant while a large difference between the United States and another education system may not be significant. The international median represents all participating ePIRLS education systems, including the United States. The international median represents the percentage at which half of the education systems have that percentage of students at or above the median and half have that percentage of students below the median. Benchmarking participants are not included in the international median. The standard errors of the estimates are shown in figure 3 available at https://nces.ed.gov/surveys/pirls/pirls2016/tables/pirls2016_figure03.asp.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), 2016.

Figure 6. ePIRLS online informational reading average scale scores of U.S. fourth-grade students, by sex, race/ethnicity, and percentage of public school students eligible for free or reduced-price lunch: 2016



* $p < .05$. Significantly different from the U.S. average score.

NOTE: The U.S. met guidelines for sample participation rates only after replacement schools were included. Black includes African American, and Hispanic includes Latino. All other races includes American Indian or Alaskan Native; Native Hawaiian or other Pacific Islander; or Two or more races. Racial categories exclude Hispanic origin. Students who identified themselves as being of Hispanic origin were classified as Hispanic, regardless of their race. Analyses for free or reduced-price lunch are limited to public schools only, based on school reports of the percentage of students in public school eligible for the federal free or reduced-price lunch program. The standard errors of the estimates are shown in table 15 available at https://nces.ed.gov/surveys/pirls/pirls2016/tables/pirls2016_table15.asp.

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), 2016.

References

- Martin, M.O., Mullis, I.V.S., and Hooper, M. (Eds.). (2017). *Methods and Procedures in PIRLS 2016*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement. Retrieved from <https://timssandpirls.bc.edu/publications/pirls/2016-methods.html>.
- Mullis, I.V., and Martin, M.O. (2015). *PIRLS 2016 Assessment Framework, 2nd Edition*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement.
- Mullis, I.V.S., Martin, M.O., Foy, P., and Hooper, M. (2017a). *PIRLS 2016 International Results in Reading*. Retrieved from Boston College. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement. Retrieved from <https://timssandpirls.bc.edu/pirls2016/international-results/>.
- Mullis, I.V.S., Martin, M.O., Foy, P., and Hooper, M. (2017b). *ePIRLS 2016 International Results in Online Informational Reading*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement. Retrieved from <https://timssandpirls.bc.edu/pirls2016/international-results/>.

Appendix A. Technical Notes

This appendix briefly describes features of the PIRLS 2016 assessment, with a particular focus on its implementation in the United States. For further details about the assessment and any of the topics discussed here, see the fuller description of technical notes on the NCES PIRLS website at <https://nces.ed.gov/surveys/pirls/pirls2016/> as well as the IEA's PIRLS 2016 Technical Report (Martin, Mullis, and Hooper 2017).

Sampling and Response Rates

The IEA required all participating education systems to adhere to the PIRLS 2016 technical standards (Martin, Mullis, and Hooper 2017), which provided detailed information about the target population, sampling, response rates, translation, assessment administration, and data submission. All students enrolled in the grade that represents the fourth year of formal schooling constitute the desired international target population, provided that their mean age at the time of testing is at least 9.5 years and that the grade level is calculated from the first year of the International Standard Classification of Education (ISCED) Level 1. (ISCED Level 1 is the equivalent of grade 1 in the United States.) For most education systems, the target grade is grade 4 or its national equivalent. All students enrolled in the target grade, regardless of their age, belong to the desired international target population. Additional detail can be found in the detailed technical notes, available online at <https://timssandpirls.bc.edu/publications/pirls/2016-methods.html>.

A representative sample of fourth-grade students was selected using a stratified two-stage systematic sample, with sampling probabilities proportional to the estimated number of fourth-grade students in the school based on grade enrollments. Classrooms within sampled schools were selected using sampling software provided by the TIMSS & PIRLS International Study Center.

PIRLS guidelines call for a minimum of 150 schools to be sampled, with a minimum of 4,000 students assessed. The U.S. PIRLS 2016 national school sample consisted of 176 schools, which was higher than the international sampling minimum of 150 to offset anticipated school nonresponse and ineligibility. A total of 158 U.S. schools agreed to participate in PIRLS 2016, including 131 from the original sample and 27 sampled as replacements for nonparticipating schools from the original sample. Of the 158 U.S. schools that participated in PIRLS, 153 also participated in ePIRLS. In total, 4,425 U.S. students participated in PIRLS and 4,090 of these students also participated in ePIRLS.

The overall weighted school response rate was 92 percent for PIRLS and 89 percent for ePIRLS. The overall weighted student response rate was 94 percent for PIRLS and 90 percent for ePIRLS, and the overall U.S. student exclusion rate was 4.8 percent for PIRLS and 4.9 percent for ePIRLS. Additionally, as required by NCES standards, a nonresponse bias analysis was conducted because the U.S. school-level response rate (before substitution) for PIRLS fell below 85 percent. The purpose of this analysis was to examine whether the participation status of schools was related to various characteristics and thus introduced the potential for bias in the results. The results suggested some potential for nonresponse bias in the U.S. sample (prior to substitution) based on the characteristics studied. It also suggested that the use of substitute schools reduced the potential for bias in all variables except Census region, such that schools in the Central Region are represented at higher rates among participating schools relative to all eligible schools. Nonresponse bias results for the final sample (after substitution) indicated that schools with larger fourth-grade enrollment were more likely to respond than those with smaller fourth-grade enrollment. Applying nonresponse adjustments lowered estimates of bias, relative bias, and effect size associated with bias. After the application of school nonresponse adjustments, bias estimates remained on a few variables; however, the magnitudes of these bias estimates were quite small. Variables that indicated bias include Census region and percentage school enrollment of American Indian or Alaska Native students.

As indicated by the footnotes in the cross-education system tables and figures in this report, there were sampling or other issues in the United States and 19 education systems. For the current administration, these specific issues are detailed in table A-1. For the 2001, 2006, and 2011 administrations, these issues are addressed in table A-2.

See <https://timssandpirls.bc.edu/publications/pirls/2016-methods.html> for the PIRLS international sampling guidelines and requirements regarding accommodations, exclusions, and response rate requirements, as well as the response rates of all participating education systems.

Table A-1. Sampling or other issues for PIRLS and ePIRLS, by education system: 2016

	Coverage			Sampling		
	National Target Population does not include all of the International Target Population.	National Defined Population covers 90 to 95 percent of the National Target Population.	National Defined Population covers less than 90 percent of National Target Population (but at least 77 percent).	Met guidelines for sample participation rates only after replacement schools were included.	Nearly satisfied guidelines for sample participation rates after replacement schools were included.	Did not satisfy guidelines for sample participation rates.
PIRLS						
Austria		•				
<i>Belgium (French)-BEL</i>		•				
Canada	•	•				
Denmark		•				
Georgia	•					
<i>Hong Kong-CHN</i>		•		•		
Israel			•			
Latvia		•				
Malta		•				
Netherlands				•		
Portugal		•				
Singapore			•			
United States				•		
<i>Quebec-CAN</i>						•
<i>Madrid-ESP</i>		•				
ePIRLS						
Canada	•	•				
Denmark						•
Georgia	•					
Israel			•			
Portugal		•				
Singapore			•			
United States				•		

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), 2016.

Table A-2. Sampling or other issues for PIRLS, by education system: 2001, 2006, and 2011

	Coverage				Sampling		
	National Target Population does not include all of the International Target Population.	National Defined Population covers 90 to 95 percent of the National Target Population.	National Defined Population covers less than 90 percent of National Target Population (but at least 77 percent).	National Defined Population covers less than 80 percent of National Target Population	Met guidelines for sample participation rates only after replacement schools were included.	Nearly satisfied guidelines for sample participation rates after replacement schools were included.	Did not satisfy guidelines for sample participation rates.
PIRLS							
<i>England-GBR (2001)</i>		●			●		
<i>England-GBR (2011)</i>					●		
<i>Hong Kong-CHN (2011)</i>			●				
Lithuania (2001)	●						
Lithuania (2011)	●	●					
Netherlands (2001)					●		
Netherlands (2011)					●		
Norway(4) (2011)						●	
Russian Federation (2001)		●					
Singapore (2011)		●					
United States (2001)					●		
United States (2006)					●		
United States (2011)		●					
<i>Ontario-CAN (2001)</i>	●						
<i>Ontario-CAN (2011)</i>		●					
<i>Quebec-CAN (2001)</i>	●						

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), 2016.

Assessment and Questionnaires

The 2016 assessment instruments were developed by international experts and PIRLS international test developers; the instruments included items submitted by participating education systems. Each reading passage and its accompanying items were assigned to a block and then systematically distributed among individual student assessment booklets. The 2016 assessment consisted of 12 passages or blocks, distributed across 15 booklets (with questions printed after each passage) and one reader (presented in a magazine-type format with the questions in a separate booklet). Six blocks were included from previous PIRLS assessments to provide a link for measuring trends in reading achievement. Approximately 39 percent of assessment score points were based on multiple choice questions and 61 percent of the assessment score points were based on open ended questions. Open-ended items were coded by trained scorers following international coding guidelines.

All participating education systems field-tested the assessment instruments in spring 2015. After the field test, passages and items were dropped if they did not meet the established measurement criteria or were otherwise found to include intrinsic biases. Dropped passages and items were not included in the main assessment.

For the main assessment, each student received a booklet containing two reading passages: one literary experience passage and one informational passage. Students had 40 minutes to complete each passage and all questions about that passage, with a 5- to 10-minute break between each passage. Then they received a questionnaire asking about their backgrounds, attitudes, and experiences in school. Students had 30 minutes to complete the questionnaire.

The ePIRLS assessment presented students with a simulated internet environment and five possible tasks. Typically, students who participated in the ePIRLS assessment took the assessment on the day following PIRLS. Like PIRLS, ePIRLS was delivered in two 40-minute segments with a 5-minute break between them, followed by a brief questionnaire about computer use.

As part of the PIRLS study, principals were asked to complete a questionnaire about their school's structure, resources, instruction, climate, and policies. Language Arts teachers of assessed classrooms were also asked to complete a questionnaire about their experience, available resources, and instructional practices at the school.

See <https://nces.ed.gov/surveys/pirls/pirls2016/> for more information about the PIRLS 2016 assessment design and questionnaires.

Reporting Results

In PIRLS 2016, results are generally reported in two ways: scale scores and international benchmarks of achievement. Scores on PIRLS for each administration since 2001 have been scaled to range from 0 to 1,000, with an international centerpoint of 500 and a standard deviation of 100. By centering results in this manner, comparisons can be made from 2001 to 2006, to 2011, and now to 2016. The ePIRLS scale also ranges from 0 to 1,000, with an international centerpoint of 500 and a standard deviation of 100. This scaling facilitates comparisons across education systems of overall printed reading achievement and overall online reading achievement.

Along with scale scores, PIRLS has international benchmarks that group achievement into four levels: *Advanced*, *High*, *Intermediate*, and *Low*. The distribution of student scores—and the kinds of skills and knowledge that students demonstrate—determines the score cut-points for these benchmarks. The benchmarks offer an interpretation of what the scale scores mean using scale anchoring. Scale anchoring involves selecting benchmarks on the PIRLS achievement scales to be described in relation to student performance. Items within each of these benchmarks identify what students are likely to answer correctly; experts then examine the data to provide a sense of what students know and can do. Examples of items at each benchmark level are provided at <https://timssandpirls.bc.edu/pirls2016/frameworks.html>.

Sampling Weights and Standard Errors

The use of sampling weights is necessary to compute statistically sound estimates. Survey weights adjust for the probability of selection of individual schools and students, for school or student nonresponse, and for errors in estimating the size of a school or the number of fourth-grade students in a school at the time of sampling. As with any study, estimates produced using data from PIRLS 2016 are subject to two types of error: nonsampling errors and sampling errors. The sources of nonsampling errors are typically problems such as unit and item nonresponse, differences in respondents' interpretations of the meaning of survey questions, and mistakes in data preparation. Sampling errors arise when a sample of the population, rather than the whole population, is used to estimate some statistic. Different samples from the same population would likely produce somewhat different estimates of the statistic in question. This uncertainty is referred to as sampling variance and is usually expressed as the standard error of a statistic estimated from sample data. Standard errors for all statistics reported in this report are available online at <https://nces.ed.gov/surveys/pirls/pirls2016/>.

Statistical Comparisons

The comparisons made in this report have been tested for statistical significance. For example, in the commonly made comparison of international averages to U.S. averages, tests of statistical significance were used to establish whether or not the observed differences from the U.S. average were statistically significant. In all instances, the tests for significance used were standard *t* tests. These fell into two categories according to the nature of the comparison being made: comparisons of independent samples and comparisons of nonindependent samples. A difference is “significant” if the probability associated with the *t* test is less than .05. If a test is significant, it implies that the difference in the observed value in the sample represents a real difference in the population. No adjustments were made for multiple comparisons.

More information on statistical tests and more detailed information on the implementation of PIRLS in the United States are available on the NCES website at <https://nces.ed.gov/surveys/pirls/pirls2016/>.



www.ed.gov



ies.ed.gov