Research Note

Synthesizing NAEP and International Large-Scale Assessment Score Trends

A Pattern of Diverging Performance

Samantha Burg, Maria Stephens, Lydia Malley, and Frank Fonseca

OCTOBER 2022

American Institutes for Research® | AIR.ORG

AIR®
Advancing Evidence.
Improving Lives.

Research Note

Synthesizing NAEP and International Large-Scale Assessment Score Trends

A Pattern of Diverging Performance

OCTOBER 2022

Authors:
Samantha Burg, Maria Stephens, Lydia Malley, and Frank Fonseca

AIR® Headquarters
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
+1.202.403.5000 | **AIR.ORG**

Contents

Figures

Tables

Abstract

Nearly every year, the U.S. Department of Education releases new data about U.S. student performance, from the National Assessment of Educational Progress (NAEP) or one of the several international large-scale assessments (ILSAs) in which the United States participates. The results are important and informative, but the need to release data in a timely manner leaves little time to reflect upon the studies together, even though they cover overlapping grades and subjects. This study brings together results from NAEP and three ILSAs to examine long-term, intermediate, and recent score trends in reading, mathematics, and science for U.S. students in 4th grade, in 8th grade, and at 15 years old. The analysis finds a relatively consistent pattern across these assessments: performance between high- and low-performing students is diverging. Understanding this pre-pandemic pattern provides a baseline picture of U.S. student performance that will be essential for accurately understanding post-pandemic findings from studies that are soon to come.

Introduction

With U.S. schools limited to remote learning for much of the 2020–21 school year and the COVID-19 pandemic still ongoing in the 2021–22 school year, one of the biggest questions that education researchers and others have is this: How has the pandemic affected students' learning outcomes? In the coming years, there will be scores of studies looking to answer this question. Important for answering it accurately, however, is having a nuanced baseline picture of students' performance prior to the pandemic. For example, we expect to find that the pandemic will have had some negative effects on students' performance, but did the pandemic create new trends or exacerbate existing ones? Do the effects differ for different groups of students, including those already at one end of the achievement distribution or the other?

This Research Note reports the results of a study that aimed to triangulate across data sources and examine long-term, intermediate, and recent score trends for U.S. students in the core subjects of reading, mathematics, and science. It brings together data from the United States' National Assessment of Educational Progress (NAEP) and three international large-scale assessments (ILSAs) in which the United States participates: the Progress in International Reading Literacy Study (PIRLS), the Trends in International Mathematics and Science Study (TIMSS), and the Program for International Student Assessment (PISA). This triangulation across studies extends the lessons we learned from the regularly released study-specific reports and

allows a more comprehensive look at U.S. student performance over time that can serve as a pre-pandemic baseline.

Methods and Data Sources

This study examines trends in U.S. student performance using data from four large-scale assessments that regularly collect nationally representative data on U.S. students (including students in both public and private schools):

- NAEP, which provides data on 4th-, 8th-, and 12th-grade students in reading, mathematics, science, and a host of other subjects;
- PIRLS, which is an international study that provides data on 4th-grade students in reading literacy;
- TIMSS, which is an international study that provides data on 4th- and 8th-grade students in mathematics and science and, less frequently, on 12th-grade students with TIMSS Advanced; and
- PISA, which is an international study that provides data on 15-year-old students in reading, mathematics, and science literacy and other, optional domains.

NAEP, also known as "the Nation's report card," is a congressionally mandated project administered by the National Center for Education Statistics (NCES) within the U.S. Department of Education's Institute of Education Sciences (IES). By law, NCES develops and administers NAEP and reports NAEP results, while the National Assessment Governing Board (NAGB) is responsible for setting the assessment schedule, developing the frameworks that provide the blueprint for the content and design of the assessment, and setting achievement levels. PIRLS and TIMSS are sponsored by the International Association for the Evaluation of Educational Achievement (IEA) and conducted in the United States by NCES. PISA is coordinated by the Organization for Economic Cooperation and Development (OECD) and is also conducted in the United States by NCES.

Each of these assessments is based on its own framework, which describes the specific purpose of the program; the knowledge and skills to be measured; the target distribution of items across knowledge and skills; and the general test design (e.g., Mullis & Martin, 2015, 2017; NAGB, 2019a, 2019b, 2019c; OECD, 2019). What is measured by these different assessments broadly overlaps, as has been shown in numerous content comparison studies (e.g., see here and here), but each also has its own unique aspects that contribute to a multifaceted picture of U.S. student performance overall. For example, NAEP, TIMSS, and PIRLS are designed as tests of

curricular achievement at specific grade levels, whereas PISA is designed as a test of students' ability to apply the knowledge and skills they have cumulatively gained by age 15 in real-world scenarios and contextualized problems (see figure 1). Therefore, even overlapping content may be measured in different ways or with different emphases. Another distinction of note is NAEP's relatively larger sample sizes and consequent higher levels of precision, which are necessary for its goal of tracking national- and state-level subgroup differences.

Figure 1. Overview of large-scale student assessments in the study

Feature	PIRLS	TIMSS	NAEP	PISA
Target population	4th-graders	4th- and 8th-graders	4th-, 8th-, and 12th-graders	15-year-olds
Subject(s) assessed	Reading literacy	Mathematics and science	Reading, mathematics, science, and other subjects	Reading, mathematics, and science literacy
Focus of the assessment framework	Curricula common across countries		Curricula in the United States	Real-world applications of knowledge and skills
Initial year[1]	2001	1995	Early 1990s	2000
Frequency[2]	Every 5 years	Every 4 years	Every 2 years	Every 3 years
School sample size (approximate)	160	280 per grade	7,500	300
Student sample size (approximate)	4,400	8,800 per grade	148,000 per subject and grade	4,800
Scale(s)[3]	0–1,000	0–1,000	0–500	0–1,000
Sponsoring organization	IEA International Association for the Evaluation of Educational Achievement		NCES National Center for Education Statistics (implementation) NAGB National Assessment Governing Board (policy oversight)	OECD Organization for Economic Cooperation and Development

[1] This describes the first year in which data were collected for the given assessment. This may or may not coincide with the year used in this research to identify "long-term trend." See figure 2 for additional detail.
[2] This describes the frequency with which the given assessment is administered. For NAEP, "every 2 years" is a general rule; for some subjects and grades, the frequency may be 4 or more years. In TIMSS, the 4th-grade assessment was not administered in 1999. In PISA, all three subjects are administered every 3 years but what is considered a "major" versus a "minor" domain rotates, with a "major" domain occurring every 9 years.
[3] Scales are by subject (e.g., there is a composite scale for mathematics and one for science in TIMSS).
Note: International assessments include the Progress in Reading Literacy Study (PIRLS), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA). NAEP is the U.S. National Assessment of Educational Progress.

This study focuses on the subjects and ages and grades in common across all four assessments: reading, mathematics, and science for 4th-graders at the elementary level and for 8th-graders and 15-year-olds, whose modal grade is the 10th, at the secondary level. The study includes students in public and private schools across the United States. The study excludes NAEP's 12th-grade population because neither PISA nor TIMSS Advanced can be considered an international counterpart for the purpose of this study. Regarding PISA, prior content comparison studies have found that, while the cognitive complexity of PISA's assessments is high, their content is more often mapped to U.S. 8th-grade content than to 12th-grade content (e.g., Gattis et al., 2016; NCES, 2008, 2010). Regarding TIMSS Advanced, its 12th-grade assessments focus on advanced mathematics and physics, which cover only a subset of NAEP's mathematics and science framework, and they are only given to the subset of the 12th-grade population who are taking or have taken advanced coursework.

The measures examined in the study include U.S. students' average score and 10th- and 90th-percentile scores. The 10th-percentile score is the score below which the lowest 10 percent of students score, and the 90th percentile score is the score above which the highest 10 percent of students score. In this Research Note, these groups may also be referred to as low- and high-performing students, or the bottom and top ends of the distribution, respectively.

Score trends are identified by comparing the most recent available scores (average, 10th, and 90th percentile) for each assessment, grade, and subject to the analogous scores from three previous administrations of the assessments to identify a long-term trend, an intermediate trend, and a recent trend for each. For all assessments, the most recent available scores are pre-pandemic: 2016 for PIRLS, 2018 for PISA, and 2019 for TIMSS and NAEP.[1] The comparison years for the trend periods for each assessment were selected so that (1) the long term would be equivalent to a 15- to 20-year time span; (2) the intermediate term would be equivalent to roughly a 10-year time span; (3) the recent term would be a time span of 5 years or less; and (4) cross-assessment comparisons would have data points close in time to one another for the given grade and subject (see figure 2).

[1] This study was completed prior to the public availability of NAEP 2022 reading and mathematics main study data, which (like this report) were released in Fall 2022.

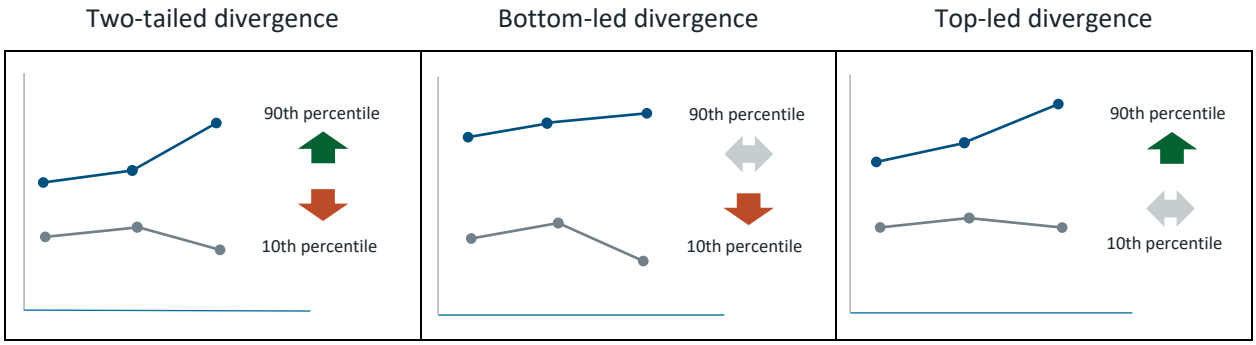
Figure 2. Overview of long-term, intermediate, and recent trends in the study, by grade, subject, and assessment

Grade	Subject	Assessment	Years examined			
4th grade	Reading	NAEP	2002	2009	2017	2019
		PIRLS	2001	2006	2011	2016
	Mathematics	NAEP	2003	2009	2017	2019
		TIMSS	2003	2011	2015	2019
	Science	NAEP[1]	—	2009	2015	2019
		TIMSS	2003	2011	2015	2019
8th grade/ 15-year-olds	Reading	NAEP	1998	2009	2017	2019
		PISA	2000	2009	2015	2018
	Mathematics	NAEP	2000	2009	2017	2019
		TIMSS	1999	2011	2015	2019
		PISA[2]	2003	2009	2015	2018
	Science	NAEP[1]	—	2009	2015	2019
		TIMSS	1999	2011	2015	2019
		PISA[2]	—	2009	2015	2018

[1] The NAEP science framework was revised in 2009, so there is no long-term trend period to report in this study.
[2] The first year in which mathematics was administered as a "major" domain in PISA was 2003 and so it is the starting point for its long-term trend. The first year in which science was administered as a "major" domain in PISA was 2006. However, since using that as a starting point for long-term trend would create a time span that is less than that for other assessments, we forgo a long-term data point in this study.

This study is interested particularly in what is happening for U.S. students performing at the bottom and top ends of the distribution and whether score changes in these groups are contributing to diverging performance at the tails of the distribution (i.e., the 10th and 90th percentiles). Divergence can occur when each end of the distribution is moving in the opposite direction (i.e., two-tailed), or when one end is unchanged while either the top end rises or the bottom end falls (see figure 3). Although not all divergence is created equal—in some cases, at least high-performing students are improving—in all cases, it is a bad sign for educational equity, as it indicates that score differences between low- and high-performing students may be growing wider.

Figure 3. Examples of diverging student performance



Additionally, the study also examined score trends for each of the eight other countries that participated in at least one subject of all of the study's ILSAs and selected comparison years, in order to place the U.S. results and patterns in a global context. These countries (Hong Kong-CHN, Hungary, Italy, Lithuania, New Zealand, the Russian Federation, Singapore, and the United Kingdom/England-GBR)[2] represent a range of performance levels, geographic regions, and language groups.

All estimates were calculated using the data analysis tools publicly available through the National Center for Education Statistics, including the NAEP Data Explorer (NDE) and the International Data Explorer (IDE). The statistical significance of changes in average scores and 10th and 90th percentile scores were also generated by these tools, based on independent *t* tests using unrounded data; all stated changes are significant at the $p < .05$ level. In the figures and tables, an asterisk (*) next to the score for a given year indicates a statistically significant difference from the score for the most recent assessment year (i.e., 2016 for PIRLS, 2018 for PISA, and 2019 for TIMSS and NAEP). The trend lines shown in the figures highlight the years of interest for this Research Note but are based on all available data points, including those from any intervening administrations.

Results

Results are presented first for 4th-graders and then for 8th-graders and 15-year-olds. Within the grade or age level, results are presented by subject. Following the two grade-/age-specific sections, results are summarized across grades for the United States as a whole, and

[2] In this study, PIRLS and TIMSS results are reported for England-GBR, whereas PISA results are reported for the United Kingdom (of which England-GBR is a part). This is because IEA studies allow for the participation of jurisdictions within countries, whereas the OECD requires a country-level score for member countries. Hong Kong and England are appended with their three-letter country codes (CHN and GBR, respectively) per IEA reporting conventions.

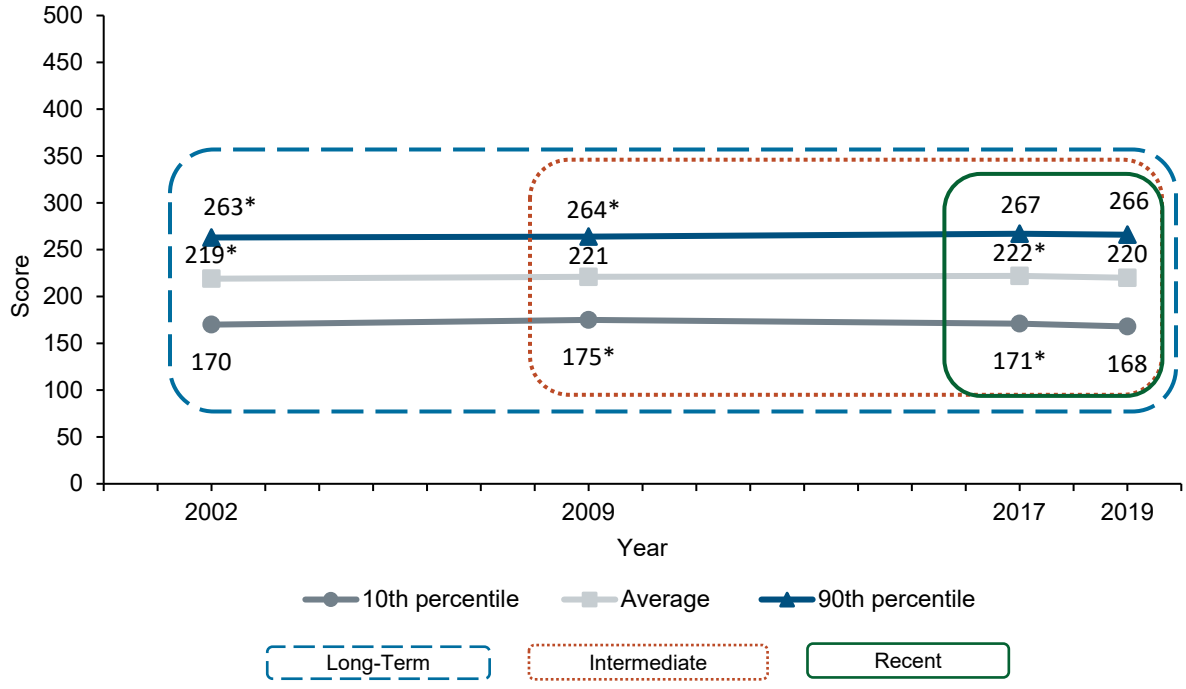
international patterns are examined. The results section also includes notes on the limitations of the analyses and considerations for future work.

Trends in 4th-grade scores

Reading

Looking first at the NAEP results, the reading scores of 4th-grade students have generally shown improvement over the long term (see figure 4). Both the average and 90th percentile scores were higher in 2019 than in 2002, while the 10th percentile score was statistically unchanged. The 90th percentile score was also higher in 2019 than it was in 2009, over the intermediate term. However, beginning in 2009, for lower-performing students (i.e., at the 10th percentile) and in 2017 for students performing at the average, NAEP 4th-grade reading scores trended downward to 2019, showing a bottom-led divergence of scores.

Figure 4. Trends in U.S. 4th-graders' average scores and 10th and 90th percentile scores in NAEP reading: Selected years 2002 to 2019



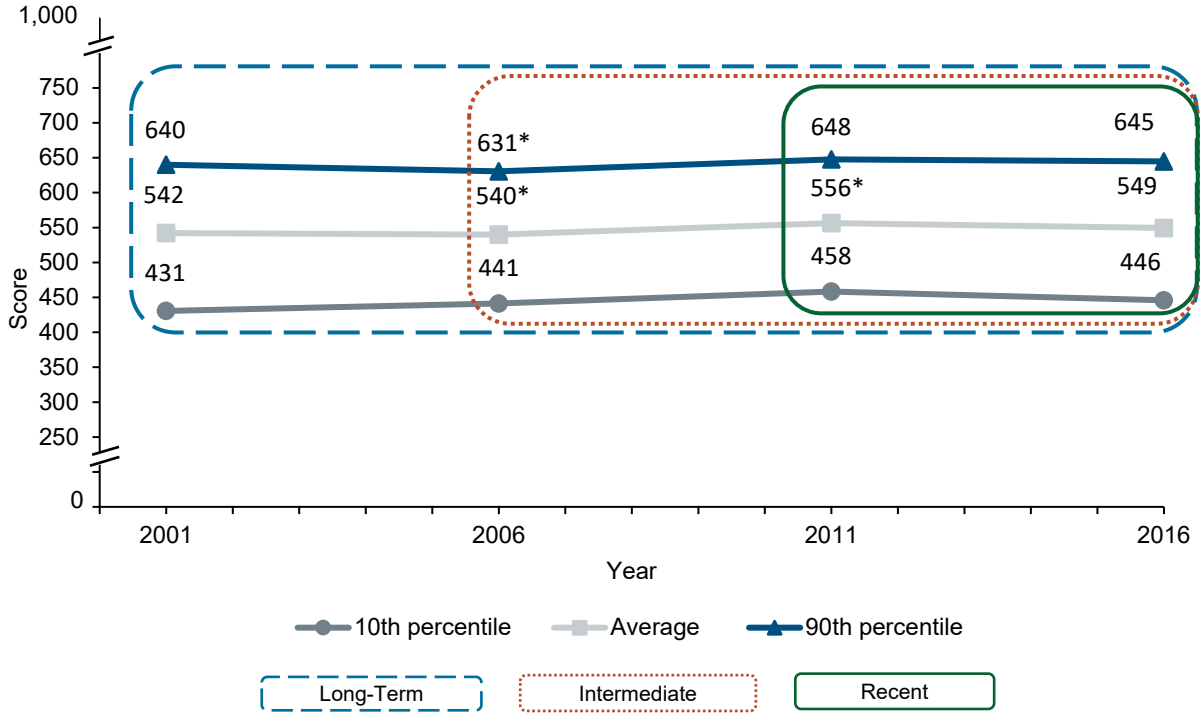
* $p < .05$. Score is significantly different from 2019 score.

Note: The boxes represent different trend periods: dashed (long-term), dotted (intermediate), and smooth (recent). The NAEP reading scale ranges from 0 to 500. The trend lines shown in this figure highlight the years of interest for this paper but are based on all available data points, including those from any intervening administrations. See table A1 for standard errors for the years of interest and https://www.nationsreportcard.gov/reading/nation/scores/?grade=4 for data and standard errors for all available years.

Source: National Center for Education Statistics (NCES), National Assessment of Educational Progress (NAEP), 2002–2019.

U.S. results for 4th-grade reading from the PIRLS international assessment detected fewer significant differences over time compared with NAEP. In PIRLS, there were no differences in the 2016 reading scores at the average, 10th, or 90th percentiles compared with 2001 (over the long term); nor were there differences in 10th percentile scores over any of the examined time periods (see figure 5). Over the intermediate term, both the average and 90th percentile scores increased (from 2006 to 2016). Over the most recent period (from 2011 to 2016), however, the 90th percentile score was flat, and the average score declined. PIRLS scores do not yet appear to be diverging, but if the current trajectories continue it could result in a worsening of not just average performance but also of equity.

Figure 5. Trends in U.S. 4th-graders' average scores and 10th and 90th percentile scores in PIRLS reading: Selected years 2001 to 2016



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

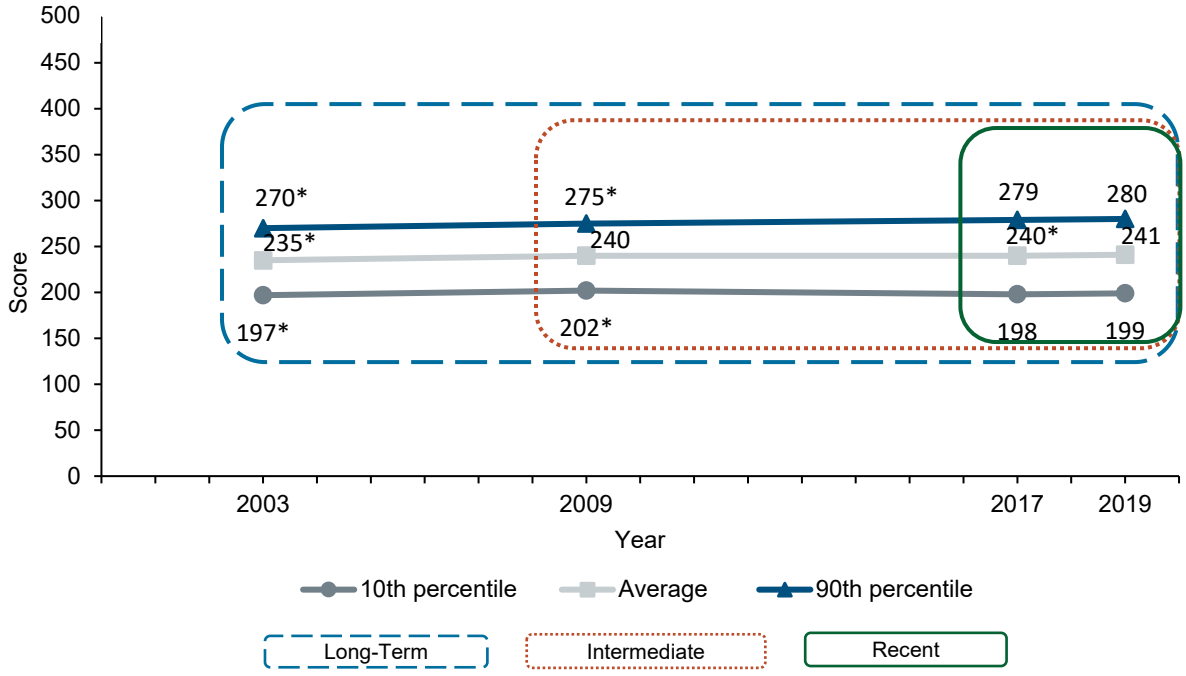
Note: The boxes represent different trend periods: dashed (long-term), dotted (intermediate), and smooth (recent). The PIRLS reading scale ranges from 0 to 1,000. The trend lines shown in this figure highlight the years of interest for this paper, which include all available data points. See table A2 for standard errors.

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

Mathematics

In 4th-grade mathematics, there were improvements in NAEP scores over the long term, with the 2019 average and 10th and 90th percentile scores all higher in 2019 than in 2003 (see figure 6). Over the intermediate time period (from 2009 to 2019), average scores did not change, but scores at the 90th percentile increased and scores at the 10th percentile decreased, indicating a two-tailed divergence of scores. However, in the most recent time period, from 2017 to 2019, the scores at the 10th and 90th percentiles were flat, while the average score increased slightly.

Figure 6. Trends in U.S. 4th-graders' average scores and 10th and 90th percentile scores in NAEP mathematics: Selected years 2003 to 2019



* $p < .05$. Score is significantly different from 2019 score.

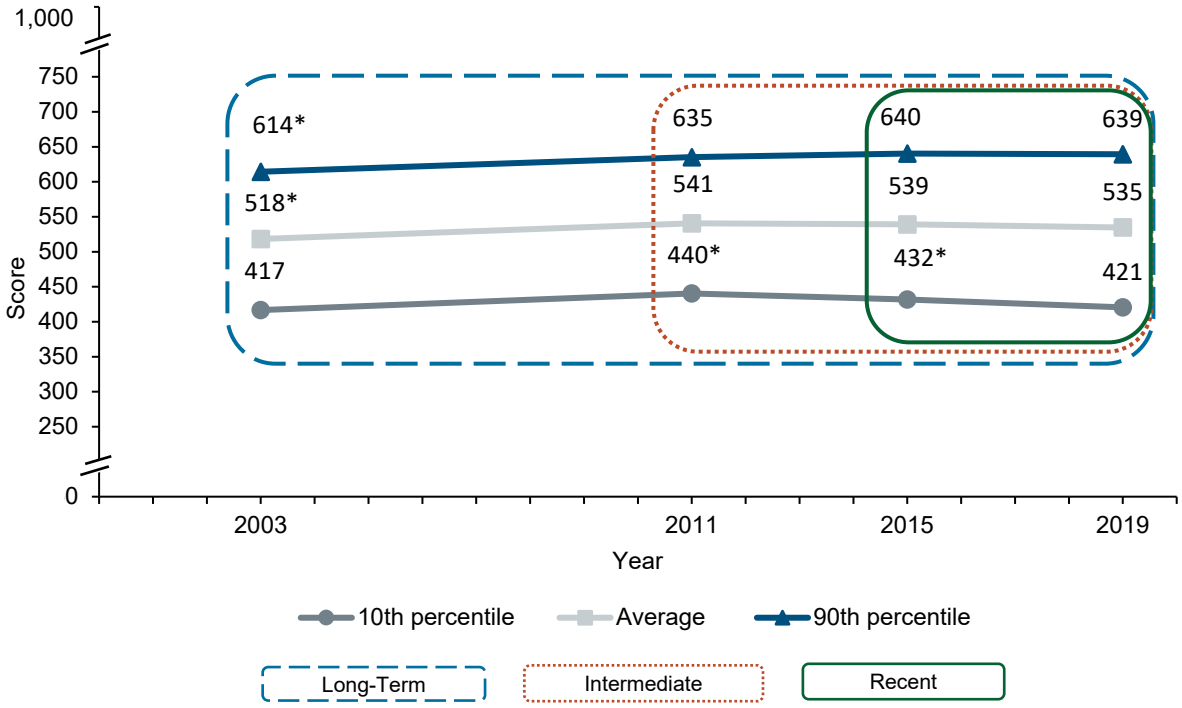
Note: The boxes represent different trend periods: dashed (long-term), dotted (intermediate), and smooth (recent). The NAEP mathematics scale ranges from 0 to 500. The trend lines shown in this figure highlight the years of interest for this paper but are based on all available data points, including those from any intervening administrations. See table A3 for standard errors for the years of interest and https://www.nationsreportcard.gov/mathematics/nation/scores/?grade=4 for data and standard errors for all available years.

Source: National Center for Education Statistics (NCES), National Assessment of Educational Progress (NAEP), 2003–2019.

The U.S. mathematics results from the TIMSS international assessment at 4th grade followed a somewhat similar pattern as the NAEP results. Over the long term, from 2003 to 2019, the average mathematics score increased, as did the 90th percentile score (see figure 7). Over the intermediate time period (from 2011 to 2019), the average and 90th percentile scores

flattened, but the 10th percentile score began to decline. This pattern continued to hold for the most recent time period (from 2015 to 2019), indicating a sustained bottom-led divergence.

Figure 7. Trends in U.S. 4th-graders' average scores and 10th and 90th percentile scores in TIMSS mathematics: Selected years 2003 to 2019



* $p < .05$. Score is significantly different from 2019 score.

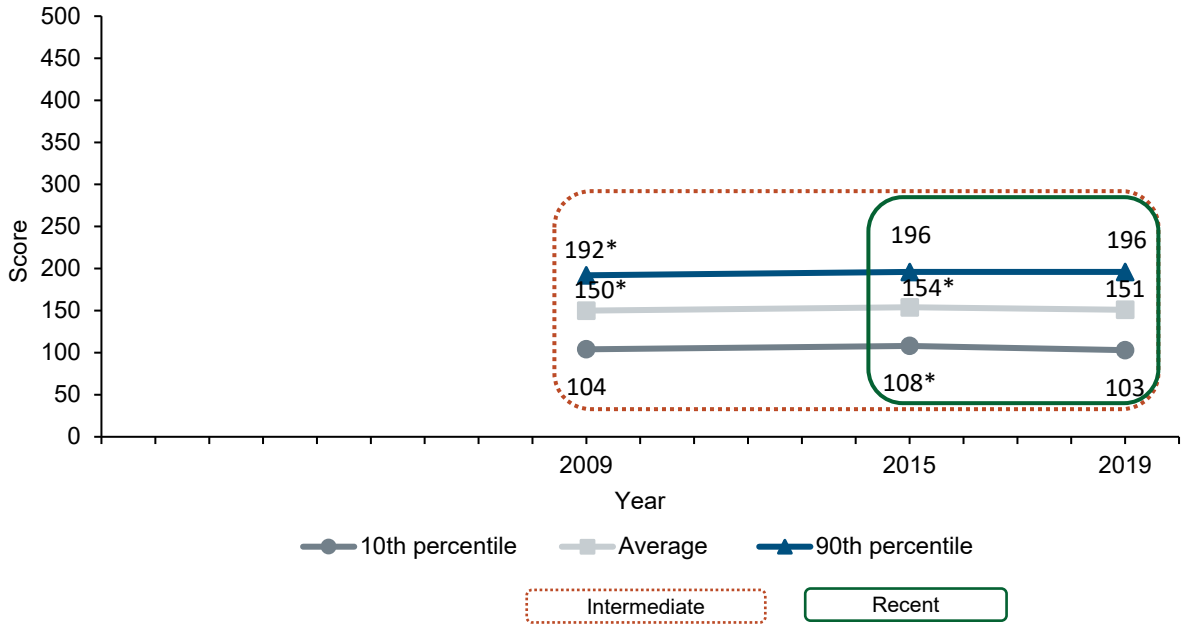
Note: The boxes represent different trend periods: dashed (long-term), dotted (intermediate), and smooth (recent). The TIMSS mathematics scale ranges from 0 to 1,000. The trend lines shown in this figure highlight the years of interest for this paper but are based on all available data points, including those from any intervening administrations. See table A4 for standard errors for the years of interest and https://nces.ed.gov/timss/results19/index.asp for data and standard errors for all years.

Source: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003–2019.

Science

The current NAEP 4th-grade science trend starts in 2009 when a new assessment framework was first implemented, and so reporting in this Research Note begins with the intermediate term. From 2009 to 2019, the average score and the 90th percentile score increased, while the 10th percentile scores in these years did not differ (see figure 8). During the recent time period (from 2015 to 2019), the average score decreased, drawn down by a decrease at the 10th percentile with no change at the 90th percentile, indicating a bottom-led divergence.

Figure 8. Trends in U.S. 4th-graders' average scores and 10th and 90th percentile scores in NAEP science: Selected years 2009 to 2019



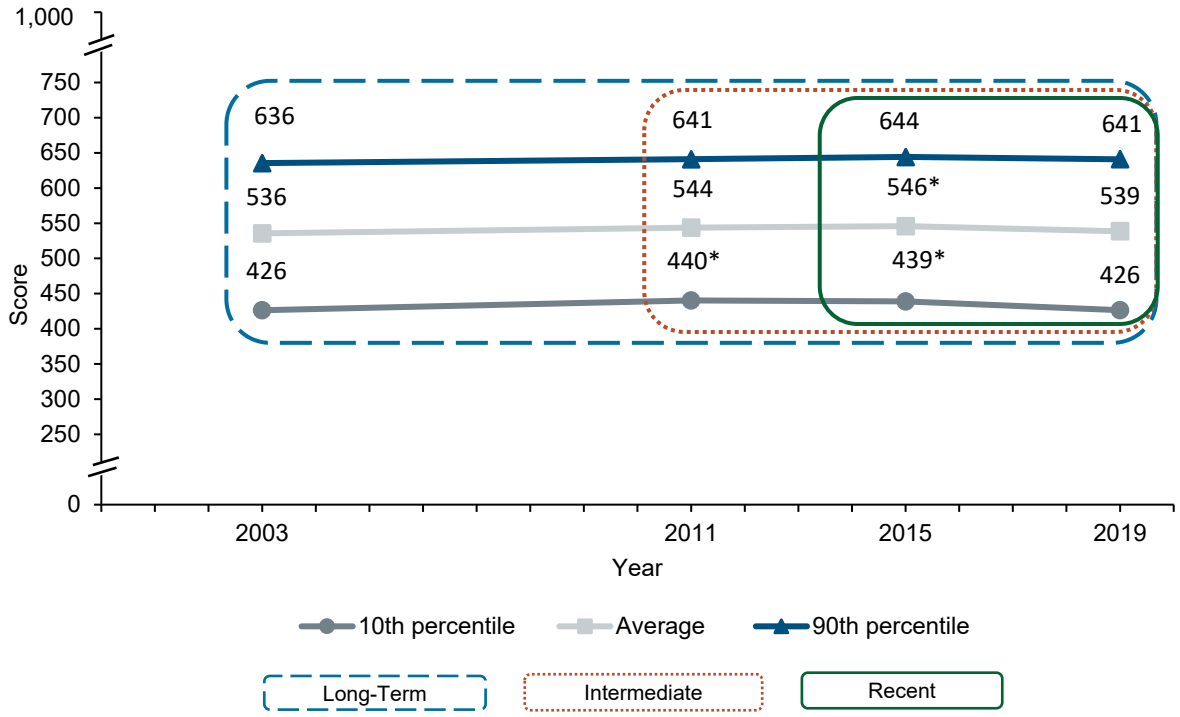
* $p < .05$. Score is significantly different from 2019 score.

Note: The boxes represent different trend periods: dotted (intermediate) and smooth (recent). The NAEP science scale ranges from 0 to 500. The NAEP science framework was revised in 2009, so there is no long-term trend period to report. The trend lines shown in this figure highlight the years of interest for this paper, which include all available data points. See table A5 for standard errors.

Source: National Center for Education Statistics (NCES), National Assessment of Educational Progress (NAEP), 2009–2019.

As noted above, the NAEP science framework was revised in 2009, and so the TIMSS international assessment is our only source for data on U.S. long-term trends in science. Unlike the U.S. 4th-grade mathematics trends, which showed improvements over the long term for at least two groups of students, the 2019 scores in TIMSS 4th-grade science were not different from those in 2003 for any group of students: 90th percentile, 10th percentile, or average (see figure 9). Turning to the intermediate time period (from 2011 to 2019), the 10th percentile score declined and the average and 90th percentile scores were unchanged. Looking at the most recent time period (from 2015 to 2019), the TIMSS 4th-grade science results showed a continued decline at the 10th percentile, along with a new decline in the average score, while the 90th percentile score remained flat.

Figure 9. Trends in U.S. 4th-graders' average scores and 10th and 90th percentile scores in TIMSS science: Selected years 2003 to 2019



* $p < .05$. Score is significantly different from 2019 score.

Note: The boxes represent different trend periods: dashed (long-term), dotted (intermediate), and smooth (recent). The TIMSS science scale ranges from 0 to 1,000. The trend lines shown in this figure highlight the years of interest for this paper but are based on all available data points, including those from any intervening administrations. See table A6 for standard errors for the years of interest and https://nces.ed.gov/timss/results19/index.asp for data and standard errors for all years.

Source: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003–2019.

SUMMARY OF 4TH-GRADE SCORE TRENDS

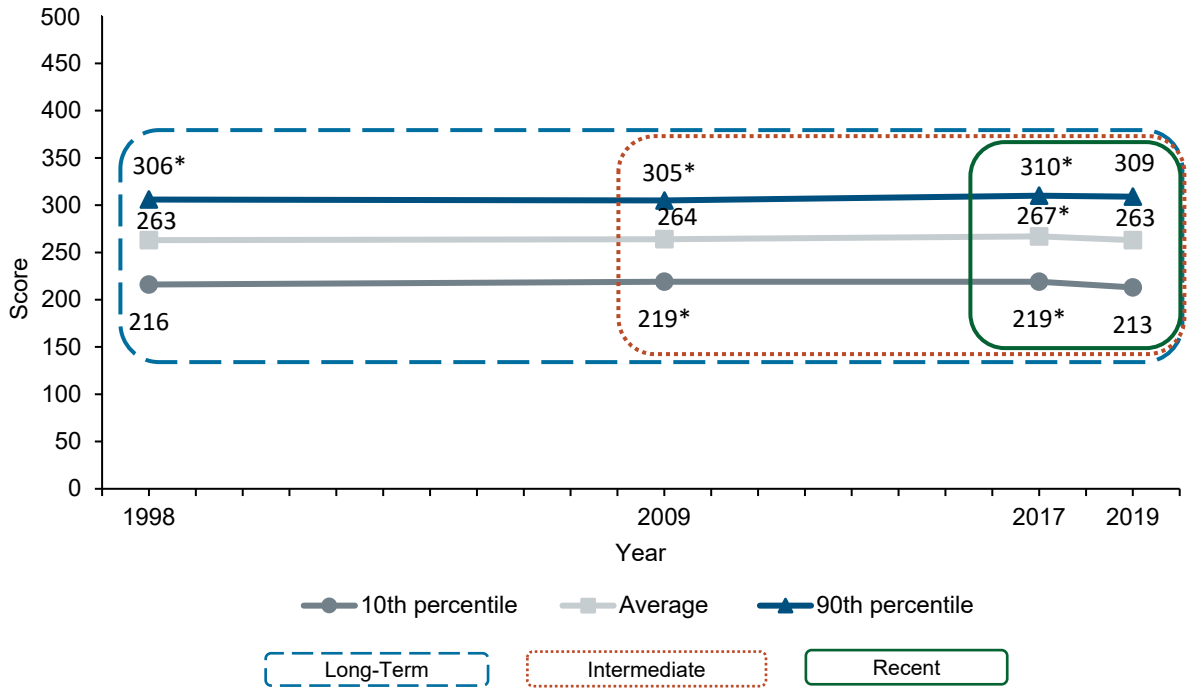
- U.S. 4th-grade students have shown improvement over the long term in both reading and mathematics at the middle and top ends of the distribution. Students performing at the bottom of the distribution in mathematics have also shown improvement over the long term in NAEP.
- However, in the intermediate and recent terms, the scores of low-performing 4th-graders have declined in all three subjects (reading, mathematics, and science).
- In all subjects and assessments, scores for high-performing students have stalled in the recent term.

Trends in 8th-grade and 15-year-olds' scores

Reading

Moving to the upper grades, we first look at trends for NAEP 8th-grade reading. Over the long term (from 1998 to 2019), the 90th percentile score rose, while the scores of students at the 10th percentile and on average were statistically unchanged (see figure 10). The average score was also unchanged over the intermediate term (from 2009 to 2019), when again the 90th percentile increased but the 10th percentile also decreased—indicating divergence. Over the recent term (from 2017 to 2019), there was a decrease in the average score, as well as in the 10th and 90th percentile scores, an across-the-board pattern of decline that was not seen in any other subjects or grades in the assessments included in this study.

Figure 10. Trends in U.S. 8th-graders' average scores and 10th and 90th percentile scores in NAEP reading: Selected years 1998 to 2019



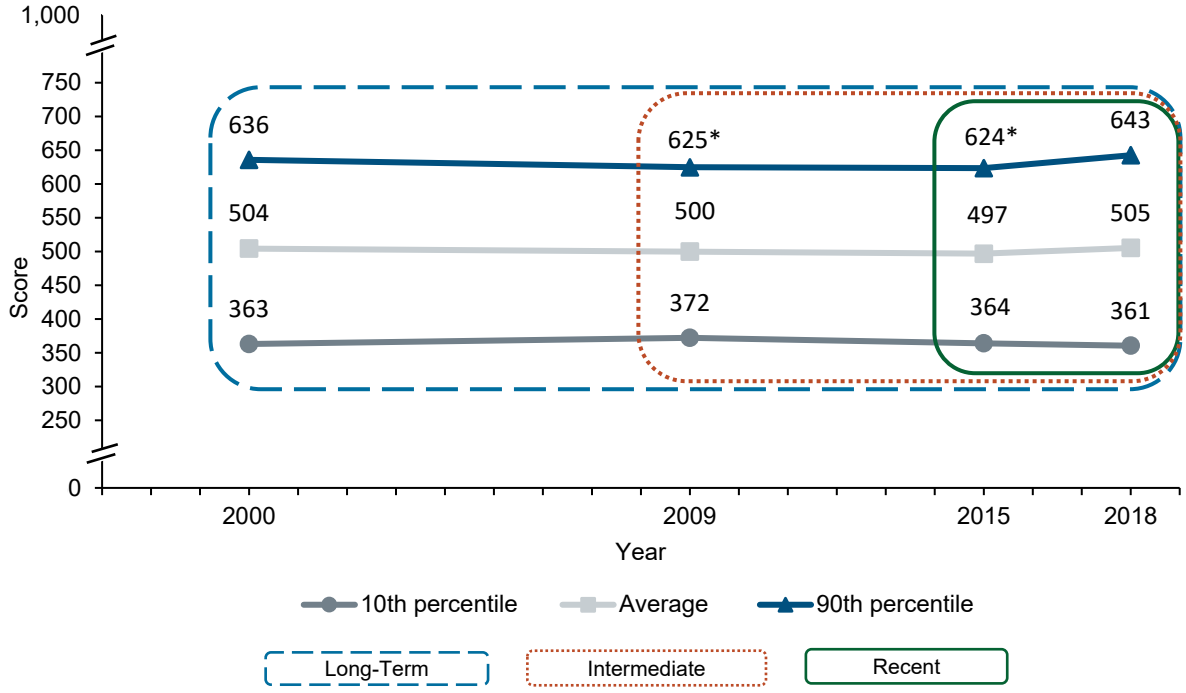
* $p < .05$. Score is significantly different from 2019 score.

Note: The boxes represent different trend periods: dashed (long-term), dotted (intermediate), and smooth (recent). The NAEP reading scale ranges from 0 to 500. The trend lines shown in this figure highlight the years of interest for this paper but are based on all available data points, including those from any intervening administrations. See table A7 for standard errors for the years of interest and https://www.nationsreportcard.gov/reading/nation/scores/?grade=8 for data and standard errors for all available years.

Source: National Center for Education Statistics (NCES), National Assessment of Educational Progress (NAEP), 1998–2019.

Next, we examine PISA results, which are for 15-year-olds (most commonly in 10th grade). The PISA reading scores were flat over the long term, with no significant differences at the average or at the 10th and 90th percentiles between 2000 and 2018 (see figure 11). While the average and 10th percentile scores remained statistically unchanged from 2009 and 2015 to 2018, the 90th percentile score rose in both periods. To some extent, the higher 90th percentile score was a recovery from decreases after 2009 (not shown), but it suggests additional growth as well. This indicates a top-led divergence in reading scores at the high school level, which was not seen in the NAEP 8th-grade data due to its small drop at the 90th percentile.

Figure 11. Trends in U.S. 15 year olds' average scores and 10th and 90th percentile scores in PISA reading: Selected years 2000 to 2018



* $p < .05$. Score is significantly different from 2018 score.

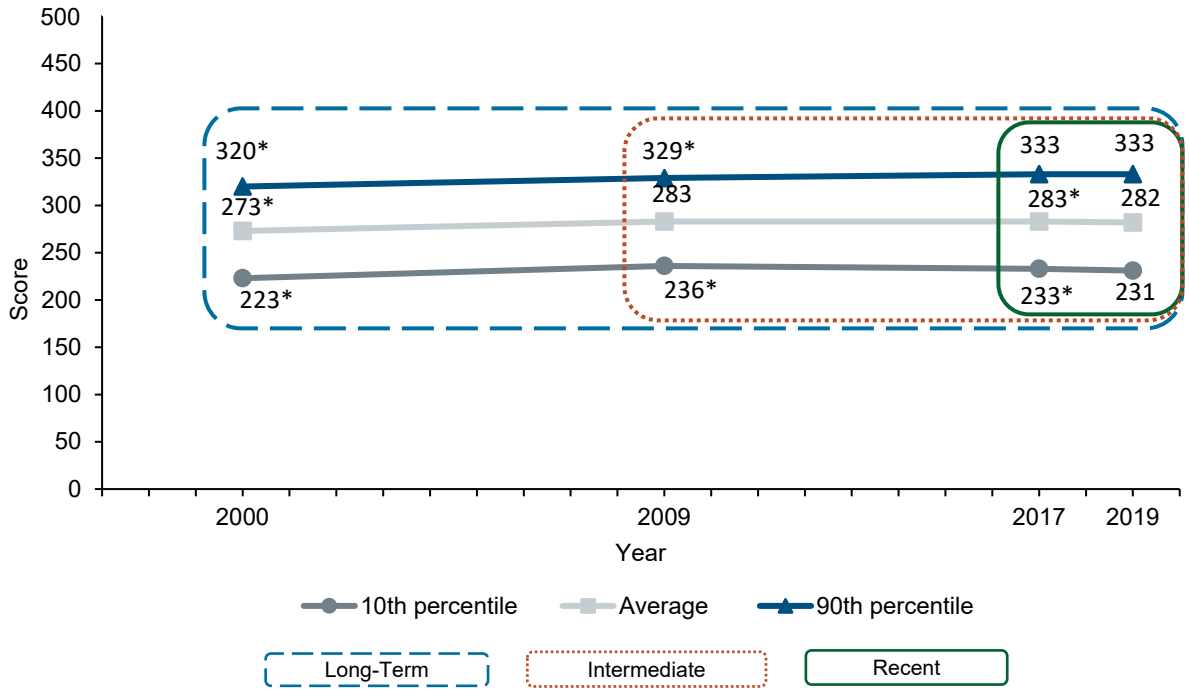
Note: The boxes represent different trend periods: dashed (long-term), dotted (intermediate), and smooth (recent). The PISA reading scale ranges from 0 to 1,000. The trend lines shown in this figure highlight the years of interest for this paper but are based on all available data points, including those from any intervening administrations. See table A8 for standard errors for the years of interest and https://nces.ed.gov/surveys/pisa/pisa2018/index.asp for data and standard errors for all years.

Source: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2000–2018.

Mathematics

Moving to NAEP 8th-grade mathematics, the average score increased over the long term, from 2000 to 2019 (see figure 12). There were also score increases at both the 10th and 90th percentiles over this roughly 20-year period. Over the intermediate term, from 2009 to 2019, there was no change in the average score; however, there was some movement at the 10th and 90th percentiles. As in other assessments, the 10th and 90th percentile scores moved in opposite directions over the intermediate term, with the 90th percentile score increasing and the 10th percentile score decreasing. Across the most recent assessment period, from 2017 to 2019, the average score and the 10th percentile score both decreased. However, even though the 90th percentile score was unchanged, the scores of low- and high-performing students continued to diverge because of the size of the drop for low-performing students.

Figure 12. Trends in U.S. 8th-graders' average scores and 10th and 90th percentile scores in NAEP mathematics: Selected years 2000 to 2019



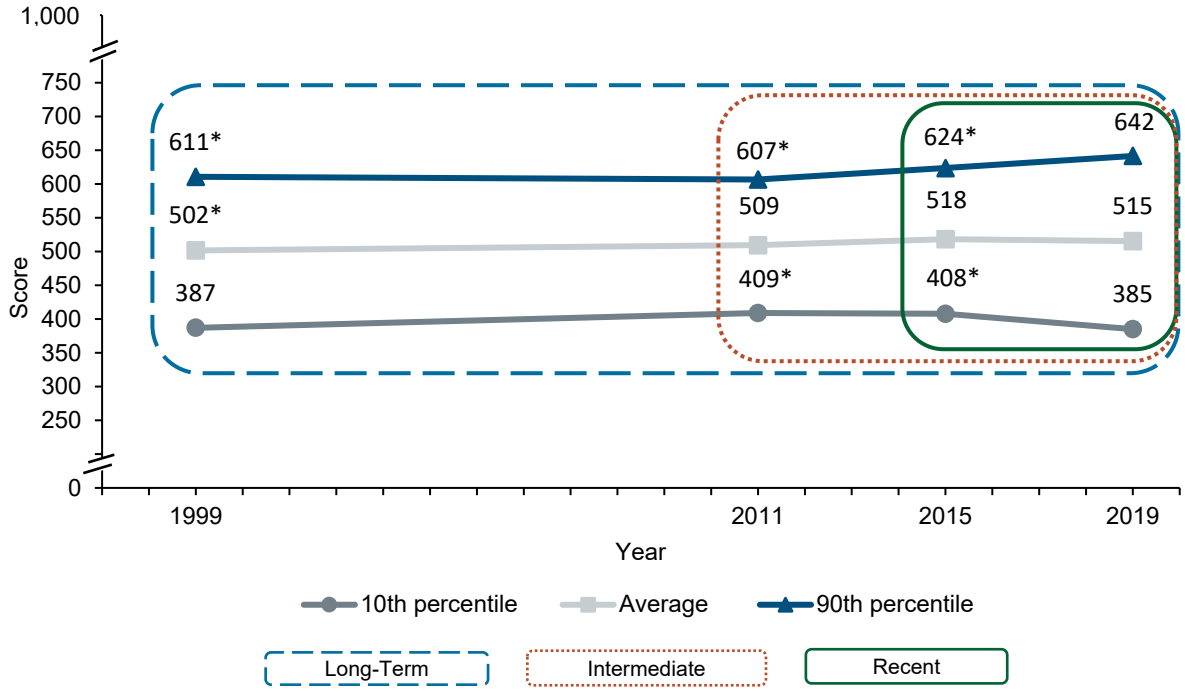
* $p < .05$. Score is significantly different from 2019 score.

Note: The boxes represent different trend periods: dashed (long-term), dotted (intermediate), and smooth (recent). The NAEP mathematics scale ranges from 0 to 500. The trend lines shown in this figure highlight the years of interest for this paper but are based on all available data points, including those from any intervening administrations. See table A9 for standard errors for the years of interest and https://www.nationsreportcard.gov/mathematics/nation/scores/?grade=8 for data and standard errors for all available years.

Source: National Center for Education Statistics (NCES), National Assessment of Educational Progress (NAEP), 2000–2019.

Similar to the NAEP trends in mathematics at 8th grade, the U.S. TIMSS results showed improvements in the average score and 90th percentile score over the long term (from 1999 to 2019), although the 10th percentile score was statistically unchanged (see figure 13). The intermediate TIMSS trends in mathematics at 8th grade (from 2011 to 2019) also mirrored the changes shown in NAEP. TIMSS scores decreased at the 10th percentile and increased at the 90th percentile, thus diverging, while the average score did not significantly change. The same pattern continued into the most recent time period (from 2015 to 2019), as TIMSS continued its two-tailed divergence of scores.

Figure 13. Trends in U.S. 8th-graders' average scores and 10th and 90th percentile scores in TIMSS mathematics: Selected years 1999 to 2019

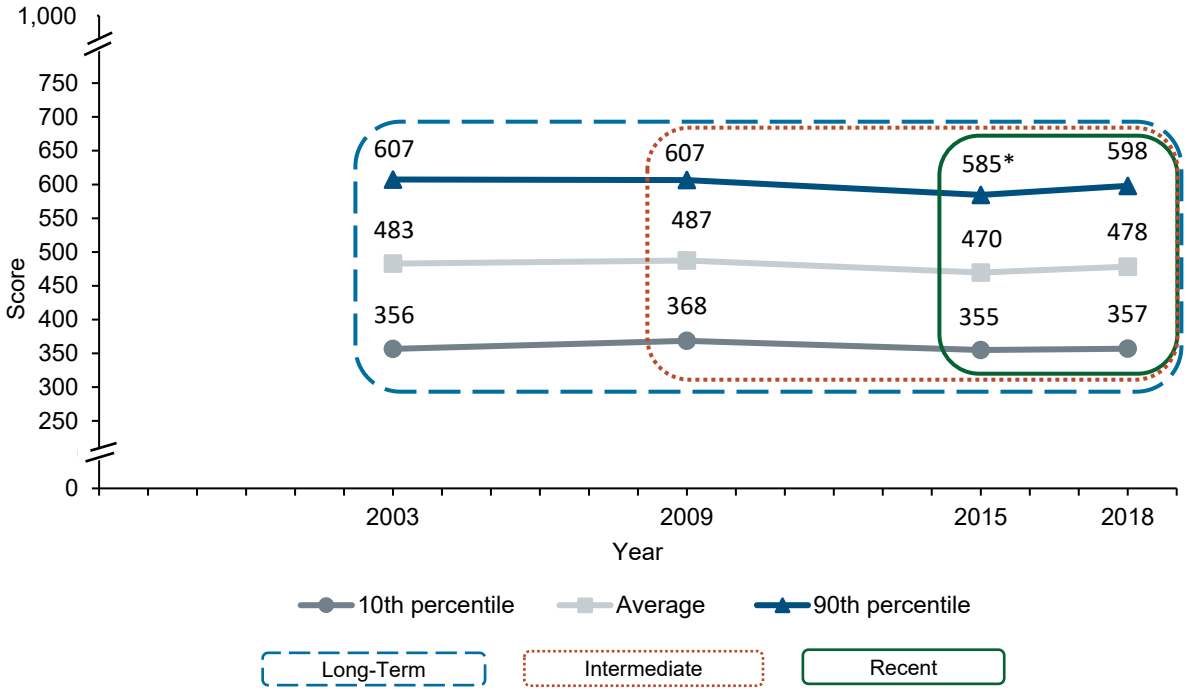


* $p < .05$. Score is significantly different from 2019 score.
Note: The boxes represent different trend periods: dashed (long-term), dotted (intermediate), and smooth (recent). The TIMSS mathematics scale ranges from 0 to 1,000. The trend lines shown in this figure highlight the years of interest for this paper but are based on all available data points, including those from any intervening administrations. See table A10 for standard errors for the years of interest and https://nces.ed.gov/timss/results19/index.asp for data and standard errors for all years.
Source: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1999–2019.

In PISA, the mathematics scores of U.S. 15-year-olds were flat over both the long and intermediate terms, with no significant differences in the 2018 average or 10th and 90th percentile scores from either the 2003 or 2009 scores (see figure 14). In the recent term, too,

the average and 10th percentile scores remained flat. However, 90th percentile scores increased from 2015 to 2018.

Figure 14. Trends in U.S. 15 year olds' average scores and 10th and 90th percentile scores in PISA mathematics: Selected years 2003 to 2018



* $p < .05$. Score is significantly different from 2018 score.

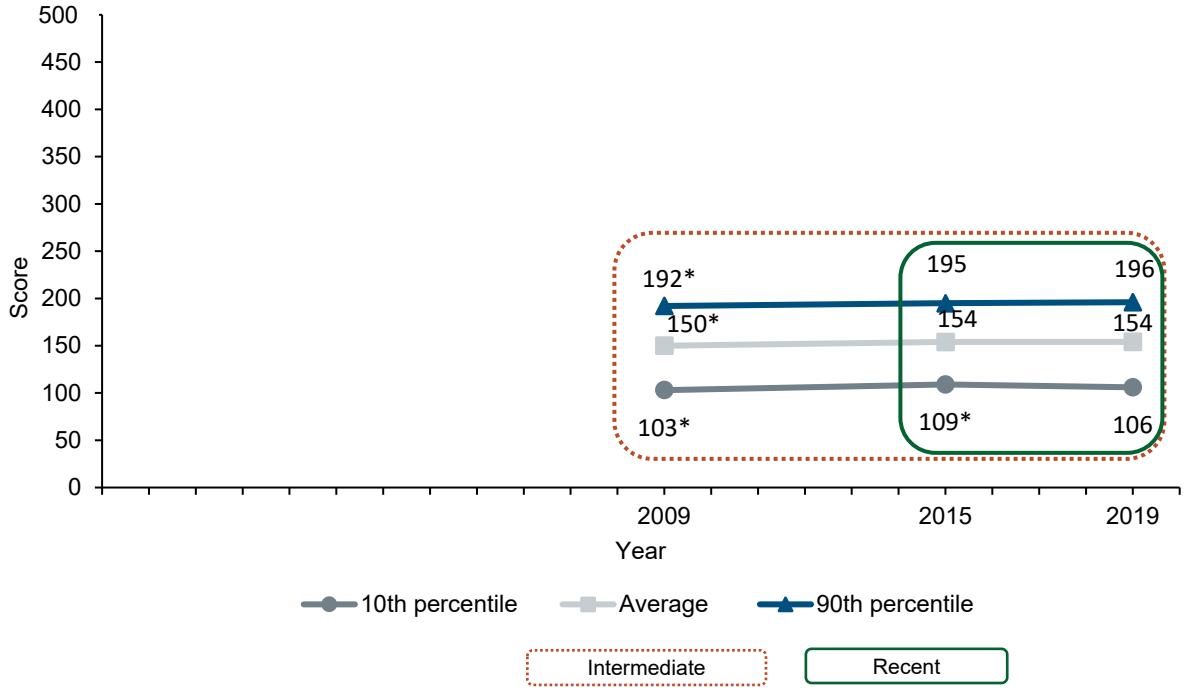
Note: The boxes represent different trend periods: dashed (long-term), dotted (intermediate), and smooth (recent). The PISA mathematics scale ranges from 0 to 1,000. The trend lines shown in this figure highlight the years of interest for this paper but are based on all available data points, including those from any intervening administrations. See table A11 for standard errors for the years of interest and https://nces.ed.gov/surveys/pisa/pisa2018/index.asp for data and standard errors for all years.

Source: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2003–2018.

Science

Like the NAEP 4th-grade science trends, the current NAEP 8th-grade science trends start in 2009 due to the framework change. Over the intermediate term, from 2009 to 2019, NAEP 8th-grade science scores rose on average and at both the 10th and 90th percentiles (see figure 15). This pattern, in which there were increases at all three data points, was mirrored only in the NAEP 4th- and 8th-grade mathematics results over the long term. Over the recent term (from 2015 to 2019), the average 8th-grade science score was flat, as was the score at the 90th percentile. However, the score at the 10th percentile decreased, indicating bottom-led divergence.

Figure 15. Trends in U.S. 8th-graders' average scores and 10th and 90th percentile scores in NAEP science: Selected years 2009 to 2019



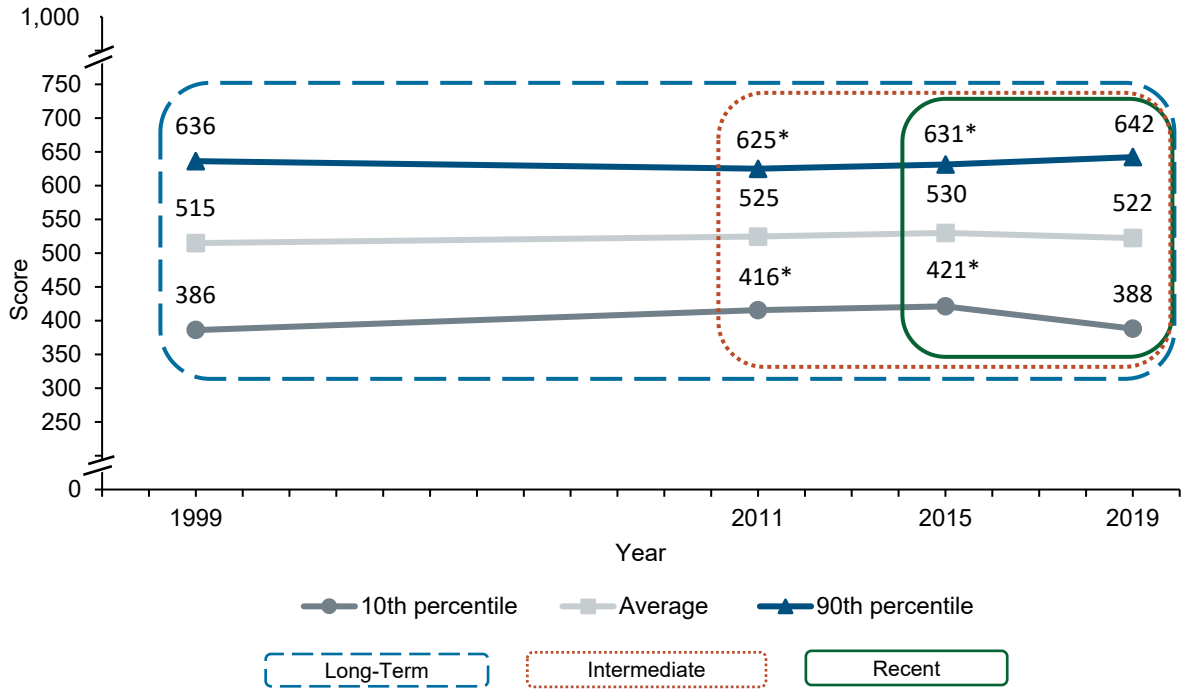
* $p < .05$. Score is significantly different from 2019 score.

Note: The boxes represent different trend periods: dotted (intermediate) and smooth (recent). The NAEP science scale ranges from 0 to 500. The NAEP science framework was revised in 2009, so there is no long-term trend period to report in this study. The trend lines shown in this figure highlight the years of interest for this paper, which include all available data points. See table A12 for standard errors.

Source: National Center for Education Statistics (NCES), National Assessment of Educational Progress (NAEP), 2009–2019.

Turning to the TIMSS 8th-grade science results, the long-term trends (from 1999 to 2019) showed no changes in the average score or the 10th and 90th percentile scores (see figure 16). Similarly, no changes were detected in the average science score over the intermediate or the recent time periods (2011–2019 and 2015–2019, respectively). The static average, however, masks the divergence that occurred over the intermediate and recent terms. Over both time periods, the 10th percentile score declined while the 90th percentile score rose, indicating a sustained two-tailed divergence in the scores.

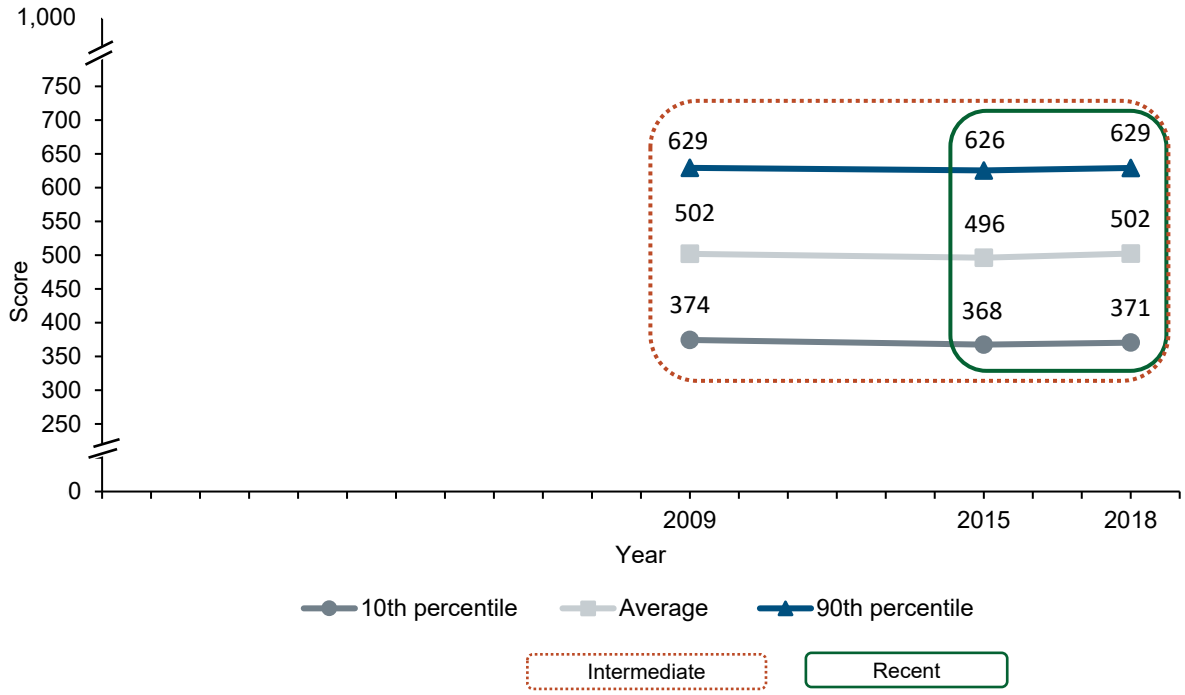
Figure 16. Trends in U.S. 8th-graders' average scores and 10th and 90th percentile scores in TIMSS science: Selected years 1999 to 2019



* $p < .05$. Score is significantly different from 2019 score.
Note: The boxes represent different trend periods: dashed (long-term), dotted (intermediate), and smooth (recent). The TIMSS science scale ranges from 0 to 1,000. The trend lines shown in this figure highlight the years of interest for this paper but are based on all available data points, including those from any intervening administrations. See table A13 for standard errors for the years of interest and https://nces.ed.gov/timss/results19/index.asp for data and standard errors for all years.
Source: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1999–2019.

In PISA science, for the two periods examined for this study (2009–2018 and 2015–2018), there were no changes in the average score or 10th and 90th percentile scores for U.S. 15-year-olds (see figure 17). However, if we were to include PISA data from 2006, which is the first year in which the full science framework was implemented, we would see that the 2018 science scores on average and at the 10th percentile have improved from that initial point (Sen et al., 2019). Since 2009, however, science scores have been flat across the distribution.

Figure 17. Trends in U.S. 15 year olds' average scores and 10th and 90th percentile scores in PISA science: Selected years 2009 to 2018



* $p < .05$. Score is significantly different from 2018 score.

Note: The boxes represent different trend periods: dotted (intermediate) and smooth (recent). The PISA science scale ranges from 0 to 1,000. The first year that science was administered as a "major" domain in PISA was 2006. However, since using that as a starting point for long-term trend would create a narrower time span than that for other assessments, we forgo a long-term data point in this study. The trend lines shown in this figure highlight the years of interest for this paper but are based on all available data points, including those from any intervening administrations. See table A14 for standard errors for the years of interest and https://nces.ed.gov/surveys/pisa/pisa2018/index.asp for data and standard errors for all years.

Source: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009–2018.

SUMMARY OF 8TH-GRADE AND 15-YEAR-OLDS' SCORE TRENDS

- Over the long term, U.S. students' scores were generally unchanged (reading and science) or improved (mathematics) among 8th-graders and 15-year-olds.
- However, over the intermediate term across all three subjects, the scores of low-performing students tended to decline, while the scores of high-performing students rose. In 8th-grade TIMSS, this exact pattern persisted over the recent term, whereas in NAEP, the scores of high-performing students flattened or reversed while declines for low-performing students continued.
- In any case, with two exceptions out of the eight subject/grade/assessment combinations, U.S. student performance was shown to be diverging: top-led in PISA, bottom-led in NAEP, and two-tailed in TIMSS. The two exceptions are NAEP reading, which showed recent declines for all groups, and PISA science, which showed no significant score changes for any groups of students.

Trends summarized across grades, subjects, and programs

Yearly state assessment scores are important to consider but ultimately lead to comparisons—how did this year's results compare to last year's? And if scores change from year to year, what does it indicate—is it good? Should the alarm bells be ringing? In addition, for a number of reasons, states change their state assessments frequently. So, one of the benefits of NAEP and ILSA results is that they provide longer-term trends than most state assessments do, and they give us the ability to place scores in a more wholistic context (and, for ILSAs, the ability to place U.S. scores in an international context). It is also important to note that while each assessment framework differs in specifics, they also broadly overlap, and new insight can be gained when considering all the results of NAEP and ILSA results together over time.

In this section, a discussion of the trends across these multiple assessments brings together the NAEP and ILSA data to develop a comprehensive understanding of pre-pandemic student achievement in the United States. It is followed by a look at the prevalence of the U.S. patterns among other countries.

U.S. Results

In the previous sections, score patterns over the long, intermediate, and recent terms were discussed by grade, subject, and program. This section considers the score patterns altogether, examining them across the grades, subjects, and programs, as shown in figure 18. This figure, which is a matrix, summarizes the earlier data symbolically.

HOW TO READ FIGURE 18

In the figure, the row headings are the assessments, and the column headings are the scores of interest (10th percentile, average, and 90th percentile), grouped by the three time periods of interest (long, intermediate, and recent term). Each cell then shows whether there has been a significant score change for the given assessment and time period. Specifically: If the most recent score is significantly higher than the score in the reference year of the given period, the cell is green with an up arrow. If the most recent score is lower, the cell is red with a down arrow. If there is no difference between the two years' scores, the cell is gray with side-to-side arrows. (The years corresponding to the long-term, intermediate, and recent trend periods for each assessment, grade, and subject are shown in figure 2.)

Over the long term, three patterns emerge. First, there were gains across the distribution—at the 10th percentile, average, and 90th percentile—in NAEP 4th- and 8th-grade mathematics over the long term. Second, there were no score declines for low-performing students in any of the assessments: their most recent scores were not statistically different from their long-term starting points. Third, there were no long-term score changes for any group of students in PIRLS and PISA reading, PISA mathematics, or TIMSS science (either grade): these scores were also not different from the long-term starting points.

Turning to the intermediate term, see a prevalent pattern of divergence between the 10th and 90th percentiles, which occurs on all assessments except NAEP 8th-grade science and PISA mathematics and science. In all other assessments, the scores of high- and low-performing students diverged—most commonly because high-performing students' scores rose while low-performing students' scores dropped.

Finally, looking at the most recent term, we see a number of patterns. First, there are more red down-arrow cells than green up-arrow cells, indicating more score declines than increases. Second, scores for low-performing students did not increase in any assessment. And finally, we see a story similar to what was observed across the intermediate-term—a prevalence of diverging scores between low- and high-performing students, with a few exceptions for PIRLS, NAEP 4th-grade mathematics, NAEP 8th-grade reading, and PISA science.

Figure 18. Overview of U.S. score trends across assessments and subjects, by grade

Assessments and subjects	Long-term trend ~20-year span			Intermediate trend ~10-year span			Recent trend ~ 2- to 5-year span		
	10th percentile	Average	90th percentile	10th percentile	Average	90th percentile	10th percentile	Average	90th percentile
Grade 4									
NAEP reading	↔	↑	↑	↓	↔	↑	↓	↓	↔
PIRLS reading	↔	↔	↔	↔	↑	↑	↔	↓	↔
NAEP mathematics	↑	↑	↑	↓	↔	↑	↔	↑	↔
TIMSS mathematics	↔	↑	↑	↓	↔	↔	↓	↔	↔
NAEP science[1]	–	–	–	↔	↑	↑	↓	↓	↔
TIMSS science	↔	↔	↔	↓	↔	↔	↓	↓	↔
Grade 8 and 15-year-olds[2]									
NAEP reading	↔	↔	↑	↓	↔	↑	↓	↓	↓
PISA reading	↔	↔	↔	↔	↔	↑	↔	↔	↑
NAEP mathematics	↑	↑	↑	↓	↔	↑	↓	↓	↔
TIMSS mathematics	↔	↑	↑	↓	↔	↑	↓	↔	↑
PISA mathematics	↔	↔	↔	↔	↔	↔	↔	↔	↑
NAEP science[1]	–	–	–	↑	↑	↑	↓	↔	↔
TIMSS science	↔	↔	↔	↓	↔	↑	↓	↔	↑
PISA science[3]	–	–	–	↔	↔	↔	↔	↔	↔

↑ Upward trend (Most recent score is higher than score in year of referenced term, $p < .05$)

↔ No change (Most recent score is not significantly different from score in year of referenced term, $p < .05$)

↓ Downward trend (Most recent score is lower than score in year of referenced term, $p < .05$)

[1] The NAEP science framework was revised in 2009, so there is no long-term trend period to report in this study.
[2] NAEP and TIMSS results are for 8th-graders. PISA results are for 15-year-olds.
[3] The first year in which science was administered as a "major" domain in PISA was 2006. However, since using that as a starting point for long-term trend would create a time span that is less than that for the other assessments, we forgo a long-term data point in this study.
Note: See figure 2 for years corresponding to long-term, intermediate, and recent trend periods for each assessment, grade, and subject. See tables A1-A14 for data and standard errors on which the figure is based.

Cross-Country Comparisons

Broadening to a global perspective, the ILSA studies—PIRLS, TIMSS, and PISA—provide cross-country comparisons in student achievement and can provide answers to the following questions. Are the patterns of diverging scores identified in the United States common or unique? What other kinds of score trend patterns are seen in other countries? To synthesize the cross-country comparisons and to place the U.S. results in a global context, our research focused on the eight countries, plus the United States, that participated in at least one subject of all three of the study's ILSAs and selected comparison years. These countries (Hong Kong-CHN, Hungary, Italy, Lithuania, New Zealand, the Russian Federation, Singapore, and the United Kingdom/England-GBR) represent a range of performance levels, geographic regions, and language groups.

4th-grade score patterns. Beginning with 4th-grade reading trends from PIRLS, most of these eight countries, unlike the United States, made improvements for at least one group of students over the long term (see supplemental figure A1). Several countries (Hungary, Lithuania, the Russian Federation, and England-GBR) saw improvements in either the 10th percentile score, the average score, or the 90th percentile score in each of the time periods (long term, intermediate, and recent). The Russian Federation showed improvements across all three time periods for all three groups of students. Only New Zealand, like the United States, had a declining average score in the recent time period. New Zealand also showed declining 10th percentile and average scores in the intermediate time period, while Italy's 90th percentile score dropped in the intermediate, but not in the recent, time period.

Turning to 4th-grade mathematics trends from TIMSS, most of the eight countries, plus the United States, showed improvements for at least one group of students over the long term. Three countries (Italy, Lithuania, and England-GBR) saw improvements for at least one group in every time period, and two countries (the Russian Federation and Singapore) had gains for every group over both the long and intermediate terms but none in the recent term. The United States stands out among the group as the only country showing a decline at the 10th percentile in the intermediate term; the U.S. decline continued over the recent time period, when there was also a decline at the 10th percentile (and average) for Hong Kong-CHN. Conversely, Hungary showed a decline at the 90th percentile in the recent term.

The 4th-grade science trends from TIMSS across countries were more mixed than were those in mathematics. Only three countries showed improvements with no declines over the long term (Lithuania, the Russian Federation, and Singapore), while four countries showed declines for at least one group of students (Hong Kong-CHN, Italy, New Zealand, and England-GBR). As in mathematics, the United States began to decline at the 10th percentile in the intermediate time period, which continued over into the recent term. The pattern in the United States is not

represented in any of the other eight countries, but Hong Kong's 10th percentile, average, and 90th percentile scores all declined over the recent time period, and Hungary's average and 90th percentile scores declined as well. Most other countries' scores remained flat for all three groups of students over the recent time period, with only Lithuania showing an improvement in its average and 90th percentile scores.

8th-grade and 15-year-old score patterns. In PISA reading in the United States, 90th percentile scores increased in the intermediate and recent terms (see supplemental figure A2). Rises at the top end of the distribution were not unique across the countries in the study, with score increases for high-performing students in Hong Kong-CHN and Singapore (recently) and Hong Kong-CHN, Lithuania, the Russian Federation, Singapore, and the United Kingdom (over the intermediate term). However, in Hong Kong-CHN, gains for high-performing students were accompanied by declines for low-performing students in both of these periods. In the Russian Federation, the increases seen for high-performing students (and on average) over the intermediate term, flattened (or reversed) over the recent term along with a decrease at the bottom end of the distribution.

In upper-grade mathematics, data are available from both TIMSS and PISA. In TIMSS, the U.S. pattern of two-tailed divergence—rising scores at the top and falling scores at the bottom of the distribution—was not mirrored in any other countries. Of note, Lithuania improved in all periods at one or more points on the distribution; and of concern, Hong Kong-CHN exhibited recent score drops at the 10th percentile that impacted its average score. In PISA, the scores of low-performing U.S. students held steady over time, but recent score increases for high-performing students indicate increasing divergence. Elsewhere in PISA, this pattern was only seen in the United Kingdom, with the scores of all other countries' high-performing students remaining flat over the recent time period.

In upper-grade science, data are again available from both TIMSS and PISA. One striking observation is that, compared to their performance in other subjects, more countries experienced declines in science over one or more time periods—although for the United States, this was only in TIMSS and only at the 10th percentile. The U.S. performance patterns in TIMSS (two-tailed divergence in the intermediate and recent terms) and PISA (flat over these periods) were not precisely mirrored in any other countries. However, three countries (Hong Kong-CHN, New Zealand, and England-GBR) had declining TIMSS scores for low-performing students (among others) in both periods, and most countries' PISA science performance was flat across the distribution at least in the recent term. Lithuania was again a standout in TIMSS, with consistent score increases across time periods and across the distribution.

LIMITATIONS OF ANALYSES AND CONSIDERATIONS FOR FUTURE WORK

This study provides a useful, cross-assessment look at patterns in U.S. student performance and identifies a troubling, generalized trend toward diverging scores and, thus, rising inequity. Several limitations are important to keep in mind in interpreting the data.

- First, as mentioned earlier, the various assessment programs—while broadly overlapping—are built on their own unique frameworks and are not equivalent measures. It says something that, despite potential content differences, similar patterns are emerging across programs. However, it should be kept in mind that content differences may impact differences in the results observed and may be worthy of exploration themselves. For example, PISA and 8th-grade TIMSS mathematics both show increases for high-performing students, but only TIMSS shows decreases for low-performing students. Can this be solely attributed to TIMSS' greater precision in measuring the low end of the scale or is there, additionally, particular content with which U.S. students struggle on TIMSS? As another, international example, what does it mean for Lithuania that students routinely improve in TIMSS but their performance is generally static on PISA? To what extent might PISA's design, which has rotating major and minor domains, impact sensitivity to small changes over time?
- Second, while the time spans were set to be roughly similar for the assessments in direct comparison, neither the years of assessment nor the time spans between them are perfectly equivalent. In general, these differences are minor and unlikely to be an issue given the slow pace of educational change; however, it is more notable with PIRLS, for which the most recent data point available (2016) is 3 years prior to that for NAEP 4th-grade reading (2019).
- Finally, there are differences in sensitivity between NAEP and ILSAs. Because NAEP is designed to detect national- and state-level subgroup differences over time, its sample sizes are large and measurement precision high. The ILSAs, rather, are intended to monitor national-level achievement and thus may not pick up small changes. The results seem to bear this out, as typically, the NAEP results in this study were statistically significant more frequently.

Future analyses might unpack some of the questions above, as well as dig deeper into the question of who are the low- and high-performing students. Are there changes in the composition of these groups—in terms of student characteristics—over time? Would the results be different if the average of the top and bottom deciles of students were tracked instead of the cut scores of the 90th and 10th percentiles? Additionally, it would be useful to add to these analyses when the next rounds of ILSA and NAEP data are released, as PIRLS 2021, NAEP 2022, PISA 2022, and TIMSS 2023 each will provide the first post-pandemic data points.

Conclusions

When the same pattern is seen repeatedly over time and particularly from multiple, independent sources, the convergence of information strengthens the argument that the pattern is valid. NAEP and ILSA results offer a unique opportunity to triangulate an emerging pattern—a divergence between the tails of the score distribution in the United States. This widening pattern became prevalent over the last decade (the intermediate trend), occurring in almost every subject and grade combination, and it continues to be seen in almost all of the recent trend comparisons prior to the COVID-19 pandemic. The international data indicate that this pattern is fairly unique to the United States. As post-pandemic results become available, it will also be critical to update this research and determine whether this pattern of divergence has been exacerbated, held steady, or improved. This study provides an important baseline for tracking the achievement of high- and low-performing students so that any equity concerns can be understood and addressed.

Authors and Acknowledgments

AIR contributing authors include Maria Stephens (senior researcher) and Frank Fonseca (researcher). The lead author Samantha Burg (psychometrician) and Lydia Malley (education research analyst) are with the National Center for Education Statistics (NCES).

The authors would like to thank Brian Cramer and Sheila Thompson at NCES for their input in the initial development of this work, and the reviewers for their insightful suggestions, particularly Stephen Provasnik (NCES) and Markus Broer (AIR). In addition, the authors would like to specifically thank Thanh Mai (AIR) for her technical assistance.

References

Gattis, K., Kim, Y. Y., Stephens, M., Hall, L. D., Liu, F., & Holmes, J. (2016). *A comparison study of the Program for International Student Assessment (PISA) 2012 and the National Assessment of Educational Progress (NAEP) 2013 mathematics assessments* (AIR-NAEP Working Paper #02-2016). American Institutes for Research. https://www.air.org/sites/default/files/2021-06/Comparison-NAEP-PISA-Mathematics-May-2016.pdf

Mullis, I. V. S., & Martin, M. O. (Eds.). (2015). *PIRLS 2016 assessment framework* (2nd ed.). Boston College, TIMSS & PIRLS International Study Center. http://timssandpirls.bc.edu/pirls2016/framework.html

Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 assessment frameworks*. Boston College, TIMSS & PIRLS International Study Center. http://timssandpirls.bc.edu/timss2019/frameworks/

National Assessment Governing Board. (2019a). *Mathematics framework for the 2019 National Assessment of Educational Progress*. U.S. Department of Education. https://www.nagb.gov/naep-frameworks/mathematics.html

National Assessment Governing Board. (2019b). *Reading framework for the 2019 National Assessment of Educational Progress*. U.S. Department of Education. https://www.nagb.gov/naep-frameworks/mathematics.html

National Assessment Governing Board. (2019c). *Science framework for the 2019 National Assessment of Educational Progress*. U.S. Department of Education. https://www.nagb.gov/naep-frameworks/mathematics.html

National Center for Education Statistics. (2008). *Comparing TIMSS with NAEP and PISA in mathematics and science*. U.S. Department of Education. https://nces.ed.gov/surveys/international/cross-study-comparisons.asp

National Center for Education Statistics. (2010). Comparison of the PISA 2009 and NAEP 2009 reading assessments. U.S. Department of Education. https://nces.ed.gov/surveys/international/cross-study-comparisons.asp

Organization for Economic Cooperation and Development. (2019). *PISA 2018 assessment and analytical framework.* OECD Publishing. https://doi.org/10.1787/b25efab8-en

Sen, A., Mai, T., Kastberg, D., Cummings, L., and Perkins. R. (2019). *Highlights of U.S. PISA 2018 results web report* (NCES 2020-166). National Center for Education Statistics, U.S. Department of Education. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2020166

Data Tables and Supplementary Figures

Table A1. Trends in U.S. 4th-graders' average scores and 10th and 90th percentile scores in NAEP reading: Selected years 2002 to 2019

Year	10th percentile		s.e.	Average score		s.e.	90th percentile		s.e.
2019	168		(0.4)	220		(0.2)	266		(0.4)
2017	171	*	(0.4)	222	*	(0.3)	267		(0.3)
2009	175	*	(0.5)	221		(0.3)	264	*	(0.3)
2002	170		(0.9)	219	*	(0.4)	263	*	(0.4)

* $p < .05$. Score is significantly different from 2019 score.

Note: The NAEP reading scale ranges from 0 to 500. Data in this table are for figure 4.

Source: National Center for Education Statistics (NCES), National Assessment of Educational Progress (NAEP), 2002, 2009, 2017, and 2019.

Table A2. Trends in U.S. 4th-graders' average scores and 10th and 90th percentile scores in PIRLS reading: Selected years 2001 to 2016

Year	10th percentile		s.e.	Average score		s.e.	90th percentile		s.e.
2016	446		(6.3)	549		(3.1)	645		(4.0)
2011	458		(3.1)	556	*	(1.6)	648		(2.2)
2006	441		(5.0)	540	*	(3.4)	631	*	(3.7)
2001	431		(8.8)	542		(3.8)	640		(3.1)

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

Note: The PIRLS reading scale ranges from 0 to 1,000. Data in this table are for figure 5.

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

Table A3. Trends in U.S. 4th-graders' average scores and 10th and 90th percentile scores in NAEP mathematics: Selected years 2003 to 2019

Year	10th percentile		s.e.	Average score		s.e.	90th percentile		s.e.
2019	199		(0.3)	241		(0.2)	280		(0.4)
2017	198		(0.3)	240	*	(0.2)	279		(0.4)
2009	202	*	(0.4)	240		(0.2)	275	*	(0.2)
2003	197	*	(0.3)	235	*	(0.2)	270	*	(0.3)

* $p < .05$. Score is significantly different from 2019 score.

Note: The NAEP mathematics scale ranges from 0 to 500. Data in this table are for figure 6.

Source: National Center for Education Statistics (NCES), National Assessment of Educational Progress (NAEP), 2003, 2009, 2017, and 2019.

Table A4. Trends in U.S. 4th-graders' average scores and 10th and 90th percentile scores in TIMSS mathematics: Selected years 2003 to 2019

Year	10th percentile		s.e.	Average score		s.e.	90th percentile		s.e.
2019	421		(4.4)	535		(2.5)	639		(2.8)
2015	432	*	(3.1)	539		(2.3)	640		(2.6)
2011	440	*	(3.6)	541		(1.9)	635		(2.7)
2003	417		(3.3)	518	*	(2.4)	614	*	(2.8)

* $p < .05$. Score is significantly different from 2019 score.

Note: The TIMSS mathematics scale ranges from 0 to 1,000. Data in this table are for figure 7.

Source: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003, 2011, 2015, and 2019.

Table A5. Trends in U.S. 4th-graders' average scores and 10th and 90th percentile scores in NAEP science: Selected years 2009 to 2019

Year	10th percentile		s.e.	Average score		s.e.	90th percentile		s.e.
2019	103		(0.9)	151		(0.6)	196		(0.6)
2015	108	*	(0.6)	154	*	(0.3)	196		(0.4)
2009	104		(0.6)	150	*	(0.3)	192	*	(0.3)

* $p < .05$. Score is significantly different from 2019 score.

Note: The NAEP science scale ranges from 0 to 500. The NAEP science framework was revised in 2009, so there is no long-term trend period to report. Data in this table are for figure 8.

Source: National Center for Education Statistics (NCES), National Assessment of Educational Progress (NAEP), 2009, 2015, and 2019.

Table A6. Trends in U.S. 4th-graders' average scores and 10th and 90th percentile scores in TIMSS science: Selected years 2003 to 2019

Year	10th percentile		s.e.	Average score		s.e.	90th percentile		s.e.
2019	426		(4.8)	539		(2.7)	641		(2.5)
2015	439	*	(3.3)	546	*	(2.2)	644		(2.7)
2011	440	*	(3.3)	544		(2.1)	641		(2.3)
2003	426		(3.5)	536		(2.5)	636		(3.0)

* $p < .05$. Score is significantly different from 2019 score.

Note: The TIMSS science scale ranges from 0 to 1,000. Data in this table are for figure 9.

Source: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003, 2011, 2015, and 2019.

Table A7. Trends in U.S. 8th-graders' average scores and 10th and 90th percentile scores in NAEP reading: Selected years 1998 to 2019

Year	10th percentile		s.e.	Average score		s.e.	90th percentile		s.e.
2019	213		(0.5)	263		(0.3)	309		(0.3)
2017	219	*	(0.5)	267	*	(0.3)	310	*	(0.3)
2009	219	*	(0.5)	264		(0.3)	305	*	(0.4)
1998	216		(1.7)	263		(0.8)	306	*	(0.8)

* $p < .05$. Score is significantly different from 2019 score.

Note: The NAEP reading scale ranges from 0 to 500. Data in this table are for figure 10.

Source: National Center for Education Statistics (NCES), National Assessment of Educational Progress (NAEP), 1998, 2009, 2017, and 2019.

Table A8. Trends in U.S. 15-year-olds' average scores and 10th and 90th percentile scores in PISA reading: Selected years 2000 to 2018

Year	10th percentile		s.e.	Average score		s.e.	90th percentile		s.e.
2018	361		(5.3)	505		(3.6)	643		(3.9)
2015	364		(5.4)	497		(3.4)	624	*	(3.8)
2009	372		(3.9)	500		(3.7)	625	*	(5.0)
2000	363		(11.4)	504		(7.0)	636		(6.5)

* $p < .05$. Score is significantly different from 2018 score.

Note: The PISA reading scale ranges from 0 to 1,000. Data in this table are for figure 11.

Source: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2000, 2009, 2015, and 2018.

Table A9. Trends in U.S. 8th-graders' average scores and 10th and 90th percentile scores in NAEP mathematics: Selected years 2000 to 2019

Year	10th percentile		s.e.	Average score		s.e.	90th percentile		s.e.
2019	231		(0.4)	282		(0.3)	333		(0.4)
2017	233	*	(0.4)	283	*	(0.3)	333		(0.4)
2009	236	*	(0.5)	283		(0.3)	329	*	(0.5)
2000	223	*	(1.7)	273	*	(0.8)	320	*	(1.0)

* $p < .05$. Score is significantly different from 2019 score.

Note: The NAEP mathematics scale ranges from 0 to 500. Data in this table are for figure 12.

Source: National Center for Education Statistics (NCES), National Assessment of Educational Progress (NAEP), 2000, 2009, 2017, and 2019.

Table A10. Trends in U.S. 8th-graders average scores and 10th and 90th percentile scores in TIMSS mathematics: Selected years 1999 to 2019

Year	10th percentile		s.e.	Average score		s.e.	90th percentile		s.e.
2019	385		(7.2)	515		(4.8)	642		(4.8)
2015	408	*	(4.5)	518		(3.1)	624	*	(4.4)
2011	409	*	(3.5)	509		(2.7)	607	*	(4.7)
1999	387		(5.0)	502	*	(3.9)	611	*	(5.7)

* $p < .05$. Score is significantly different from 2019 score.

Note: The TIMSS mathematics scale ranges from 0 to 1,000. Data in this table are for figure 13.

Source: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1999, 2011, 2015, and 2019.

Table A11. Trends in U.S. 15-year-olds' average scores and 10th and 90th percentile scores in PISA mathematics: Selected years 2003 to 2018

Year	10th percentile		s.e.	Average score		s.e.	90th percentile		s.e.
2018	357		(4.6)	478		(3.2)	598		(4.3)
2015	355		(3.9)	470		(3.2)	585	*	(4.2)
2009	368		(4.3)	487		(3.6)	607		(4.6)
2003	356		(4.5)	483		(2.9)	607		(3.9)

* $p < .05$. Score is significantly different from 2018 score.

Note: The PISA mathematics scale ranges from 0 to 1,000. Data in this table are for figure 14.

Source: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2003, 2009, 2015, and 2018.

Table A12. Trends in U.S. 8th-graders' average scores and 10th and 90th percentile scores in NAEP science: Selected years 2009 to 2019

Year	10th percentile		s.e.	Average score		s.e.	90th percentile		s.e.
2019	106		(0.8)	154		(0.5)	196		(0.6)
2015	109	*	(0.6)	154		(0.3)	195		(0.3)
2009	103	*	(0.6)	150	*	(0.3)	192	*	(0.3)

* $p < .05$. Score is significantly different from 2019 score.

Note: The NAEP science scale ranges from 0 to 500. The NAEP science framework was revised in 2009, so there is no long-term trend period to report in this study. The data in this table are for figure 15.

Source: National Center for Education Statistics (NCES), National Assessment of Educational Progress (NAEP), 2009, 2015, and 2019.

Table A13. Trends in U.S. 8th-graders' average scores and 10th and 90th percentile scores in TIMSS science: Selected years 1999 to 2019

Year	10th percentile		s.e.	Average score		s.e.	90th percentile		s.e.
2019	388		(8.8)	522		(4.7)	642		(4.2)
2015	421	*	(4.2)	530		(2.8)	631	*	(2.9)
2011	416	*	(3.4)	525		(2.4)	625	*	(2.5)
1999	386		(6.9)	515		(4.4)	636		(4.7)

* $p < .05$. Score is significantly different from 2019 score.

Note: The TIMSS science scale ranges from 0 to 1,000. The data in this table are for figure 16.

Source: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1999, 2011, 2015, and 2019.

Table A14. Trends in U.S. 15-year-olds' average scores and 10th and 90th percentile scores in PISA science: Selected years 2009 to 2018

Year	10th percentile		s.e.	Average score		s.e.	90th percentile		s.e.
2018	371		(4.9)	502		(3.3)	629		(3.9)
2015	368		(3.9)	496		(3.2)	626		(3.9)
2009	374		(4.5)	502		(3.6)	629		(5.1)

* $p < .05$. Score is significantly different from 2018 score.

Note: The PISA science scale ranges from 0 to 1,000. The first year science was administered as a "major" domain in PISA was 2006. However, since using that as a starting point for long-term trend would create a narrower timespan than that for other assessments, we forgo a long-term data point in this study. The data in this table are for figure 17.

Source: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009, 2015, and 2018.

Figure A1. Overview of cross-country score trends across 4th-grade ILSAs, by assessment and country

	Long-term trend ~20-year span			Intermediate trend ~10-year span			Recent trend ~ 2- to 5-year span		
	10th percentile	Average	90th percentile	10th percentile	Average	90th percentile	10th percentile	Average	90th percentile
PIRLS Reading									
Hong Kong-CHN	↑	↑	↑	↔	↔	↑	↔	↔	↔
Hungary	↔	↑	↑	↔	↔	↑	↑	↑	↑
Italy	↑	↑	↔	↔	↔	↓	↔	↑	↔
Lithuania	↔	↔	↑	↔	↑	↑	↑	↑	↑
New Zealand	↔	↔	↔	↓	↓	↔	↔	↓	↔
Russian Federation	↑	↑	↑	↑	↑	↑	↑	↑	↑
Singapore	↑	↑	↑	↔	↑	↑	↔	↔	↔
England-GBR[1]	↑	↔	↔	↑	↑	↑	↑	↑	↔
United States	↔	↔	↔	↔	↑	↑	↔	↓	↔
TIMSS Mathematics									
Hong Kong-CHN	↑	↑	↑	↔	↔	↔	↓	↓	↔
Hungary	↔	↔	↔	↑	↔	↔	↔	↔	↓
Italy	↑	↑	↔	↑	↑	↔	↑	↑	↔
Lithuania	↔	↑	↑	↔	↑	↑	↔	↔	↑
New Zealand	↔	↔	↔	↔	↔	↑	↔	↔	↔
Russian Federation	↑	↑	↑	↑	↑	↑	↔	↔	↔
Singapore	↑	↑	↑	↑	↑	↑	↔	↔	↔
England-GBR[1]	↑	↑	↑	↑	↑	↑	↔	↑	↑
United States	↔	↑	↑	↓	↔	↔	↓	↔	↔
TIMSS Science									
Hong Kong-CHN	↓	↓	↔	↔	↔	↔	↓	↓	↓
Hungary	↔	↔	↔	↔	↔	↓	↔	↓	↓
Italy	↑	↔	↓	↔	↓	↓	↔	↔	↔
Lithuania	↑	↑	↑	↑	↑	↑	↔	↑	↑
New Zealand	↔	↓	↓	↔	↔	↔	↔	↔	↔
Russian Federation	↑	↑	↑	↑	↑	↔	↔	↔	↔
Singapore	↑	↑	↑	↑	↑	↔	↔	↔	↔
England-GBR[1]	↔	↔	↓	↑	↑	↔	↔	↔	↔
United States	↔	↔	↔	↓	↔	↔	↓	↓	↔

↑ Upward trend (Most recent score is higher than score in year of referenced term, $p < .05$)

↔ No change (Most recent score is not significantly different from score in year of referenced term, $p < .05$)

↓ Downward trend (Most recent score is lower than score in year of referenced term, $p < .05$)

[1] England is shown out of alphabetical order to match the positioning of the United Kingdom (of which England is a part) in figure A2. IEA studies allow for the participation of jurisdictions within countries, whereas the OECD requires a country-level score for member countries.

Note: See figure 2 for years corresponding to long-term, intermediate, and recent trend periods for each assessment, grade, and subject. See tables A15, A16, and A17 for the data and standard errors on which this figure is based.

Figure A2. Overview of cross-country score trends across 8th-grade and 15-year-old ILSAs, by assessment and country

	Long-term trend ~20-year span			Intermediate trend ~10-year span			Recent trend ~ 2- to 5-year span		
	10th percentile	Average	90th percentile	10th percentile	Average	90th percentile	10th percentile	Average	90th percentile
PISA Reading									
Hong Kong-CHN	↓	↔	↑	↓	↔	↑	↓	↔	↑
Hungary	↔	↔	↔	↓	↓	↔	↔	↔	↔
Italy	↓	↓	↔	↔	↓	↔	↔	↔	↔
Lithuania	—	—	—	↔	↔	↑	↔	↔	↔
New Zealand	↓	↓	↓	↓	↓	↔	↔	↔	↔
Russian Federation	↑	↑	↑	↔	↑	↑	↓	↓	↔
Singapore	—	—	—	↔	↑	↑	↔	↑	↑
United Kingdom	‡	‡	‡	↔	↑	↑	↔	↔	↔
United States	↔	↔	↔	↔	↔	↑	↔	↔	↑
TIMSS Mathematics									
Hong Kong-CHN	↓	↔	↑	↔	↔	↔	↓	↓	↔
Hungary	↓	↓	↔	↔	↑	↑	↔	↔	↔
Italy	↑	↑	↔	↔	↔	↔	↔	↔	↔
Lithuania	↑	↑	↑	↑	↑	↑	↔	↑	↑
New Zealand	↔	↔	↔	↔	↔	↔	↔	↓	↔
Russian Federation	↑	↑	↔	↔	↔	↔	↔	↔	↔
Singapore	↔	↔	↔	↔	↔	↔	↔	↔	↔
England-GBR[1]	↔	↑	↑	↔	↔	↔	↔	↔	↔
United States	↔	↑	↑	↓	↔	↑	↓	↔	↑
PISA Mathematics									
Hong Kong-CHN	↔	↔	↔	↔	↔	↔	↔	↔	↔
Hungary	↔	↔	↓	↔	↔	↔	↔	↔	↔
Italy	↑	↑	↑	↔	↔	↔	↔	↔	↔
Lithuania	—	—	—	↔	↔	↔	↔	↔	↔
New Zealand	↓	↓	↓	↓	↓	↓	↔	↔	↔
Russian Federation	↑	↑	↔	↑	↑	↑	↔	↔	↔
Singapore	—	—	—	↑	↔	↔	↔	↔	↔
United Kingdom	‡	‡	‡	↔	↔	↑	↔	↑	↑
United States	↔	↔	↔	↔	↔	↔	↔	↔	↑

Continued on next page.

Figure A3. Overview of cross-country score trends across 8th-grade and 15-year-old ILSAs, by assessment and country—*continued*

	Long-term trend ~20-year span			Intermediate trend ~10-year span			Recent trend ~ 2- to 5-year span		
	10th percentile	Average	90th percentile	10th percentile	Average	90th percentile	10th percentile	Average	90th percentile
TIMSS Science									
Hong Kong-CHN	↓	↓	↔	↓	↓	↔	↓	↓	↔
Hungary	↓	↓	↓	↔	↔	↑	↔	↔	↔
Italy	↑	↔	↔	↔	↔	↔	↔	↔	↔
Lithuania	↑	↑	↑	↑	↑	↑	↑	↑	↑
New Zealand	↔	↔	↔	↓	↓	↔	↓	↓	↓
Russian Federation	↑	↔	↔	↔	↔	↔	↔	↔	↔
Singapore	↑	↑	↑	↑	↑	↔	↔	↑	↑
England-GBR[1]	↓	↓	↓	↓	↓	↔	↓	↓	↔
United States	↔	↔	↔	↓	↔	↑	↓	↔	↑
PISA Science									
Hong Kong-CHN	—	—	—	↓	↓	↓	↔	↔	↔
Hungary	—	—	—	↓	↓	↔	↔	↔	↔
Italy	—	—	—	↓	↓	↓	↔	↓	↓
Lithuania	—	—	—	↓	↔	↔	↔	↔	↔
New Zealand	—	—	—	↓	↓	↓	↔	↔	↔
Russian Federation	—	—	—	↔	↔	↔	↔	↓	↔
Singapore	—	—	—	↑	↑	↔	↔	↔	↓
United Kingdom	—	—	—	↔	↔	↔	↔	↔	↔
United States	—	—	—	↔	↔	↔	↔	↔	↔

↑ Upward trend (Most recent score is higher than score in year of referenced term, $p < .05$)

↔ No change (Most recent score is not significantly different from score in year of referenced term, $p < .05$)

↓ Downward trend (Most recent score is lower than score in year of referenced term, $p < .05$)

— Not available. Lithuania and Singapore did not participate in PISA 2000. In addition, there is no long-term trend point for PISA science because the science framework was not fully developed until 2006.

‡ Reporting standards not met.

[1] England is shown out of alphabetical order to match the positioning of the United Kingdom (of which England is a part) elsewhere in this figure. IEA studies allow for the participation of jurisdictions within countries, whereas the OECD requires a country-level score for member countries.

Note: NAEP and TIMSS results are for 8th-graders. PISA results are for 15-year-olds. See figure 2 for years corresponding to long-term, intermediate, and recent trend periods for each assessment, grade, and subject. See tables A18-A23 for the data and standard errors on which this figure is based.

Table A15. Fourth-graders' average scores and 10th and 90th percentile scores in PIRLS reading, by education system and year: 2001, 2006, 2011, and 2016

Education system	Year	10th percentile		s.e.	Average		s.e.	90th percentile		s.e.
Hong Kong - CHN	2016	487		(4.4)	569		(2.7)	645		(3.0)
	2011	492		(3.8)	571		(2.3)	643		(2.1)
	2006	486		(4.6)	564		(2.4)	637	*	(2.7)
	2001	445	*	(5.6)	528	*	(3.1)	603	*	(2.8)
Hungary	2016	452		(5.8)	554		(2.9)	645		(3.1)
	2011	435	*	(5.4)	539	*	(2.8)	633	*	(2.6)
	2006	459		(5.8)	551		(2.9)	637	*	(2.5)
	2001	457		(4.0)	543	*	(2.2)	624	*	(2.9)
Italy	2016	461		(5.4)	548		(2.2)	627		(2.8)
	2011	456		(5.3)	541	*	(2.2)	623		(3.9)
	2006	462		(4.5)	551		(2.9)	637	*	(4.0)
	2001	446	*	(4.9)	541	*	(2.4)	627		(3.6)
Lithuania	2016	459		(5.5)	548		(2.6)	632		(2.9)
	2011	440	*	(4.2)	528	*	(2.0)	609	*	(2.7)
	2006	461		(3.3)	537	*	(1.7)	608	*	(2.6)
	2001	458		(4.4)	543		(2.6)	622	*	(3.6)
New Zealand	2016	400		(5.3)	523		(2.2)	630		(3.0)
	2011	410		(3.8)	531	*	(1.9)	639		(3.8)
	2006	415	*	(3.8)	532	*	(2.1)	637		(3.0)
	2001	400		(8.6)	529		(3.7)	640		(4.3)
Russian Federation	2016	495		(4.1)	581		(2.2)	663		(2.6)
	2011	482	*	(4.1)	568	*	(2.7)	649	*	(3.6)
	2006	474	*	(5.7)	565	*	(3.4)	649	*	(4.1)
	2001	443	*	(8.0)	528	*	(4.3)	608	*	(3.8)
Singapore	2016	469		(6.2)	576		(3.2)	673		(4.2)
	2011	459		(6.2)	567		(3.3)	665		(4.0)
	2006	456		(5.0)	558	*	(2.9)	652	*	(3.6)
	2001	402	*	(12.7)	528	*	(5.2)	634	*	(4.7)

Table continued on next page.

Education system	Year	10th percentile		s.e.	Average		s.e.	90th percentile		s.e.
England-GBR[1]	2016	455		(3.3)	559		(1.9)	655		(2.9)
	2011	440	*	(5.0)	552	*	(2.6)	652		(2.7)
	2006	423	*	(4.9)	539	*	(2.5)	645	*	(3.7)
	2001	437	*	(6.6)	553		(3.5)	658		(4.2)
United States	2016	446		(6.3)	549		(3.1)	645		(4.0)
	2011	458		(3.1)	556	*	(1.6)	648		(2.2)
	2006	441		(5.0)	540	*	(3.4)	631	*	(3.7)
	2001	431		(8.8)	542		(3.8)	640		(3.1)

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

[1] England-GBR is shown out of alphabetical order to match the positioning of the United Kingdom (of which England is a part) in figure A2. IEA studies allow for the participation of jurisdictions within countries, whereas the OECD requires a country-level score for member countries.

Note: The PIRLS reading scale ranges from 0 to 1,000.

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

Table A16. Fourth-graders' average scores and 10th and 90th percentile scores in TIMSS mathematics, by education system and year: 2003, 2011, 2015, and 2019

Jurisdiction	Year	10th percentile		s.e.	Average		s.e.	90th percentile		s.e
Hong Kong-CHN	2019	509		(5.8)	602		(3.3)	689		(3.6)
	2015	531	*	(5.0)	615	*	(2.9)	696		(4.1)
	2011	519		(5.8)	602		(3.4)	681		(3.1)
	2003	492	*	(4.5)	575	*	(3.1)	653	*	(3.7)
Hungary	2019	418		(5.4)	523		(2.6)	621		(3.7)
	2015	412		(6.7)	529		(3.2)	635	*	(2.7)
	2011	397	*	(6.6)	515		(3.4)	623		(3.7)
	2003	425		(5.6)	529		(3.2)	624		(4.5)
Italy	2019	427		(3.6)	515		(2.4)	598		(3.4)
	2015	413	*	(4.8)	507	*	(2.6)	596		(2.6)
	2011	414	*	(3.7)	508	*	(2.6)	598		(4.0)
	2003	394	*	(4.8)	503	*	(3.7)	604		(5.3)
Lithuania	2019	442		(4.2)	542		(2.8)	636		(3.5)
	2015	441		(5.1)	535		(2.5)	624	*	(4.7)
	2011	436		(3.9)	534	*	(2.4)	626	*	(3.0)
	2003	433		(4.5)	534	*	(2.7)	625	*	(3.2)

Table continued on next page.

Jurisdiction	Year	10th percentile		s.e.	Average		s.e.	90th percentile		s.e
New Zealand	2019	368		(4.4)	487		(2.6)	602		(3.6)
	2015	371		(4.6)	491		(2.3)	602		(3.2)
	2011	374		(4.9)	486		(2.6)	589	*	(2.8)
	2003	379		(5.8)	493		(2.2)	597		(3.6)
Russian Federation	2019	479		(4.8)	567		(3.3)	653		(4.6)
	2015	470		(5.4)	564		(3.4)	656		(5.8)
	2011	447	*	(4.9)	542	*	(3.7)	635	*	(5.0)
	2003	430	*	(6.3)	532	*	(4.8)	632	*	(7.7)
Singapore	2019	519		(7.0)	625		(3.9)	720		(3.5)
	2015	502		(8.6)	618		(3.8)	722		(3.8)
	2011	502	*	(4.9)	606	*	(3.2)	701	*	(3.3)
	2003	483	*	(9.7)	594	*	(5.6)	696	*	(6.5)
England-GBR[1]	2019	445		(5.1)	556		(3.0)	665		(4.8)
	2015	438		(5.1)	546	*	(2.8)	651	*	(3.6)
	2011	423	*	(6.1)	542	*	(3.5)	652	*	(4.3)
	2003	416	*	(5.8)	531	*	(3.7)	640	*	(4.8)
United States	2019	421		(4.4)	535		(2.5)	639		(2.8)
	2015	432	*	(3.1)	539		(2.3)	640		(2.6)
	2011	440	*	(3.6)	541		(1.9)	635		(2.7)
	2003	417		(3.3)	518	*	(2.4)	614	*	(2.8)

* $p < .05$. Score is significantly different from 2019 score.

[1] England-GBR is shown out of alphabetical order to match the positioning of the United Kingdom (of which England is a part) in figure A2. IEA studies allow for the participation of jurisdictions within countries, whereas the OECD requires a country-level score for member countries.

Note: The TIMSS mathematics scale ranges from 0 to 1,000.

Source: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003, 2011, 2015, and 2019.

Table A17. Fourth-graders' average scores and 10th and 90th percentile scores in TIMSS science, by education system and year: 2003, 2011, 2015, and 2019

Jurisdiction	Year	10th percentile		s.e.	Average		s.e.	90th percentile		s.e
Hong Kong-CHN	2019	438		(5.3)	531		(3.3)	620		(4.0)
	2015	465	*	(6.4)	557	*	(2.9)	644	*	(4.0)
	2011	443		(7.2)	535		(3.7)	622		(3.2)
	2003	465	*	(4.7)	542	*	(3.0)	615		(3.7)

Table continued on next page.

Jurisdiction	Year	10th percentile		s.e.	Average		s.e.	90th percentile		s.e
Hungary	2019	425		(5.0)	529		(2.7)	624		(2.5)
	2015	433		(8.9)	542	*	(3.3)	639	*	(3.6)
	2011	420		(6.5)	534		(3.7)	637	*	(2.9)
	2003	424		(4.3)	530		(2.8)	626		(4.4)
Italy	2019	424		(4.6)	510		(3.0)	592		(4.1)
	2015	429		(4.7)	516		(2.6)	597		(2.3)
	2011	429		(4.4)	524	*	(2.7)	615	*	(3.7)
	2003	405	*	(5.6)	516		(3.8)	620	*	(5.3)
Lithuania	2019	443		(4.3)	538		(2.5)	628		(3.2)
	2015	437		(4.2)	528	*	(2.5)	613	*	(3.1)
	2011	426	*	(3.9)	515	*	(2.4)	598	*	(3.5)
	2003	422	*	(4.2)	512	*	(2.6)	593	*	(2.4)
New Zealand	2019	392		(4.5)	503		(2.3)	605		(3.4)
	2015	389		(5.8)	506		(2.7)	608		(3.1)
	2011	381		(5.1)	497		(2.4)	602		(3.5)
	2003	406		(5.6)	520	*	(2.4)	621	*	(2.8)
Russian Federation	2019	484		(4.4)	567		(3.0)	646		(3.1)
	2015	479		(5.1)	567		(3.2)	653		(4.1)
	2011	458	*	(4.3)	552	*	(3.4)	643		(4.3)
	2003	417	*	(6.5)	526	*	(5.3)	630	*	(6.9)
Singapore	2019	493		(5.8)	595		(3.4)	687		(3.2)
	2015	476		(6.6)	590		(3.7)	692		(4.2)
	2011	469	*	(6.1)	583	*	(3.4)	689		(4.0)
	2003	452	*	(9.5)	565	*	(5.5)	668	*	(5.5)
England-GBR[1]	2019	444		(4.5)	537		(2.7)	626		(4.6)
	2015	445		(4.7)	536		(2.4)	623		(3.0)
	2011	420	*	(5.2)	529	*	(3.0)	629		(3.3)
	2003	430		(6.0)	540		(3.5)	642	*	(5.0)
United States	2019	426		(4.8)	539		(2.7)	641		(2.5)
	2015	439	*	(3.3)	546	*	(2.2)	644		(2.7)
	2011	440	*	(3.3)	544		(2.1)	641		(2.3)
	2003	426		(3.5)	536		(2.5)	636		(3.0)

* $p < .05$. Score is significantly different from 2019 score.

[1] England-GBR is shown out of alphabetical order to match the positioning of the United Kingdom (of which England is a part) in figure A2. IEA studies allow for the participation of jurisdictions within countries, whereas the OECD requires a country-level score for member countries.

Note: The TIMSS science scale ranges from 0 to 1,000.

Source: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003, 2011, 2015, and 2019.

Table A18. Fifteen-year-olds' average scores and 10th and 90th percentile scores in PISA reading, by education system and year: 2000, 2009, 2015, and 2018

Jurisdiction	Year	10th percentile		s.e.	Average		s.e.	90th percentile		s.e
Hong Kong-CHN	2018	390		(5.5)	524		(2.7)	645		(2.5)
	2015	412	*	(4.5)	527		(2.7)	632	*	(3.1)
	2009	418	*	(4.5)	533		(2.1)	634	*	(2.9)
	2000	413	*	(7.3)	525		(2.9)	624	*	(2.9)
Hungary	2018	346		(4.0)	476		(2.3)	602		(3.7)
	2015	338		(4.2)	470		(2.7)	593		(3.2)
	2009	371	*	(6.9)	494	*	(3.2)	607		(3.5)
	2000	354		(5.5)	480		(4.0)	598		(4.4)
Italy	2018	345		(4.6)	476		(2.4)	598		(3.4)
	2015	359		(4.2)	485		(2.7)	602		(2.9)
	2009	358		(2.6)	486	*	(1.6)	604		(1.7)
	2000	368	*	(5.8)	487	*	(2.9)	601		(2.7)
Lithuania	2018	351		(2.7)	476		(1.5)	597		(1.8)
	2015	347		(3.5)	472		(2.7)	593		(4.4)
	2009	353		(4.2)	468		(2.4)	580	*	(3.4)
	2000	—		†	—		†	—		†
New Zealand	2018	362		(3.7)	506		(2.0)	640		(2.9)
	2015	368		(4.5)	509		(2.4)	643		(4.3)
	2009	383	*	(4.5)	521	*	(2.4)	649		(2.7)
	2000	382	*	(5.2)	529	*	(2.8)	661	*	(4.4)
Russian Federation	2018	357		(4.8)	479		(3.1)	597		(3.6)
	2015	381	*	(3.9)	495	*	(3.1)	608		(3.5)
	2009	344		(5.5)	459	*	(3.3)	572	*	(4.5)
	2000	340	*	(5.4)	462	*	(4.2)	579	*	(4.4)
Singapore	2018	398		(3.9)	549		(1.6)	684		(2.5)
	2015	400		(3.7)	535	*	(1.6)	657	*	(2.6)
	2009	394		(3.1)	526	*	(1.1)	648	*	(2.8)
	2000	—		†	—		†	—		†

Table continued on next page.

Jurisdiction	Year	10th percentile		s.e.	Average		s.e.	90th percentile		s.e
United Kingdom[1]	2018	372		(4.3)	504		(2.6)	632		(3.5)
	2015	372		(4.0)	498		(2.8)	621		(3.6)
	2009	370		(3.1)	494	*	(2.3)	616	*	(2.6)
	2000	‡		†	‡		†	‡		†
United States	2018	361		(5.3)	505		(3.6)	643		(3.9)
	2015	364		(5.4)	497		(3.4)	624	*	(3.8)
	2009	372		(3.9)	500		(3.7)	625	*	(5.0)
	2000	363		(11.4)	504		(7.0)	636		(6.5)

— Not available. Lithuania and Singapore did not participate in PISA 2000.

† Not applicable.

‡ Reporting standards not met.

* $p < .05$. Score is significantly different from 2018 score.

[1] In the 4th- and 8th-grade tables, data are only shown for England-GBR because IEA studies allow for the participation of jurisdictions within countries, whereas the OECD requires a country-level score for member countries.

Note: The PISA reading scale ranges from 0 to 1,000.

Source: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2000, 2009, 2015, and 2018.

Table A19. Eighth-graders' average scores and 10th and 90th percentile scores in TIMSS mathematics, by education system and year: 1999, 2011, 2015, and 2019

Jurisdiction	Year	10th percentile		s.e.	Average		s.e.	90th percentile		s.e
Hong Kong-CHN	2019	455		(7.8)	578		(4.1)	686		(7.0)
	2015	489	*	(9.3)	594	*	(4.6)	686		(5.0)
	2011	470		(9.5)	586		(3.9)	684		(4.7)
	1999	489	*	(7.9)	582		(4.3)	670	*	(4.1)
Hungary	2019	398		(5.8)	517		(2.9)	631		(4.7)
	2015	390		(6.8)	514		(3.8)	632		(5.3)
	2011	385		(5.9)	505	*	(3.5)	614	*	(4.0)
	1999	420	*	(6.9)	532	*	(3.6)	636		(4.7)
Italy	2019	405		(3.4)	497		(2.7)	589		(4.1)
	2015	397		(4.4)	494		(2.5)	588		(3.5)
	2011	400		(5.1)	498		(2.3)	590		(3.4)
	1999	366	*	(6.5)	479	*	(3.9)	587		(5.3)
Lithuania	2019	415		(3.7)	520		(2.9)	624		(4.2)
	2015	409		(5.3)	511	*	(2.8)	608	*	(4.1)
	2011	401	*	(4.1)	502	*	(2.5)	602	*	(3.1)
	1999	382	*	(6.7)	482	*	(4.3)	580	*	(8.1)

Table continued on next page.

Jurisdiction	Year	10th percentile		s.e.	Average		s.e.	90th percentile		s.e
New Zealand	2019	367		(6.3)	482		(3.4)	598		(4.0)
	2015	378		(5.8)	493	*	(3.4)	605		(4.6)
	2011	375		(5.5)	488		(5.4)	598		(6.4)
	1999	375		(8.0)	491		(5.3)	604		(6.7)
Russian Federation	2019	438		(7.3)	543		(4.5)	647		(6.1)
	2015	429		(5.7)	538		(4.7)	641		(4.9)
	2011	431		(6.0)	539		(3.6)	641		(5.1)
	1999	416	*	(7.1)	526	*	(5.8)	635		(7.1)
Singapore	2019	487		(10.4)	616		(4.0)	718		(2.8)
	2015	505		(7.9)	621		(3.2)	715		(2.1)
	2011	494		(9.2)	611		(3.8)	713		(3.3)
	1999	501		(9.5)	604		(6.3)	702		(7.8)
England-GBR[1]	2019	398		(8.3)	515		(5.3)	628		(7.6)
	2015	414		(5.5)	518		(4.2)	624		(4.1)
	2011	393		(8.0)	507		(5.6)	616		(6.9)
	1999	392		(6.0)	496	*	(4.2)	602	*	(6.1)
United States	2019	385		(7.2)	515		(4.8)	642		(4.8)
	2015	408	*	(4.5)	518		(3.1)	624	*	(4.4)
	2011	409	*	(3.5)	509		(2.7)	607	*	(4.7)
	1999	387		(5.0)	502	*	(3.9)	611	*	(5.7)

* $p < .05$. Score is significantly different from 2019 score.

[1] England-GBR is shown out of alphabetical order to match the positioning of the United Kingdom (of which England is a part) in figure A2. IEA studies allow for the participation of jurisdictions within countries, whereas the OECD requires a country-level score for member countries.

Note: The TIMSS mathematics scale ranges from 0 to 1,000.

Source: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1999, 2011, 2015, and 2019.

Table A20. Fifteen-year-olds' average scores and 10th and 90th percentile scores in PISA mathematics, by education system and year: 2003, 2009, 2015, and 2018

Jurisdiction	Year	10th percentile		s.e.	Average		s.e.	90th percentile		s.e
Hong Kong-CHN	2018	426		(5.4)	551		(3.0)	667		(3.5)
	2015	426		(5.0)	548		(3.0)	659		(3.5)
	2009	428		(4.9)	555		(2.7)	673		(3.9)
	2003	417		(8.0)	550		(4.5)	672		(4.1)

Table continued on next page.

Jurisdiction	Year	10th percentile		s.e.	Average		s.e.	90th percentile		s.e
Hungary	2018	360		(4.0)	481		(2.3)	597		(3.7)
	2015	351		(4.1)	477		(2.5)	598		(3.5)
	2009	370		(7.1)	490		(3.5)	608		(5.6)
	2003	370		(4.2)	490		(2.8)	611	*	(4.7)
Italy	2018	363		(4.7)	487		(2.8)	605		(3.9)
	2015	368		(3.8)	490		(2.8)	610		(3.8)
	2009	363		(2.4)	483		(1.9)	602		(2.5)
	2003	342	*	(5.9)	466	*	(3.1)	589	*	(3.6)
Lithuania	2018	362		(3.6)	481		(2.0)	598		(2.8)
	2015	365		(3.8)	478		(2.3)	590		(3.5)
	2009	363		(4.2)	477		(2.6)	590		(4.0)
	2003	—		†	—		†	—		†
New Zealand	2018	372		(3.0)	494		(1.7)	614		(2.2)
	2015	375		(3.8)	495		(2.3)	613		(3.1)
	2009	392	*	(4.4)	519	*	(2.3)	642	*	(3.9)
	2003	394	*	(3.9)	523	*	(2.3)	650	*	(3.2)
Russian Federation	2018	376		(4.3)	488		(3.0)	597		(3.9)
	2015	387		(4.6)	494		(3.1)	601		(3.8)
	2009	360	*	(4.5)	468	*	(3.3)	576	*	(5.3)
	2003	351	*	(5.0)	468	*	(4.2)	588		(5.3)
Singapore	2018	441		(2.9)	569		(1.6)	684		(2.7)
	2015	436		(2.6)	564		(1.5)	682		(2.4)
	2009	422	*	(4.1)	562		(1.4)	693		(2.5)
	2003	—		†	—		†	—		†
United Kingdom[1]	2018	381		(4.0)	502		(2.6)	620		(3.3)
	2015	371		(3.7)	492	*	(2.5)	610	*	(3.1)
	2009	380		(3.1)	492		(2.4)	606	*	(3.9)
	2003	‡		†	‡		†	‡		†
United States	2018	357		(4.6)	478		(3.2)	598		(4.3)
	2015	355		(3.9)	470		(3.2)	585	*	(4.2)
	2009	368		(4.3)	487		(3.6)	607		(4.6)
	2003	356		(4.5)	483		(2.9)	607		(3.9)

— Not available. Lithuania and Singapore did not participate in PISA 2003.
† Not applicable.
‡ Reporting standards not met.
* $p < .05$. Score is significantly different from 2018 score.
[1] In the 4th- and 8th-grade tables, data are only shown for England because IEA studies allow for the participation of jurisdictions within countries, whereas the OECD requires a country-level score for member countries.
Note: The PISA mathematics scale ranges from 0 to 1,000.
Source: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2003, 2009, 2015, and 2018.

Table A21. Eighth-graders' average scores and 10th and 90th percentile scores in TIMSS science, by education system and year: 1999, 2011, 2015, and 2019

Jurisdiction	Year	10th percentile		s.e.	Average		s.e.	90th percentile		s.e
Hong Kong-CHN	2019	373		(8.7)	504		(5.2)	622		(6.2)
	2015	454	*	(6.8)	546	*	(3.9)	631		(4.7)
	2011	434	*	(6.4)	535	*	(3.4)	622		(3.7)
	1999	442	*	(6.8)	530	*	(3.5)	614		(4.0)
Hungary	2019	421		(5.2)	530		(2.6)	632		(4.8)
	2015	416		(7.1)	527		(3.4)	633		(4.2)
	2011	412		(5.9)	522		(3.1)	621	*	(3.3)
	1999	445	*	(5.6)	552	*	(3.6)	655	*	(5.8)
Italy	2019	404		(4.5)	500		(2.6)	593		(2.7)
	2015	397		(6.4)	499		(2.4)	593		(3.5)
	2011	400		(5.3)	501		(2.4)	594		(2.9)
	1999	379	*	(6.5)	493		(4.0)	602		(4.8)
Lithuania	2019	433		(4.9)	534		(3.0)	630		(4.2)
	2015	416	*	(5.0)	519	*	(2.8)	616	*	(4.3)
	2011	413	*	(3.0)	514	*	(2.5)	607	*	(3.2)
	1999	383	*	(6.2)	488	*	(4.1)	592	*	(5.7)
New Zealand	2019	375		(6.5)	499		(3.5)	613		(3.4)
	2015	392	*	(5.0)	513	*	(3.1)	625	*	(3.9)
	2011	399	*	(6.2)	512	*	(4.6)	621		(4.8)
	1999	388		(8.9)	510		(5.1)	623		(6.1)
Russian Federation	2019	442		(6.3)	543		(4.2)	637		(5.4)
	2015	442		(6.2)	544		(4.2)	640		(5.0)
	2011	442		(5.4)	542		(3.3)	638		(3.5)
	1999	411	*	(8.9)	529		(6.4)	647		(10.2)
Singapore	2019	485		(9.4)	608		(3.9)	708		(3.2)
	2015	475		(8.7)	597	*	(3.2)	696	*	(2.3)
	2011	453	*	(9.0)	590	*	(4.3)	705		(3.6)
	1999	439	*	(13.3)	568	*	(8.0)	688	*	(8.7)

Table continued on next page.

Jurisdiction	Year	10th percentile		s.e.	Average		s.e.	90th percentile		s.e
England-GBR[1]	2019	393		(8.6)	517		(4.8)	630		(6.6)
	2015	428	*	(5.9)	537	*	(3.8)	640		(4.6)
	2011	419	*	(9.2)	533	*	(4.9)	639		(6.0)
	1999	424	*	(6.7)	538	*	(4.8)	654	*	(8.6)
United States	2019	388		(8.8)	522		(4.7)	642		(4.2)
	2015	421	*	(4.2)	530		(2.8)	631	*	(2.9)
	2011	416	*	(3.4)	525		(2.4)	625	*	(2.5)
	1999	386		(6.9)	515		(4.4)	636		(4.7)

* $p < .05$. Score is significantly different from 2019 score.

[1] England-GBR is shown out of alphabetical order to match the positioning of the United Kingdom (of which England is a part) in figure A2. IEA studies allow for the participation of jurisdictions within countries, whereas the OECD requires a country-level score for member countries.

Note: The TIMSS science scale ranges from 0 to 1,000.

Source: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 1999, 2011, 2015, and 2019.

Table A22. Fifteen-year-olds' average scores and 10th and 90th percentile scores in PISA science, by education system and year: 2009, 2015, and 2018

Jurisdiction	Year	10th percentile		s.e.	Average		s.e.	90th percentile		s.e
Hong Kong-CHN	2018	401		(4.3)	517		(2.5)	623		(3.3)
	2015	413		(4.5)	523		(2.5)	622		(2.7)
	2009	432	*	(4.9)	549	*	(2.8)	655	*	(2.9)
Hungary	2018	356		(3.9)	481		(2.3)	602		(3.6)
	2015	347		(4.1)	477		(2.4)	601		(3.5)
	2009	388	*	(7.6)	503	*	(3.1)	609		(3.6)
Italy	2018	348		(3.9)	468		(2.4)	583		(3.7)
	2015	359		(3.8)	481	*	(2.5)	599	*	(2.8)
	2009	362	*	(2.6)	489	*	(1.8)	609	*	(2.0)
Lithuania	2018	364		(2.9)	482		(1.6)	599		(2.3)
	2015	357		(3.8)	475		(2.7)	597		(3.7)
	2009	382	*	(4.9)	491		(2.9)	600		(3.9)
New Zealand	2018	371		(3.7)	508		(2.1)	640		(2.9)
	2015	374		(3.8)	513		(2.4)	647		(3.5)
	2009	390	*	(4.3)	532	*	(2.6)	667	*	(3.3)
Russian Federation	2018	369		(4.1)	478		(2.9)	586		(3.7)
	2015	379		(3.8)	487	*	(2.9)	595		(3.5)
	2009	364		(4.7)	478		(3.3)	594		(4.6)

Table continued on next page.

Jurisdiction	Year	10th percentile		s.e.	Average		s.e.	90th percentile		s.e
Singapore	2018	416		(3.2)	551		(1.5)	670		(1.8)
	2015	412		(2.8)	556		(1.2)	683	*	(2.2)
	2009	401	*	(3.1)	542	*	(1.4)	673		(3.0)
United Kingdom[1]	2018	374		(3.8)	505		(2.6)	632		(3.2)
	2015	377		(3.2)	509		(2.6)	638		(3.2)
	2009	385		(3.6)	514		(2.5)	640		(3.3)
United States	2018	371		(4.9)	502		(3.3)	629		(3.9)
	2015	368		(3.9)	496		(3.2)	626		(3.9)
	2009	374		(4.5)	502		(3.6)	629		(5.1)

* $p < .05$. Score is significantly different from 2018 score.

[1] In the 4th- and 8th-grade tables, data are only shown for England because IEA studies allow for the participation of jurisdictions within countries, whereas the OECD requires a country-level score for member countries.

Note: The PISA science scale ranges from 0 to 1,000. The first year science was administered as a "major" domain in PISA was 2006. However, since using that as a starting point for long-term trend would create a narrower timespan than that for other assessments, we forgo a long-term data point in this study. Thus, there are only two comparison years for 2018 in this table.

Source: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009, 2015, and 2018.

About the American Institutes for Research

Established in 1946, with headquarters in Arlington, Virginia, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance to solve some of the most urgent challenges in the U.S. and around the world. We advance evidence in the areas of education, health, the workforce, human services, and international development to create a better, more equitable world. The AIR family of organizations now includes IMPAQ, Maher & Maher, and Kimetrica. For more information, visit AIR.ORG.

AIR® Headquarters
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
+1.202.403.5000 | **AIR.ORG**