Testing Theories of Why

Four Keys to Interpreting US Student Achievement Trends

Nat Malkus

JANUARY 2025

AMERICAN ENTERPRISE INSTITUTE

Executive Summary

Although national test scores provide clear evidence on student achievement across time, they do not illuminate what is driving gains or losses. Nonetheless, careful examination of test scores can corroborate some explanations for changes in student achievement and discount others.

This report examines recent trends in US student achievement, as measured by national and international assessments, and identifies four key trends that any satisfactory explanation of recent US student performance should account for: a downward trend beginning around 2013; declines driven by the bottom half of the distribution, both before and after the pandemic; higher absolute achievement gap growth in the US than other nations; and adult assessment trends that closely match those of students. The report concludes by evaluating how common explanations of student achievement trajectories align with these trends.

Testing Theories of Why

Four Keys to Interpreting US Student Achievement Trends

Nat Malkus

Once or twice a year, a familiar sequence of events repeats itself: A new round of national test scores are released. The results make the news because they are new and because test scores are important. And then gallons of ink are spilled arguing why scores changed as they did. It's cable TV, or MTV, or violent video games, or the internet, or the phones. It's the teachers unions, or the parents, or the inaction of policymakers, or the meddling of No Child Left Behind. It's the advent of MySpace—no, Facebook—or Instagram, definitely TikTok. It's the obesity epidemic, the loss of recess, the overemphasis on sports, the shriveling of arts and music, or the lack of extracurricular options. It's students' behavior, or lack of ambition, or increasing disrespect, or timidity, or mental health, or unhealthy perfectionism, or laziness.

National test scores are important. They provide us with a big picture of student achievement at the national level and tell us whether achievement is improving, falling, or holding steady. For those concerned with the effectiveness of the schools we invest so much in, the trajectory of the children we care for, or the future of the economy we depend on, national test scores are valuable indicators.

However, although national test scores do provide strong evidence of student achievement across time, they are not great at telling us why students get the scores they do. Test scores tell us where students are academically, not why they are there. Nonetheless, careful examination of test scores can provide circumstantial evidence to corroborate some possible explanations for changes in student achievement and discount others.

In the wake of the pandemic, test score releases may be more important than ever. Two years ago, scores from the 2022 administration of the National Assessment of Educational Progress (NAEP)—the self-styled "Nation's Report Card"—showed just how far students had fallen behind during the pandemic. And in late January 2025, the release of scores from the 2024 NAEP assessment, the second post-pandemic

NAEP, will help education researchers, pundits, and policymakers answer one of the biggest questions in education right now: Are students and schools recovering from COVID learning loss?

But even that anodyne question includes at least three implicit stories of student achievement that, I argue in this report, are not supported by a careful analysis of recent test score trends. The first implicit story is that declining student performance in recent years is solely attributable to the pandemic: If test scores have declined, that is the result of the pandemic's aftereffects; if test scores have improved, that is because we have moved past the pandemic.

Insofar as the pandemic was probably the biggest disruption to American schooling in over a century, this story has a lot to recommend it. However, this story suffers from pandemic myopia, leading to a woefully incomplete picture of how performance has declined in recent years and why. Indeed, average test scores began declining well before the pandemic, suggesting that the decline in test scores over the pandemic was perhaps not solely attributable to the pandemic and that test scores could continue to decline in the coming years for reasons that have little to do with COVID-19.

The second implicit story is that things are getting much worse for all students. For good reason, average scores get the headlines, and averages have been trending downward. However, averages can mask what's going on underneath, and over the past decade, test scores show a sharp divergence in how high-achieving and low-achieving students are performing. In fact, on a number of tests and in a number of subjects, higher-achieving students have held steady over the past decade.

The third implicit story is that changes to student performance are the result of what happens in schools: Tests assess students on the skills they practice in the classroom, and so changes to test scores suggest changes to what happens in the classroom. For example, although the pandemic was not something that happened only in schools, the pandemic affected instructional quality, leading to declines in test scores.

While understandable, recent test scores suggest that factors outside of school might play a considerable role. To wit, during the pandemic, test scores declined significantly for American adults, the vast majority of whom were not in school at any point during the pandemic. I argue that, while perplexing, these changes to adult test scores align too well with changes to K–12 test scores to be ignored.

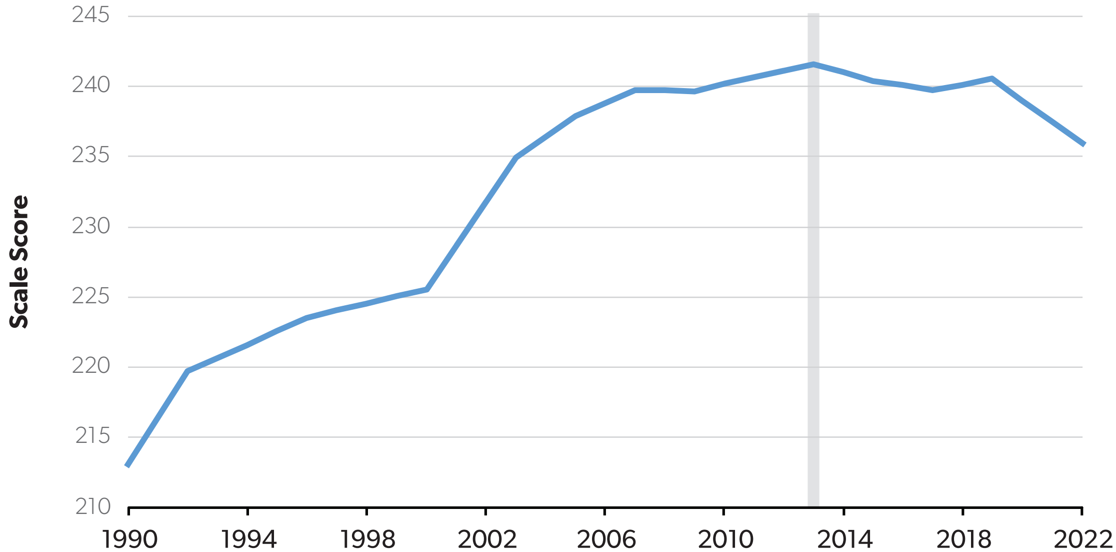
Changes to test scores are not easy to account for, and in this report I don't try to. Instead, I look at test results from a number of national and international tests and identify several patterns and trends that any satisfactory explanation of student performance over the past decade needs to account for.

In this report, I analyze longitudinal average and selected percentile scores from a range of nationally representative assessments spanning various subjects, grades, and age groups. The main NAEP assessments cover reading, math, science, and history for grades four, eight, and 12 from 1992 to 2022. The NAEP long-term trend (LTT) assessments gauge reading and math for ages nine and 13 from 1978 to 2022. The Trends in International Mathematics and Science Study (TIMSS) assesses science and math for grades four and eight from 1995 to 2003. The Program for International Student Assessment (PISA) evaluates 15-year-olds in reading, math, and science from 2000 to 2022. I also include tests of numeracy and literacy in adult populations, using the Program for the International Assessment of Adult Competencies (PIAAC) with other linked assessments, spanning 1998 to 2023. Together, these assessments provide a comprehensive view of US educational and skill trends across different populations over the past two decades.

Four Aspects of Recent Test Trends

This section highlights four important aspects of recent trends in US tests and contains several figures that display many assessments together. I begin with one assessment, NAEP fourth-grade math, to demonstrate how I transform scores in this report to look at similar trends across many datasets.

Figure 1. Average NAEP Fourth-Grade Math Scores: 1990–2022



Source: NAEP, Fourth-Grade Math, 1990–2022. See Appendix A for details.

Test Score Highs in the Early 2010s

Student test scores across numerous assessments peaked in the early 2010s. Figure 1 shows this peak for average fourth-grade math scores. Fourth-grade math scores grew rapidly from 1990 to 2006, after which growth slowed, with scores reaching their high point in 2013. After this peak, scores trended downward, declining sharply during the pandemic.

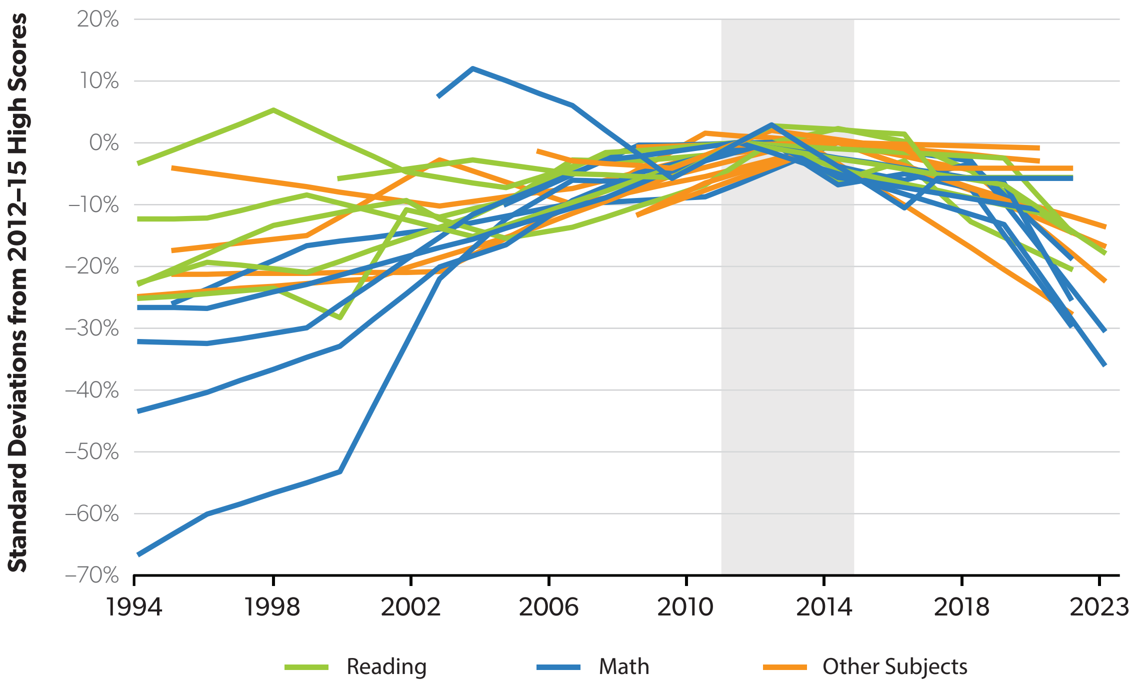
Similar patterns are evident across 21 nationally representative assessments displayed in Figure 2, which shows reading assessments in green, math in blue, and all other subjects in orange. Since these assessments were taken in different years and have different scales, I made two adjustments. First, all scores are converted into standard deviation units for a common scale. And second, each assessment is scaled relative to *its* highest score in the period from 2012 to 2015, highlighted in gray. These transformations mean that each assessment is equal to zero for at least one of the high-scoring years between 2012 and 2015 and, generally, that no scores are higher than zero during this span.[1] For example, because NAEP fourth-grade math scores reached their peak in 2013, they are equal to zero for that year. All other years for NAEP fourth-grade math show the percentage difference in standard deviations between 2013 and that year.

Several patterns stand out in this figure. First, the high point for nearly all these exams is during the highlighted 2012–15 period, with few points on any assessment above the 0 percent line. Second, before those peaks, math scores (blue) show greater improvements than do reading scores (green). Third, declines following this peak are generally steeper in math than in reading. And fourth, the post-pandemic declines are generally steeper compared with any line pre-pandemic, but the downward trends begin before the pandemic, in 2012–15.

Declines Are Driven by the Bottom Half of Scorers

Further analysis reveals that these declines were driven largely by lower-performing students. Instead of average scores, Figure 3, Panel A shows fourth-grade NAEP math scores for five different percentiles: the

Figure 2. US Average Scores on 21 Nationally Representative Achievement Tests: 1994–2023



Source: NAEP; NAEP LTT; PISA; and TIMSS. Select years from 1994 to 2023. See Appendix A for details.

10th, 25th, 50th, 75th, and 90th. Like the averages in Figure 1, the high-water mark for these percentiles is in 2013, but this figure shows larger post-2013 losses for the lower percentiles.

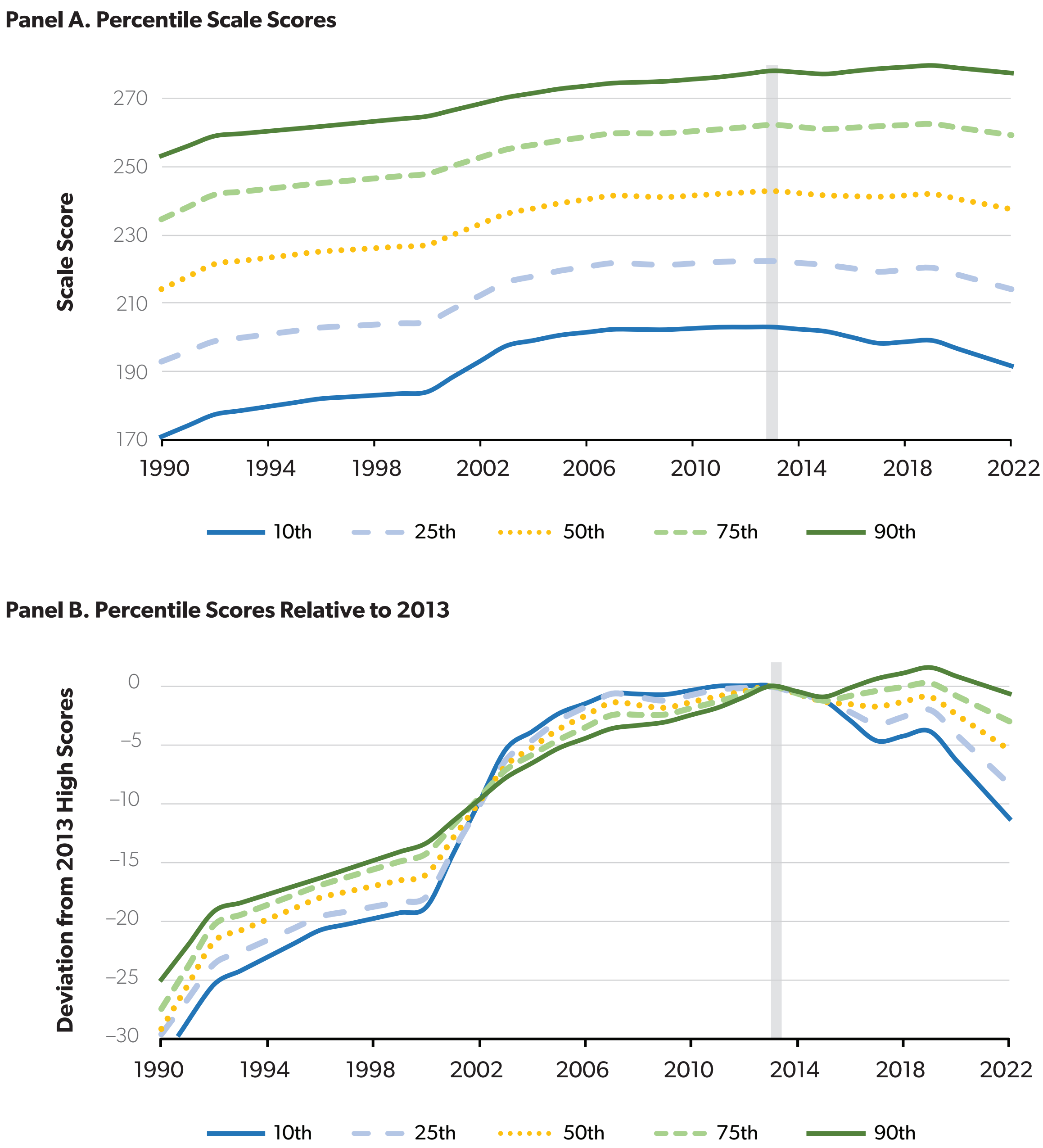
In Panel A, these differences in score declines look relatively small because the scale of the graph, at 100 points, emphasizes the differences between percentiles rather than the differences within percentiles. Accordingly, in Figure 3, Panel B, I show how percentile scores change from their peak in 2013. Subtracting the peak 2013 score for each group from that group's score in a given year makes the 2013 score for each group equal to zero and shrinks the figure's scale markedly, from 100 to 30 points.

Panel B makes the relative changes within percentile groups come into sharp relief: Between 2013 and the pandemic, scores for the 10th and 25th percentiles fell, while those for the 75th and 90th percentiles rose slightly. After the pandemic, scores fell across the board, but much more for students in the lower percentiles.

We see similar trends in other assessments. With so many assessments, displaying all percentile trends in a single figure is difficult, so I split the percentile groups into two figures. Figure 4, Panel A uses the same transformations as Figure 3, Panel B to show the 90th and 10th percentiles (in green and blue, respectively), and Figure 4, Panel B shows the 75th and 25th percentiles (in dashed green and blue lines, respectively). In both cases, the colors are jumbled before the peak period, meaning that score increases (or decreases) between percentile groups were more similar.

However, after 2012–15, there is a stark separation in trends across percentile groups: The lower percentiles trend downward, while the relatively higher percentiles are closer to zero. These lines show a remarkably consistent pattern across subjects and

Figure 3. NAEP Fourth-Grade Math Percentile Scores and Percentile Deviations from 2013: 1990–2022



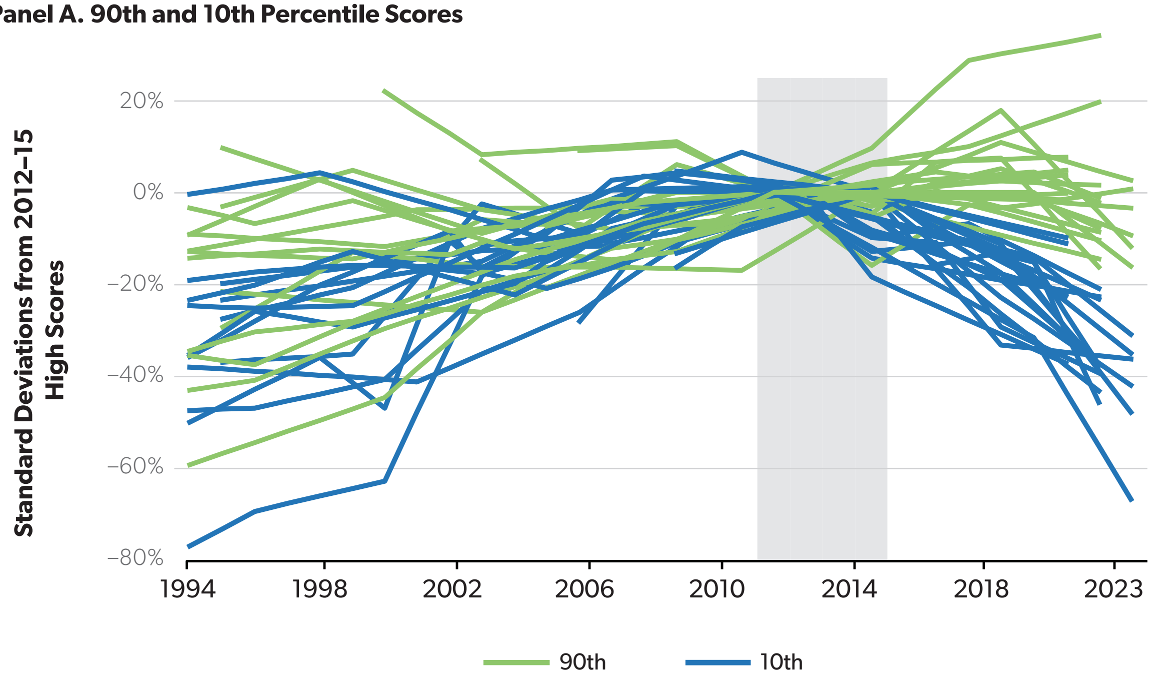
Source: NAEP, Fourth-Grade Math, 1990–2022. See Appendix A for details.

grades, whereby scores are not falling equally across the achievement distribution. Since the early 2010s, assessment scores vary, but scores for higher performers are relatively stable while scores for the bottom of the distribution fall rapidly, meaning that absolute achievement gaps have widened considerably.

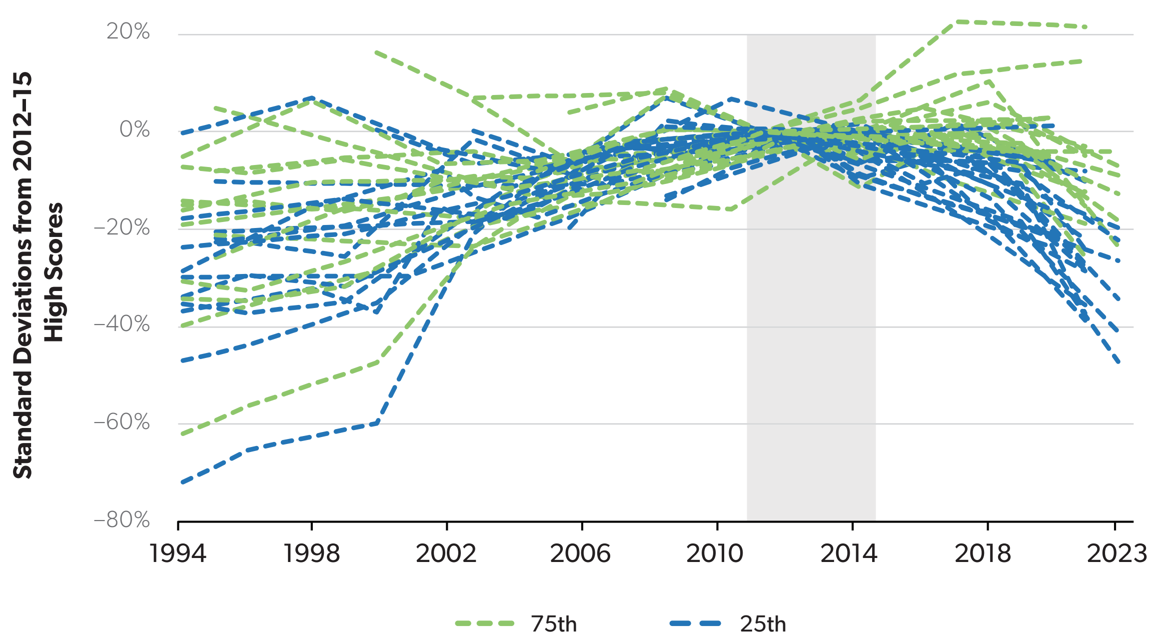
Adding to this pattern is the consistency of US subnational differences. Chad Aldeman examined

Figure 4. US Percentile Scores on 21 Nationally Representative Achievement Tests: 1994–2023

Panel A. 90th and 10th Percentile Scores



Panel B. 75th and 25th Percentile Scores



Source: NAEP; NAEP LTT; PISA; and TIMSS. Select years from 1994 to 2023. See Appendix A for details.

subnational scores on NAEP assessments for the same patterns examined above, finding that "overall, 49 of 50 states, the District of Columbia and 17 out of 20 of the large cities that participated in NAEP saw a widening of their achievement gap over the last decade."[2] That is strong evidence that these patterns are not just national but pervasive across nearly every part of the nation.

The US Is the International Leader in Achievement Gap Growth

These figures show US scores, but some of these scores come from international tests, such as the TIMSS and PISA, in which the US participates. As we might expect, US results on these international assessments show the same patterns that the NAEP assessments show; however, international results also reveal that this pattern is a peculiarly American phenomenon. Indeed, results from other countries do not reflect, either in consistency or degree, the same dual trend of a decline starting in the early 2010s and an increasing absolute achievement gap between the highest and lowest performers.

For instance, on PISA tests between 2012 and 2022, the interquartile gap (between the 25th and 75th percentiles) in the US grew by 10 points in math, 24 points in science, and 30 points in reading. However, even the 10-point growth in math is far larger than the one-point growth in the average interquartile gap across Organisation for Economic Co-operation and Development countries.[3] Likewise, TIMSS results in grades four and eight for both science and math show not only large absolute achievement gap growth in the US after 2011 but also how much the US's absolute achievement gap growth during this period stands out.

On TIMSS, achievement gap growth is common but not universal, but in the US, achievement growth was pronounced in both subjects in both grade levels. Figure 5, which depicts the raw estimates of gap growth between the 10th and 90th percentiles for both fourth-grade and eighth-grade math, shows that from 2011 to 2023, absolute achievement growth on these tests was larger in the US than in any country with comparable data.[4] Similarly, in fourth-grade science, the US led the pack in achievement gap growth between the 10th and 90th percentiles, and in eighth-grade science, the US had the fourth-largest achievement gap growth of 25 countries, after the United Arab Emirates, Hong Kong, and Sweden.[5] While not all of these differences are statistically significant, the point is not that any one comparison stands out but that the US is consistently the international leader in achievement gap growth.

The Falling Bottom Is Not Just Students

Thus far we have seen that US student scores peaked in the early 2010s, that subsequent declines were driven primarily by the lower half of scorers, and that this pattern is uncommonly large in the US. Now I turn to look at nationally representative assessment scores not for students but for US adults. These scores have potentially profound implications for explaining why student scores changed as they did.

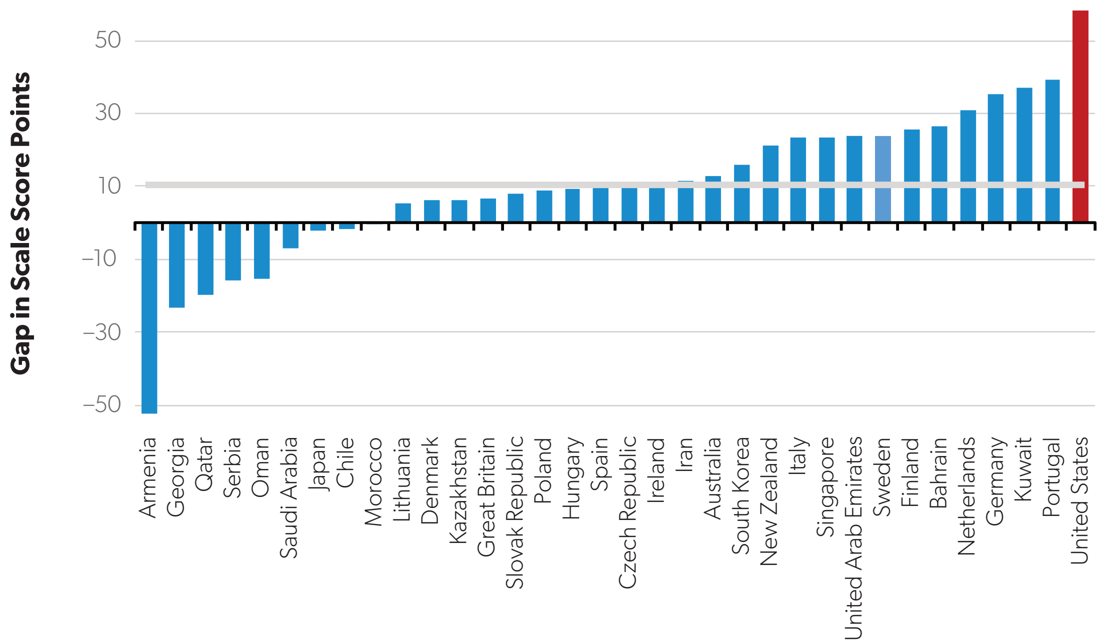
Data from the PIAAC reveal patterns in adult literacy and numeracy scores that are similar to results on student assessments like NAEP and PISA.[6] Unlike student assessments, however, which measure cross-sections of cohorts, PIAAC evaluates the adult population (ages 16–65) over time. This means that the target population is largely the same over five or even 10 years.

For three reasons, considerable caution should be taken when interpreting the most recent round of US PIAAC assessments. The first is the low US response rate, about 28 percent for 2023, which could bias results. The second is the introduction of digitally administered assessments, a mode change that could also introduce bias. A third concern is that the National Center for Education Statistics changed the PIAAC 2023 assessment design to more accurately capture the performance of the lowest-performing participants. Unlike the first two reasons, these design changes might be expected to lower the scores on the bottom percentiles of performance, though by how much is unknown. Nonetheless, the trends in PIAAC results are too large—and too similar to results on student assessments—to completely ignore.

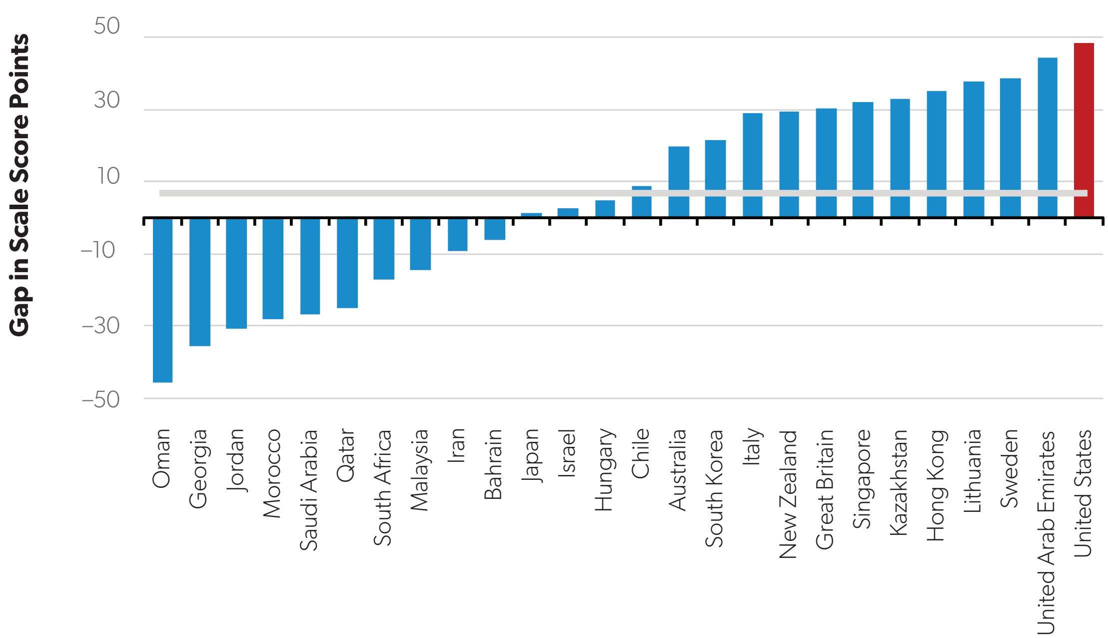
PIAAC results in Figure 6 show steep declines in both literacy and numeracy after 2017. Somewhat

Figure 5. TIMSS Math 90th and 10th Percentile Achievement Gap Growth: 2011–23

Panel A. TIMSS Fourth-Grade Math

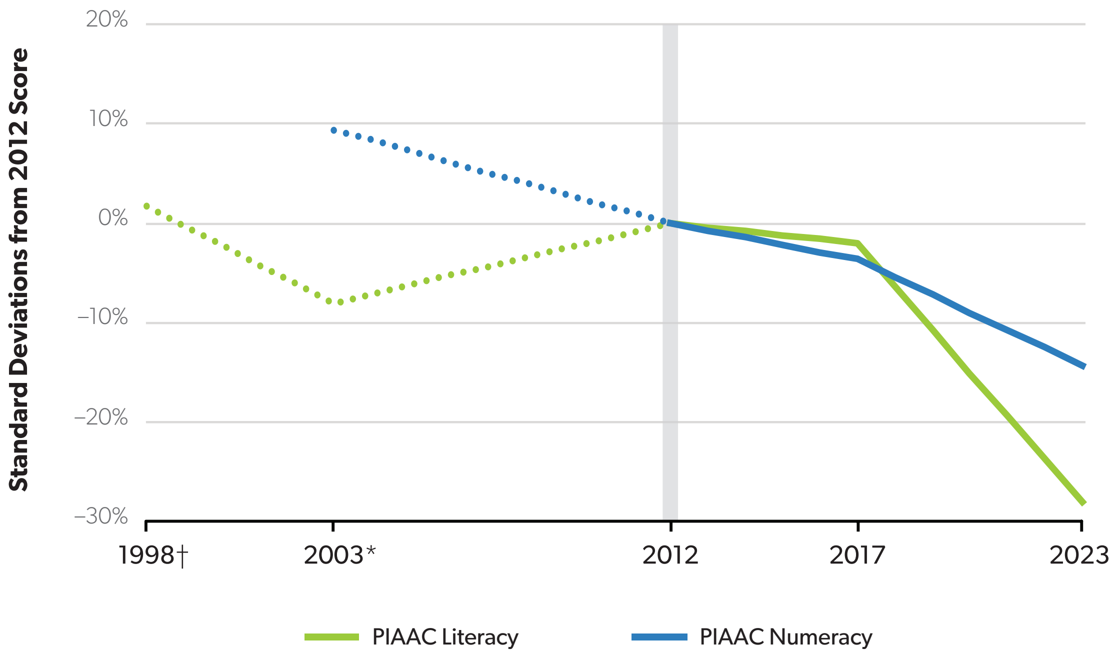


Panel B. TIMSS Eighth-Grade Math



Source: IEA; and TIMSS. Select years from 2011 to 2023. See Appendix A for details.

Figure 6. PIAAC and Equated Adult Literacy and Numeracy Standardized Scores: 1998–2023



Source: PIAAC, 2012–23; ALL, 2003–08; and IALS, 1994–98. See Appendix A for details.
Note: † Scores for 1998 are from IALS. * Scores for 2003 are from ALL.

similar to student scores, these declines occur a bit later than the 2011–15 high-point period on student assessments. The equated scores drawn from prior assessments of US adults (dashed lines) suggest that these declines are not wholly unprecedented.

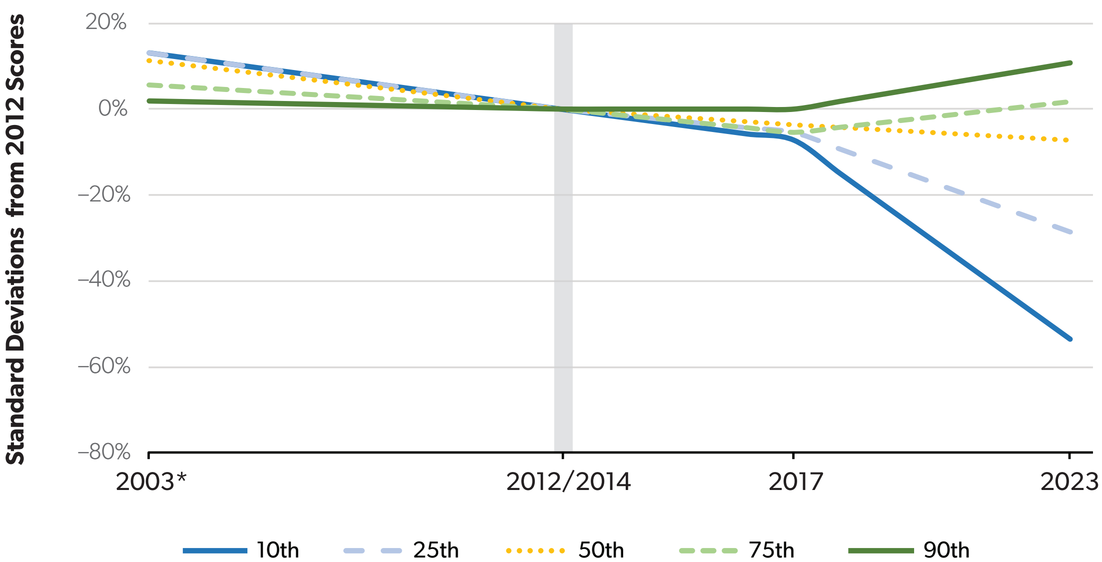
Figure 7 shows that PIAAC declines in US adult numeracy and literacy are overwhelmingly attributable to the lower half of scorers, the same pattern seen for US students, but starting slightly later. After 2017, PIAAC numeracy scores trend upward or flat for the 50th, 75th, and 90th percentiles but fall dramatically for the 25th and 10th percentiles. The percentile trajectories over the pre-2017 period are much more similar to each other than they are after 2017.

The similarities that adult PIAAC score trajectories have with student assessments are striking. Figure 8 adds PIAAC scores for the 90th percentiles (dotted green lines) and 10th perecentiles (dotted blue lines) from 2012 to 2023 to the same set of student percentiles displayed in Figure 4, Panel A. Dramatic declines for the 10th percentile scores and relative stability for the 90th percentile scores are clearly evident for both US students and adults. Even if these declines are overstated due to recent changes to US PIAAC assessments, changes half this size would be sufficiently similar to warrant concern.

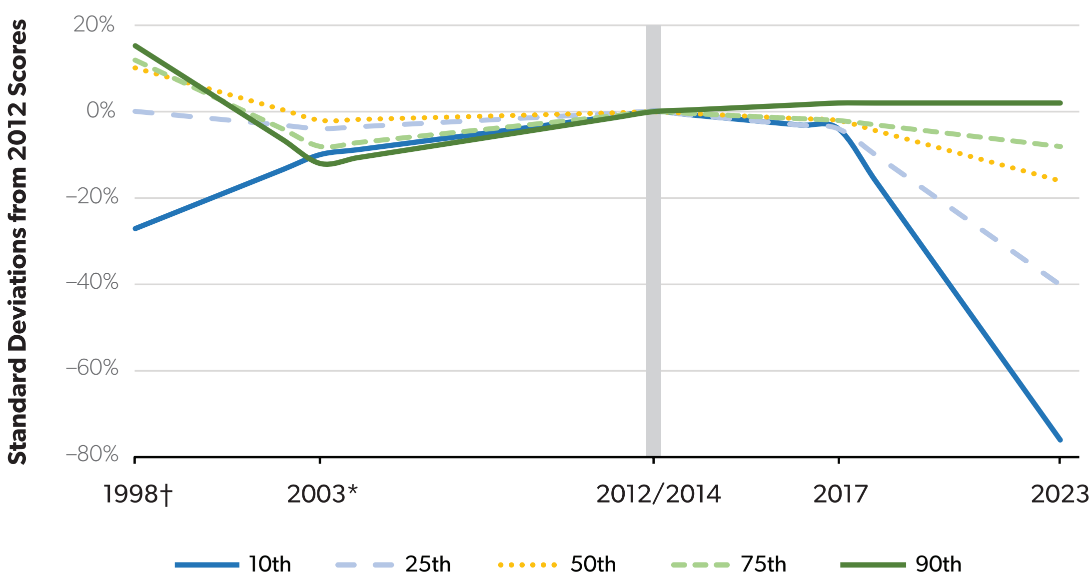
The PIAAC data highlight another important pattern: The changes observed across tests suggest a shift affecting the entire population of test takers, rather than a shift for a single cohort of students whose effects cascade over time. A cohort-specific shift—such as an event affecting fourth graders at a particular time—should be reflected in fourth-grade assessments and then show in eighth-grade results four years later and in twelfth-grade results four years after that. However, student assessment results do not fit well with a cohort cascading shift, and the PIAAC results add more weight to a shift that hit the broader population, or at least the population of lower performers.

Figure 7. US PIAAC Percentile Scores in Numeracy and Literacy: 1998–2023

Panel A. US PIAAC Numeracy Percentile Scores

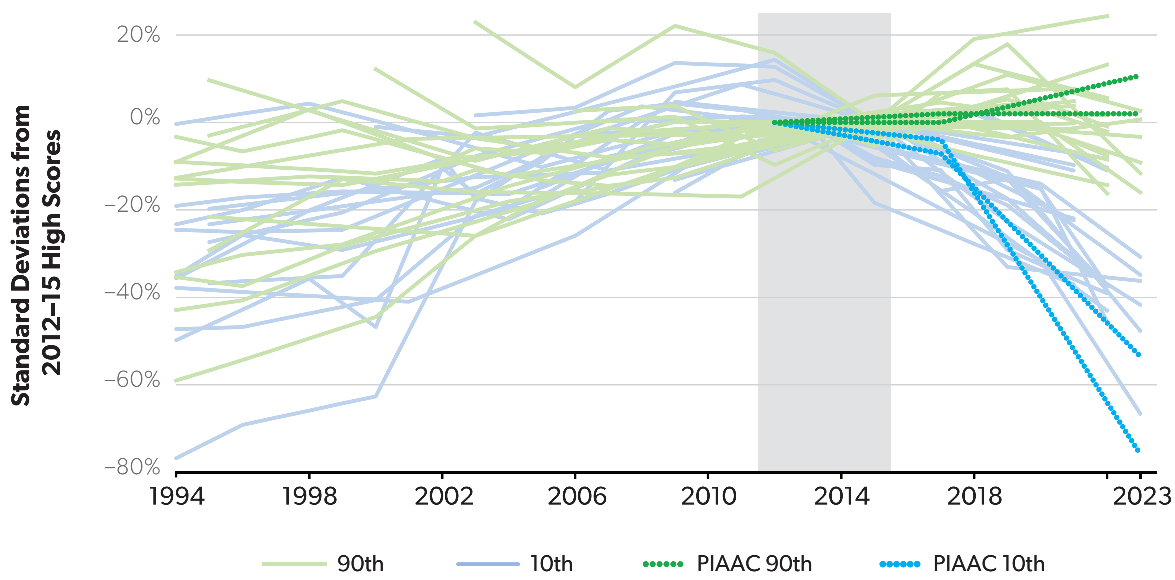


Panel B. PIAAC Percentile Scores in Literacy



Source: PIAAC, 2012–23; ALL, 2003–08; and IALS, 1994–98. See Appendix A for details.
Note: † Scores for 1998 are from IALS. * Scores for 2003 are from ALL.

Figure 8. 21 Student Assessments Plus US Adult PIAAC 90th and 10th Percentile Scores: 1995–2023



Source: NAEP; NAEP LTT; PISA; TIMSS; and PIAAC. Select years from 1994 to 2023. See Appendix A for details.

PIAAC scores show one more similarity to student scores. US adult achievement gap growth between the 10th and 90th percentiles from 2012 to 2023 is greater than in every other comparison country in numeracy and all but New Zealand in literacy (Figure 9). Across comparable countries, achievement gap growth was more common on PIAAC than on TIMSS; however, the US clearly has a peculiar case of widening gaps during this period.

The 2023 PIAAC results warrant serious caution, however: The bias that low response rates and design changes might bring is unlikely to produce such similarities in direction and magnitude to the student results, including a fall driven by the bottom half of scorers and greater achievement gap growth than seen in other countries. Indeed, attributing this pattern to nonresponse would not only require dismissing an unlikely set of parallels but also dismiss an important means of interpreting student test score changes over the past dozen years that could not be done with any other available data.

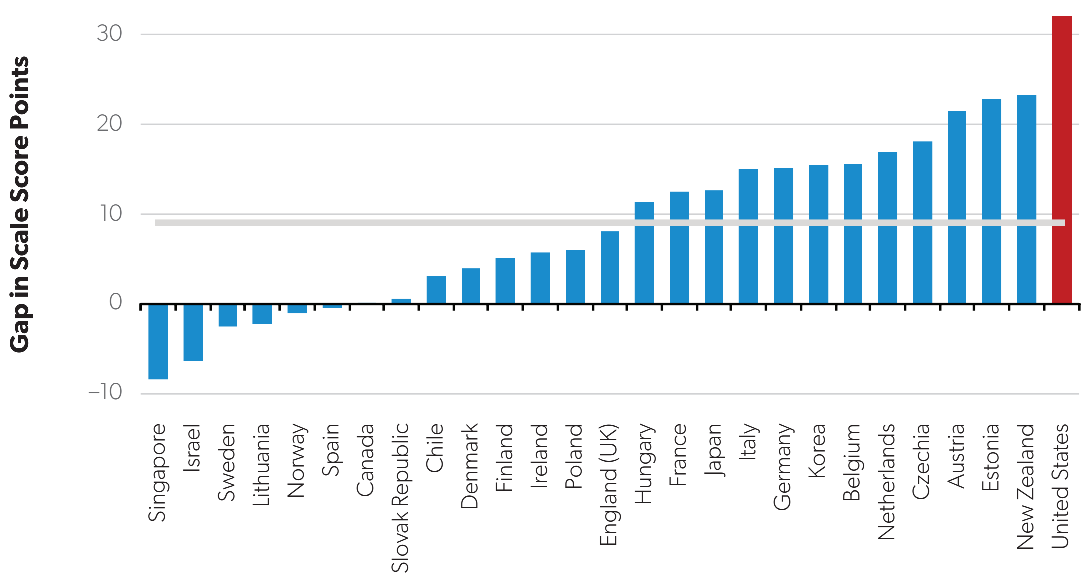
The PIAAC results suggest the pattern of US declines in numeracy and literacy in the bottom half of the distribution is not particular to students or schools. That does not mean school- and student-specific phenomena do not explain part of students' score trajectories, but it does mean any explanation of what is happening to students has to grapple with the powerful, and powerfully similar, effects occurring in the US population as a whole.

Pressure Testing Theories of Why

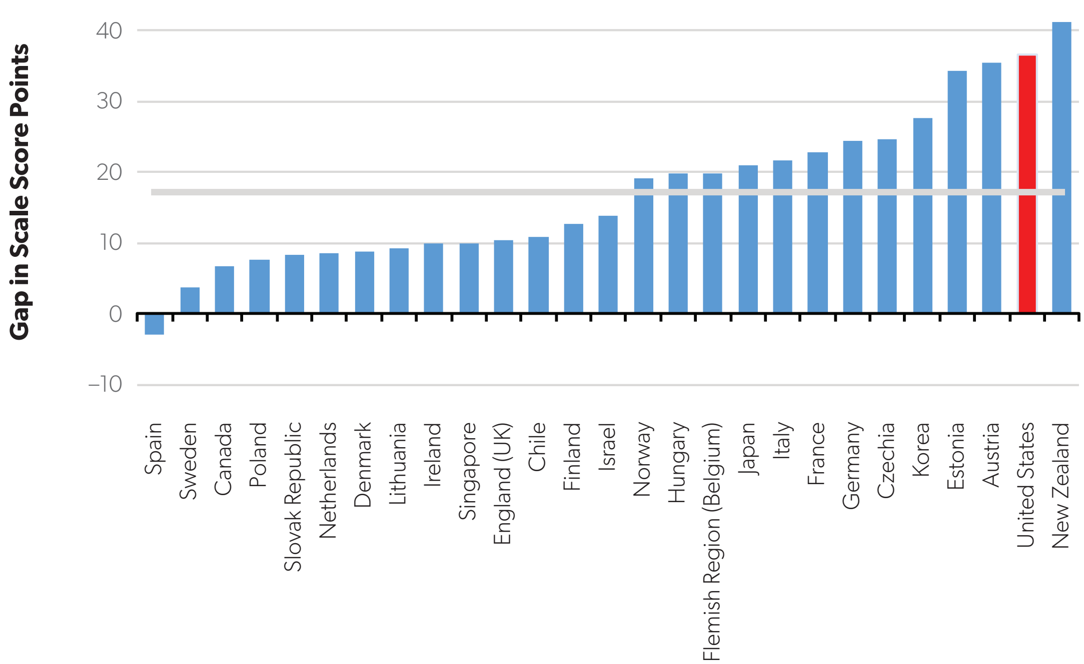
The assessment trends in this report beg for some explanation, for their size if nothing else. Declines in average scores since the early 2010s are on the order of 15 percent of a standard deviation.[7] For nationwide changes, those are very large declines and will be difficult to turn around. However, these changes were even larger for the lower end of the achievement spectrum. Indeed, while scores for the 90th and 75th percentiles declined by 0 and 6 percent of a standard deviation, respectively, scores for the 25th and 10th percentiles declined by 24 and 31 percent of a standard deviation, respectively. These declines, which

Figure 9. PIAAC 90th and 10th Percentile Gap Growth Across Countries: 2012–23

Panel A. Numeracy Score Gap Growth



Panel B. Literacy Score Gap Growth



Source: PIAAC, 2012–23. See Appendix A for details.

are more pronounced in the US and are evident for students and adults, are devastating for low achievers.

No single explanation is likely sufficient to account for the trends documented in this report, and reducing any explanation to a single cause would be misguided. Instead, plausible explanations must accommodate four key attributes: a downward trend beginning around 2013; declines driven by the bottom half of the distribution, both pre- and post-pandemic; higher absolute achievement gap growth in the US than in any other nation; and the similarity of these declines for US students and adults. In closing, I consider several potential explanations—not as definitive answers but as exercises in evaluating plausible "whys" behind these troubling trends.

National Crises: The Great Recession and the Pandemic

The Great Recession and the COVID-19 pandemic could both plausibly explain some of these changes. The Great Recession could have disproportionately affected the test scores of students and adults at the lower end of the achievement distribution. However, the connection between test scores and the Great Recession is not immediately obvious. Moreover, the timeline of the recession does not align perfectly with the onset of student score declines, and the even later onset of PIAAC declines among adults further complicates the link. Additionally, the recession was neither uniquely American nor primarily so, making it a less compelling explanation for the US-specific nature of these trends.

In theory, there should be a strong link between the pandemic and students' scores. Indeed, there is no doubt the pandemic had large effects on student schooling and achievement. The pandemic also had a larger effect on scores in the bottom half of the achievement distribution, but in general, students at both the high and low end lost ground.

However, the pandemic explanation has a few major problems. First, student declines driven by the bottom of the distribution predate the pandemic. Second, the pandemic's effects were not peculiarly American, with some international assessments—particularly the PISA—suggesting US fallout was smaller than fallout in other countries. Third, although the pandemic could explain changes to adult assessment scores, particularly given the apparent later onset of PIAAC declines, it's not clear why the effect would hit only adults on the lower half of the achievement distribution. These weaknesses in a pandemic-centric explanation don't mean the pandemic didn't play a large role, but they do suggest there is much more to the story.

Education Policy Changes: The End of No Child Left Behind and the Rise of the Common Core

No Child Left Behind, with its focus on accountability, may have modestly improved student achievement, especially among lower-performing students. Accordingly, one could argue that the Every Student Succeeds Act, with its weakened accountability structures, might have disproportionately affected the bottom half of the distribution. The timelines for these policy changes align reasonably well with the start of student score declines, and No Child Left Behind was uniquely American, making this explanation seem plausible. However, this explanation offers no insight into the concurrent drop in adult scores reflected in PIAAC data.

Similarly, the Common Core also aligns with the timeline of score declines and is another distinctly American phenomenon. Some evidence suggests that the Common Core may have hurt lower-performing students.[8] However, these observed effects are not large enough to explain the substantial subsequent shifts in scores, and it is unclear why the Common Core would have resulted in the changes we see in adult scores.

Cultural Changes: The Rise of Screens and the Fall of Reading

A final cluster of explanations relates to the concomitant rise of screen use and the decline of other, more academic student activities. Screen use increased significantly after 2012, aligning with the onset of score declines.[9] The timeline of score changes fits the timeline asserted for the connections between screens and mental health declines in the US and other countries. Moreover, the evolution of screen use—with increasing device availability and increased and escalating algorithmic content delivery—could interfere with student learning and adult skills, intensifying over time.[10] Additionally, given differences in screen time by household income, screen use could have had more of an impact on the lower part of the achievement distribution, and these effects likely became more pronounced during the pandemic when forced isolation was widespread.[11] The screen explanation, however, can't account for why these trends are distinctly American.[12]

Another possible explanation is the decline of academic pursuits. For example, reading for pleasure has fallen significantly over the past 15 years, likely displaced by the rise of screens. In 2012, 53 percent of 13-year-olds read for fun on their own time once or more weekly, but by 2020 and 2023, only 39 percent of 13-year-olds did. This decline in reading could have affected student performance and would likely have affected low-performing students more.[13] Furthermore, the decline in reading is evident for adults too. However, the size of these declines is not comparable to the size of PIAAC declines, and they were not experienced disproportionately by less-educated adults.[14] Moreover, the timelines for declines in adult reading time and PIAAC scores do not align.

Relatedly, LTT surveys show a substantial fall in the percentage of 13-year-olds taking algebra and pre-algebra after 2012, reflecting a decrease in more advanced mathematics coursework, with a substantially larger effect among low-performing students.[15] These patterns fit the timeline and hit the bottom half in ways that reflect test score declines, but they may not be sufficient to explain large portions of recent assessment trends, particularly in reading scores and for adults more generally.

Conclusion

These test score trends are important, and the explanations that aim to account for them are equally important, if harder to pin down. It is not my intention to propose a singular theory for these declines. Instead, I argue that any explanation of these trends must account for four critical factors: a downward trend beginning around 2013; declines driven by the bottom half of the distribution, both pre- and post-pandemic; higher absolute achievement gap growth in the US than in any other nation; and the similarity of these declines for US students and adults.

The potential explanations above are all viable, but not satisfying, theories, but such theories must be examined critically because they will inform how schools, families, and communities respond to score declines. Accordingly, until these patterns no longer hold, policymakers, educators, and communities should test their theories for test score declines—those seen here and in future waves of test results—against these four critical factors.

About the Author

Nat Malkus is a senior fellow, the deputy director of education policy at the American Enterprise Institute (AEI), and an affiliate of AEI's James Q. Wilson Program in K–12 Education Studies, where he specializes in empirical research on K–12 schooling.

Appendix A. Data

This report draws on nationally representative US test results from several assessments, each covering a variety of grades, ages, subjects, and time frames. It provides an overview of the assessment programs included, specifying the students' ages or grades, subjects (with abbreviations where applicable), and sources for the original data. The included assessments were selected based on their availability for a nationally representative population, recurrence over time, and inclusion of at least one set of results after 2015. All data were drawn from assessment programs administered or participated in by the National Center for Education Statistics (NCES), part of the Institute of Education Sciences in the US Department of Education. Year ranges below refer to the full span of results used in this report, though assessments were administered only in select years. Results from before 1994 are excluded from this report, even when available.

The National Assessment of Educational Progress (NAEP), often referred to as the "main NAEP" to distinguish it from the NAEP long-term trend (LTT) assessments, provides results for fourth-grade and eighth-grade students in reading (1992–2022), mathematics (1990–2022), and science (2009–19). NAEP US history results are available for eighth-grade students from 1994 to 2022. For 12th-grade students, NAEP results are available in reading (1992–2019), mathematics (2005–19), and science (2009–19).

The NAEP LTT assessments are distinct from the main NAEP and provide results that track changes in the performance of nine- and 13-year-old students in reading and mathematics over time. LTT results for 17-year-old students were available through 2012 but are not included in this report. These assessments, which began in the 1970s, focus on consistent frameworks and questions to allow for reliable comparisons across decades. The results used in this report include reading and mathematics for nine-year-olds from 1994 to 2022 and for 13-year-olds from 1993 to 2023.

The Program for International Student Assessment (PISA), administered by the Organisation for Economic Co-operation and Development (OECD), measures the performance of 15-year-old students across participating countries, including the United States, in reading, mathematics, and science. Conducted every three years beginning in 2000, PISA evaluates students' ability to apply knowledge and skills to real-world problems. This report includes PISA results for American 15-year-olds in reading from 2000 to 2022 and in mathematics and science from 2003 to 2022, offering an international comparative perspective on educational outcomes.

The Trends in International Mathematics and Science Study (TIMSS), administered by the International Association for the Evaluation of Educational Achievement (IEA), assesses mathematics and science achievement for students in grades four and eight across participating countries, including the United States. Conducted every four years since 1995, TIMSS evaluates students' knowledge and skills based on curricula and frameworks common to participating education systems. This report includes TIMSS results for American fourth- and eighth-grade students in mathematics and science from 1995 to 2023, providing consistent performance measures and international benchmarks for comparison.

The Program for the International Assessment of Adult Competencies (PIAAC), administered by the OECD, evaluates literacy and numeracy skills of adults age 16 to 65 across participating countries, including the United States. PIAAC assessments began in 2012 and have been conducted in multiple cycles through 2023. This report includes PIAAC results for US adults in literacy and numeracy from 2012 to 2023, along with linked scores in literacy and numeracy from the Adult Literacy and Lifeskills (ALL) Survey conducted between 2003 and 2008 and in numeracy from the International Adult Literacy Survey (IALS) conducted between 1994 and 1998. These linked assessments provide a longer-term perspective on adult skill levels and trends over time.

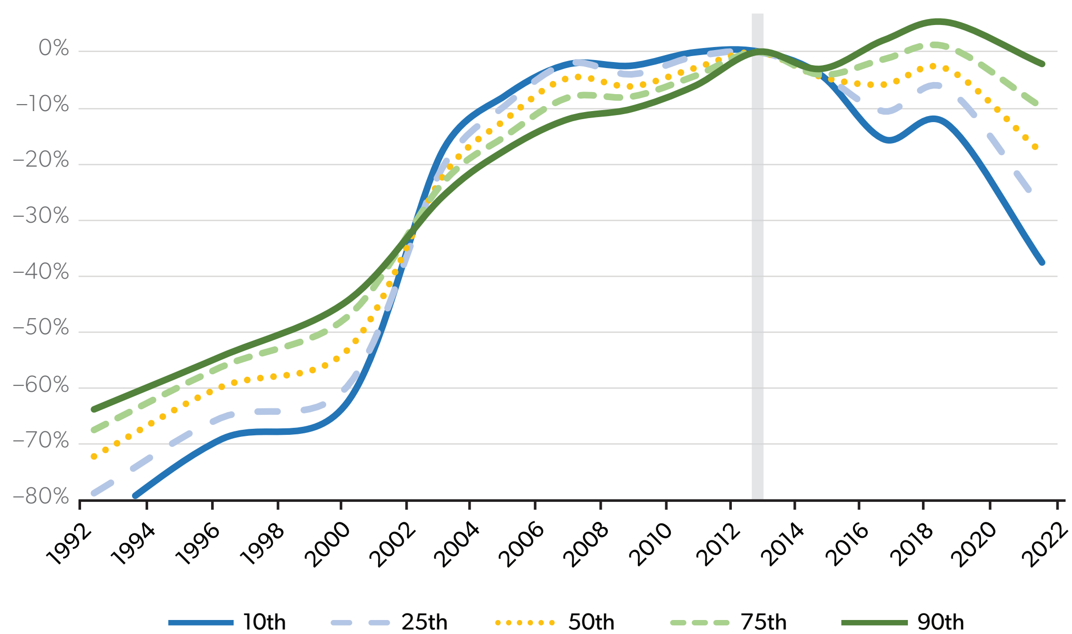
The 2023 PIAAC results should be interpreted with caution for several reasons. First, the response rate was 28 percent, significantly below Organisation for Economic Co-operation and Development standards. Second, the US PIAAC assessment introduced digitally administered assessments in 2023, a mode change that may introduce unknown biases. Finally, NCES changed the PIAAC 2023 assessment design to more accurately capture the performance of the lowest-performing participants. As discussed in this report, while the results are compelling, the decline at the lower percentiles may be overstated.

All results presented in this report are publicly available through the NAEP Data Explorers and the International Data Explorer, all produced by NCES.[16]

Appendix B. Percentile Scores on Student Assessments in This Report

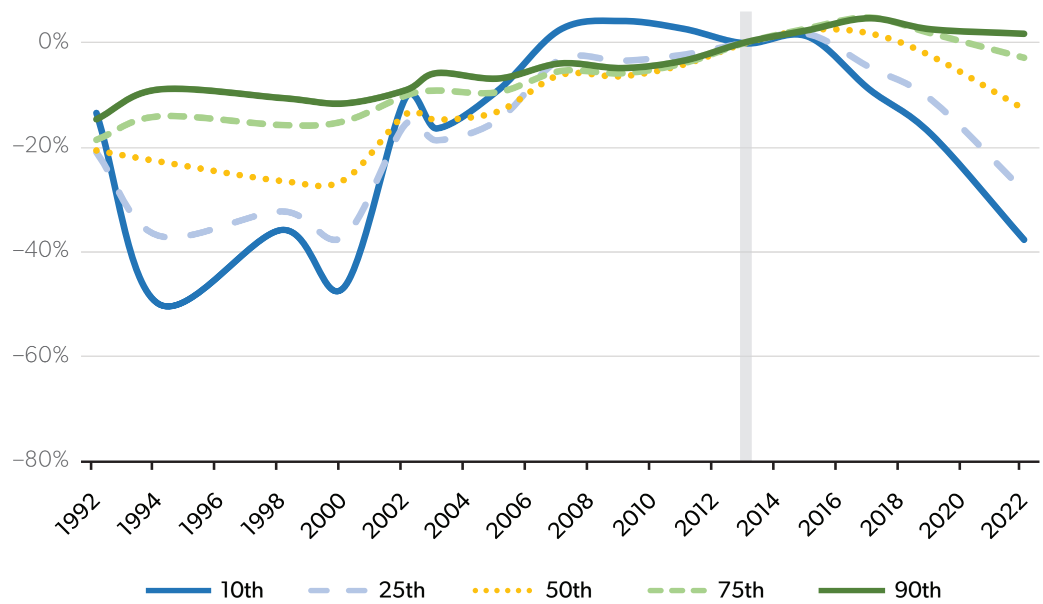
Appendix Figures B1–B21 display percentile scores for a given assessment relative to its high score in the 2012–15 period. All figures show scores in standard deviation units that are comparable across figures. The scale of each figure's vertical axis is not consistent. All apparent differences are not necessarily statistically significant.

Figure B1. NAEP Fourth-Grade Math Percentile Score Deviation from 2013



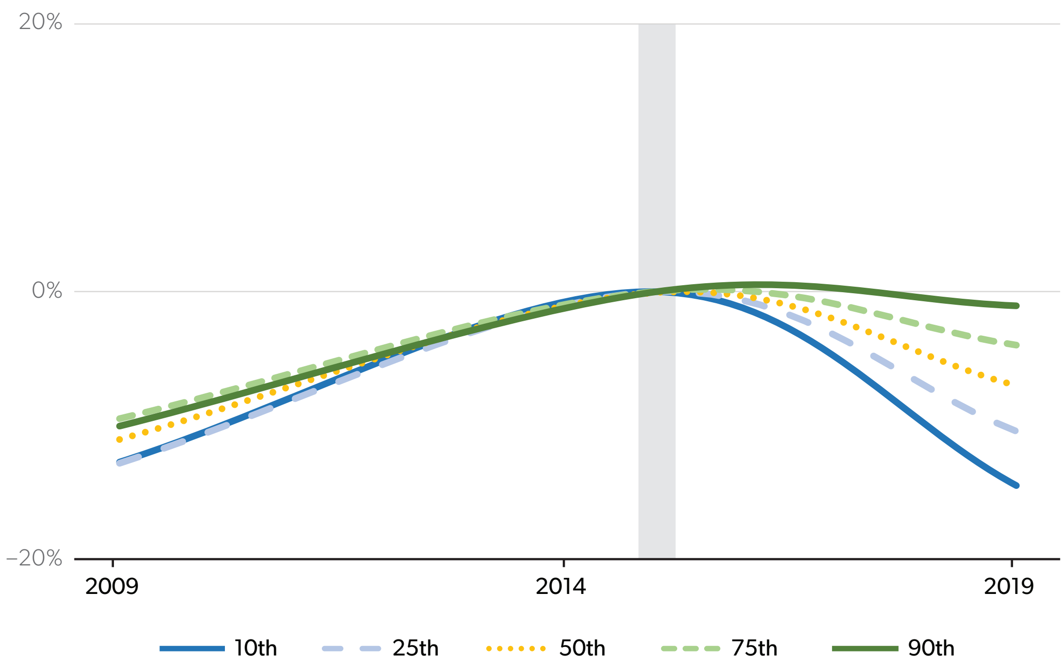
Source: NAEP, Fourth-Grade Math, 1992–2022. See Appendix A for details.

Figure B2. NAEP Fourth-Grade Reading Percentile Score Deviation from 2013



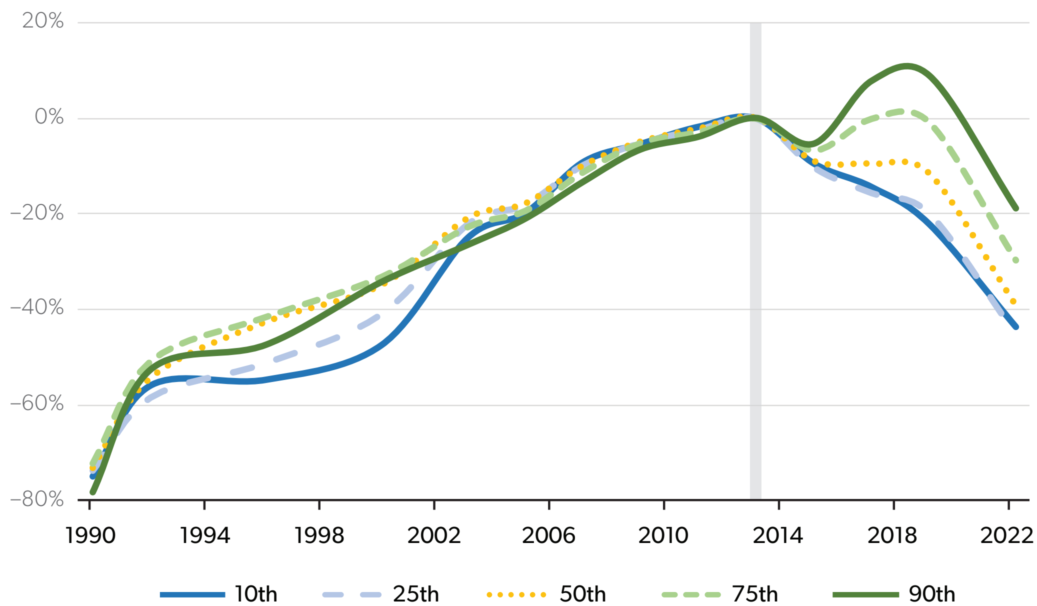
Source: NAEP, Fourth-Grade Reading, 1992–2022. See Appendix A for details.

Figure B3. NAEP Fourth-Grade Science Percentile Score Deviation from 2015



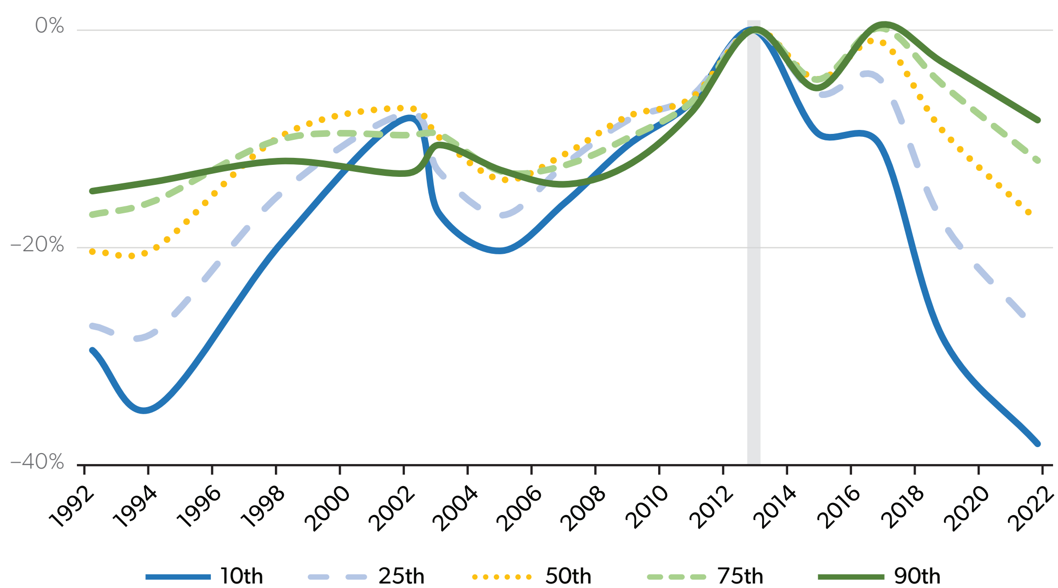
Source: NAEP, Fourth-Grade Science, 2009–19. See Appendix A for details.

Figure B4. NAEP Eighth-Grade Math Percentile Score Deviation from 2013



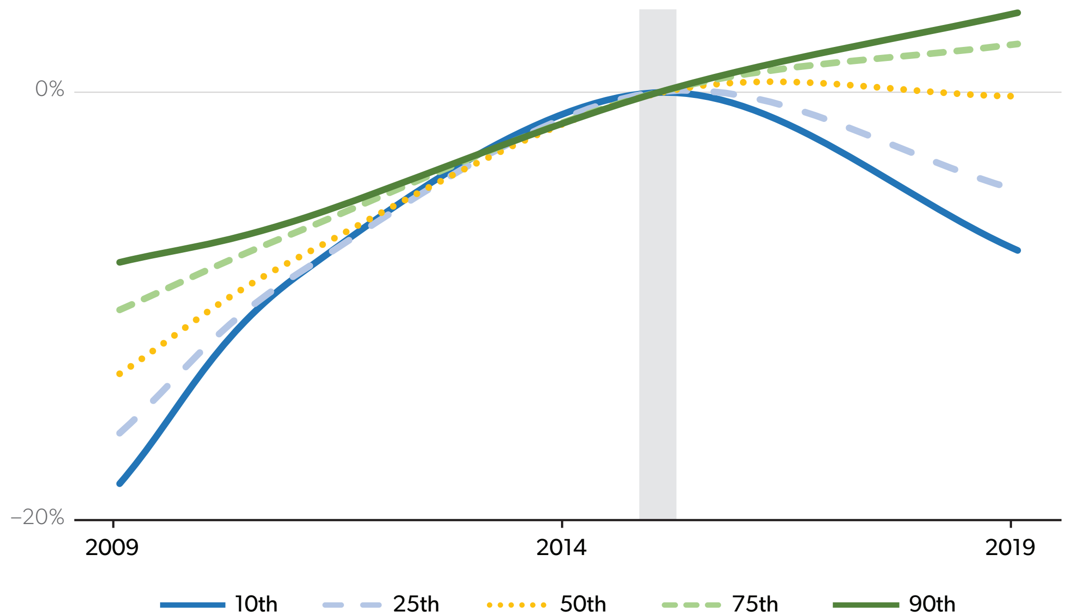
Source: NAEP, Eighth-Grade Math, 1990–2022. See Appendix A for details.

Figure B5. NAEP Eighth-Grade Reading Percentile Score Deviation from 2013



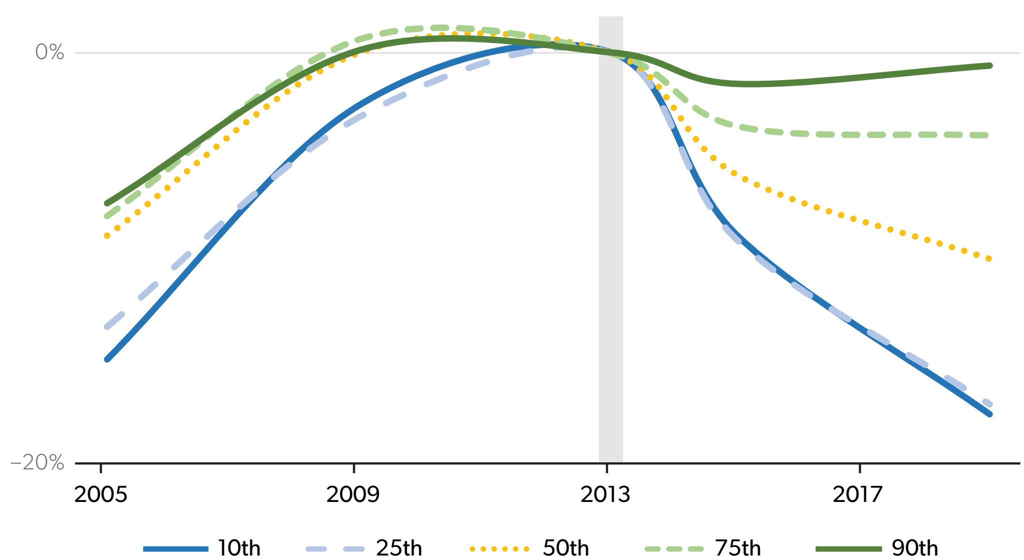
Source: NAEP, Eighth-Grade Reading, 2009–19. See Appendix A for details.

Figure B6. NAEP Eighth-Grade Science Percentile Score Deviation from 2015



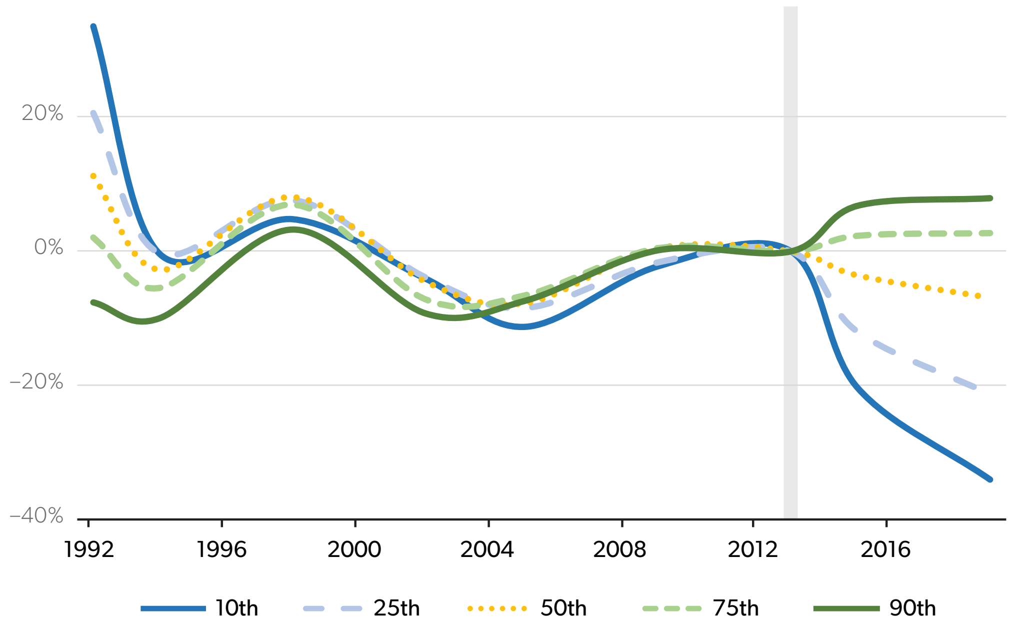
Source: NAEP, Eighth-Grade Science, 2009–19. See Appendix A for details.

Figure B7. NAEP 12th-Grade Math Percentile Score Deviation from 2013



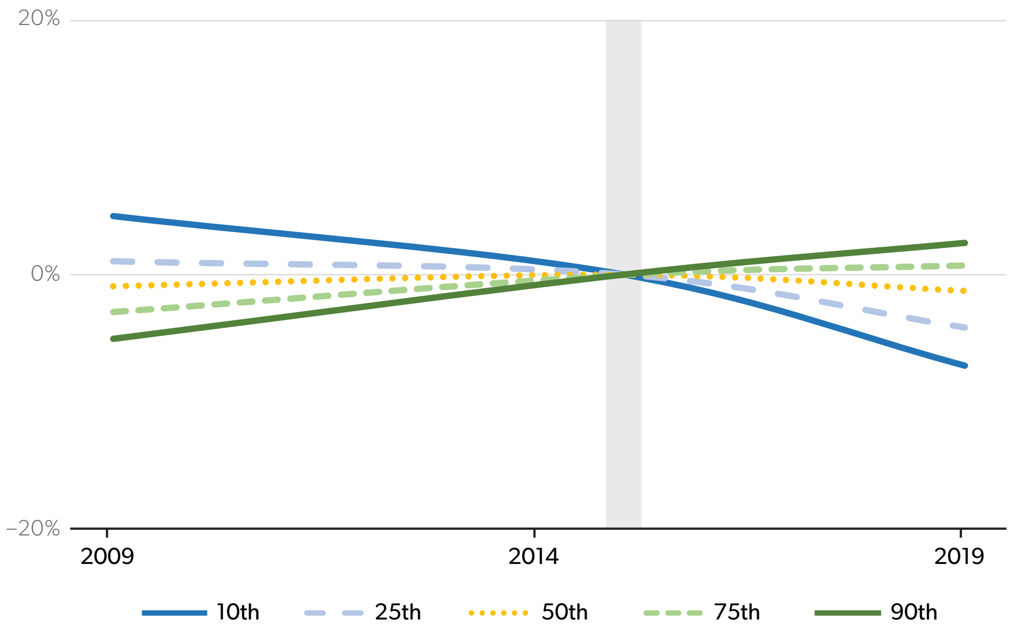
Source: NAEP, 12th-Grade Math, 2005–19. See Appendix A for details.

Figure B8. NAEP 12th-Grade Reading Percentile Score Deviation from 2013



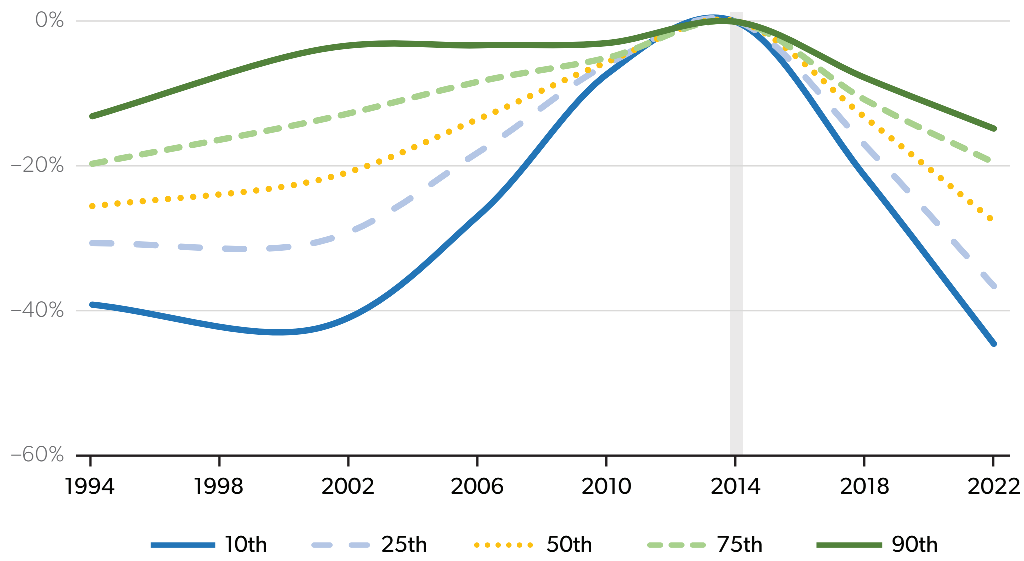
Source: NAEP, 12th-Grade Reading, 1992–2019. See Appendix A for details.

Figure B9. NAEP 12th-Grade Science Percentile Score Deviation from 2015



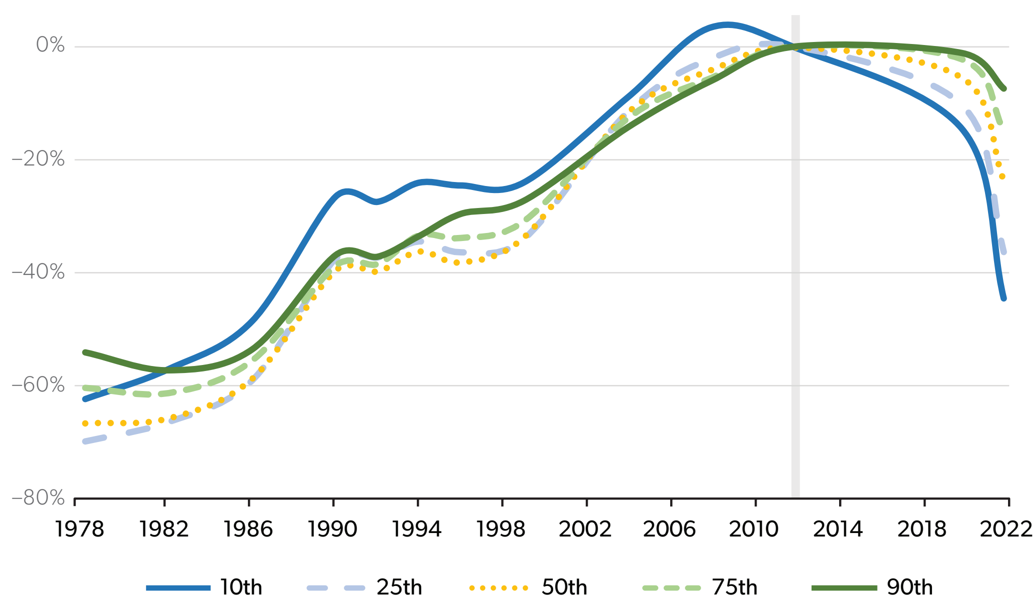
Source: NAEP, 12th-Grade Science, 2009–19. See Appendix A for details.

Figure B10. NAEP Eighth-Grade US History Percentile Score Deviation from 2014



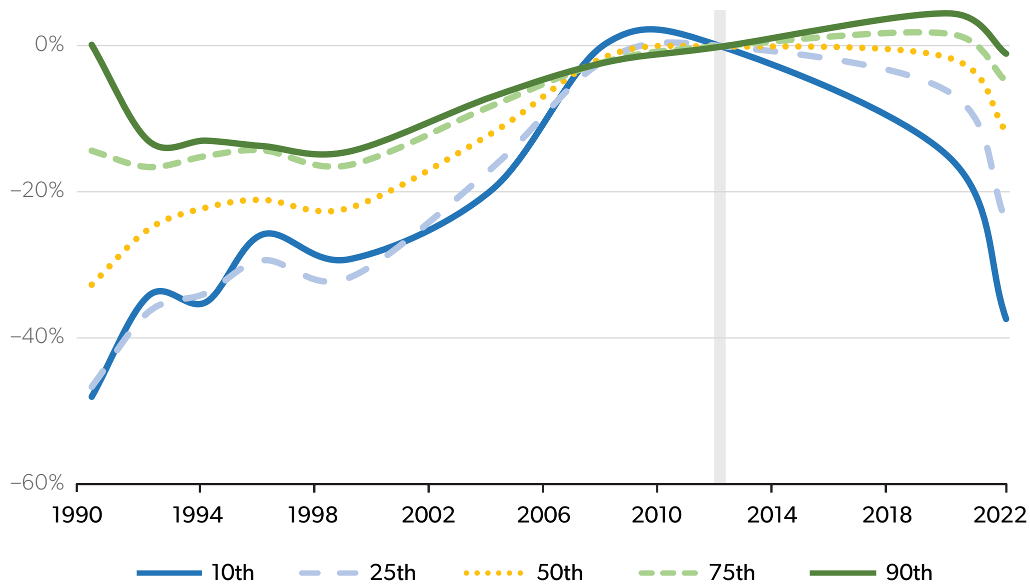
Source: NAEP, Eighth-Grade US History, 1994–2022. See Appendix A for details.

Figure B11. NAEP LTT Age Nine Math Percentile Score Deviation from 2012



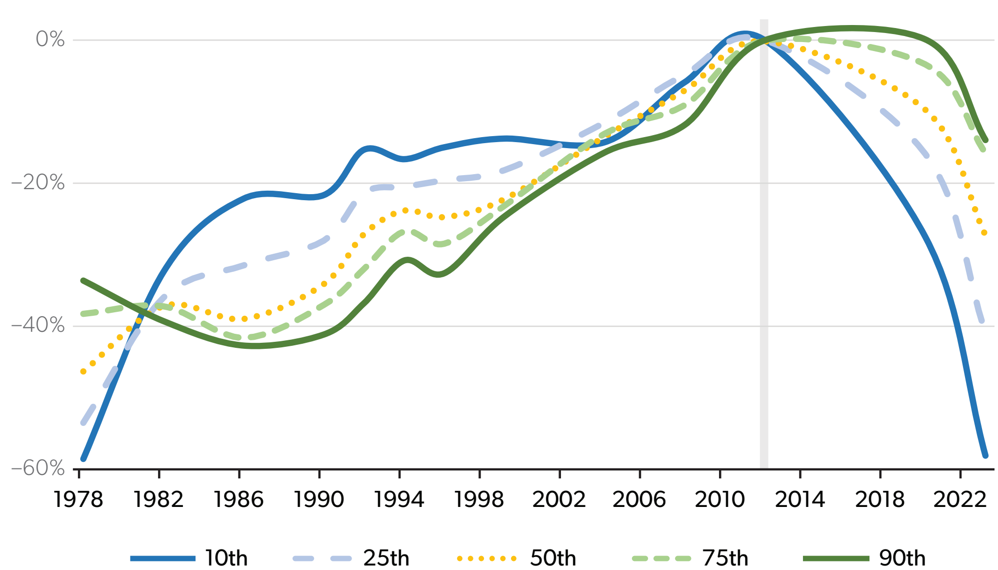
Source: NAEP LTT, Age Nine Math, 1978–2022. See Appendix A for details.

Figure B12. NAEP LTT Age Nine Reading Percentile Score Deviation from 2012



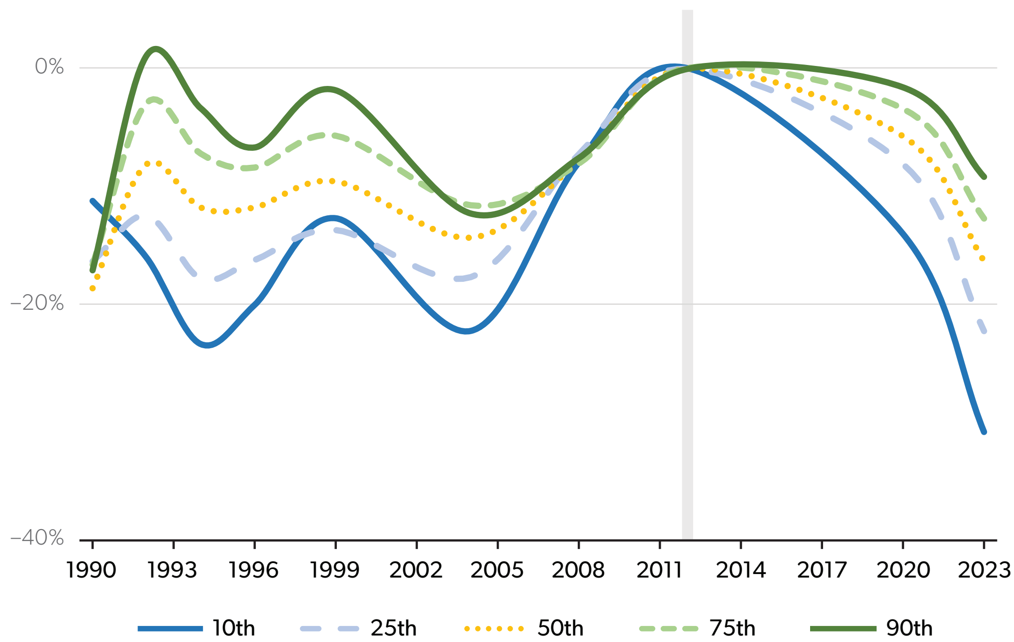
Source: NAEP LTT, Age Nine Reading, 1990–2022. See Appendix A for details.

Figure B13. NAEP LTT Age 13 Math Percentile Score Deviation from 2012



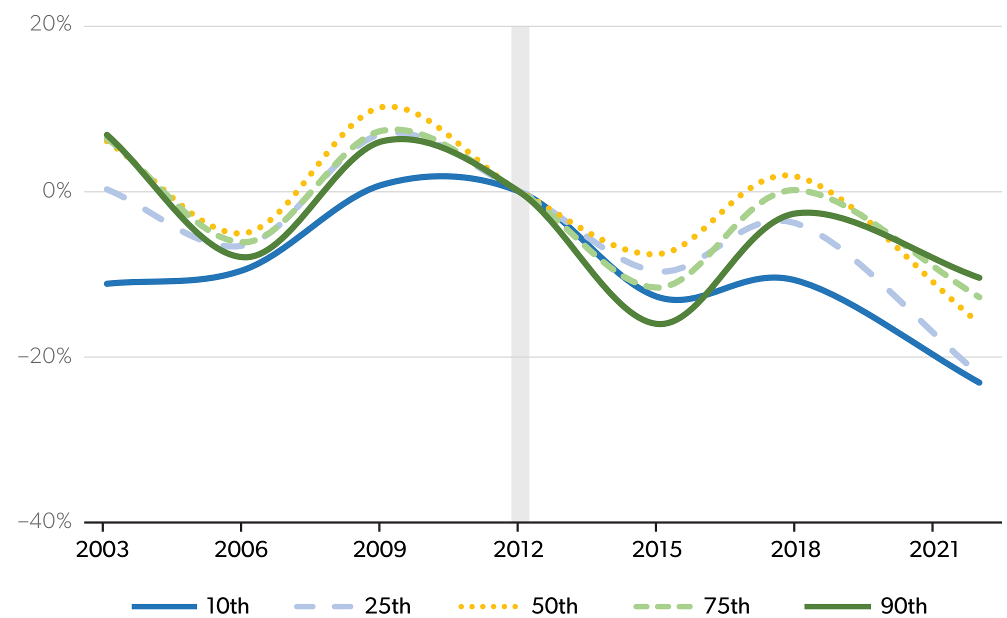
Source: NAEP LTT, Age 13 Math, 1978–2023. See Appendix A for details.

Figure B14. NAEP LTT Age 13 Reading Percentile Score Deviation from 2012



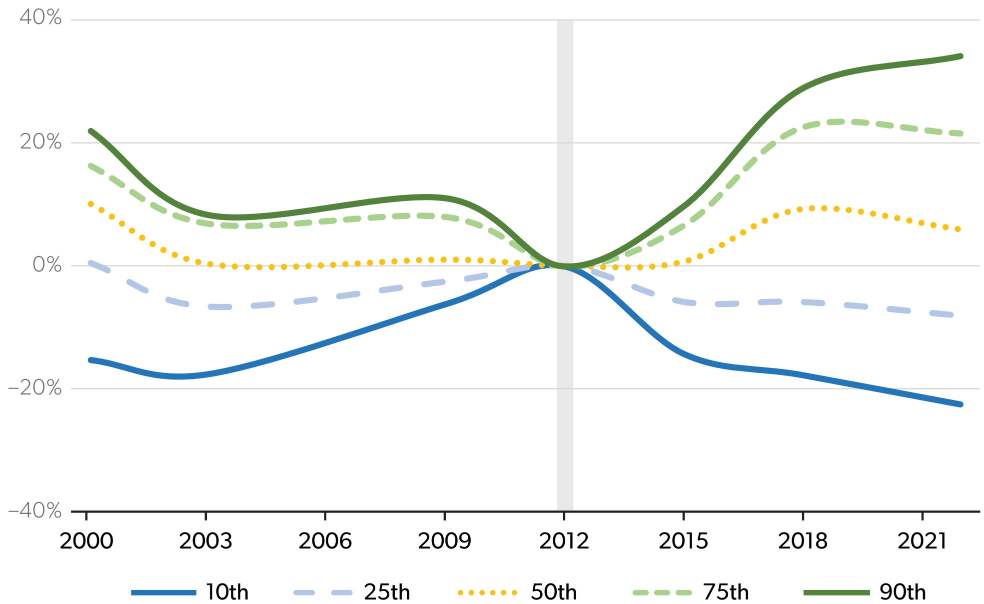
Source: NAEP LTT, Age 13 Reading, 1990–2023. See Appendix A for details.

Figure B15. PISA Age 15 Math Percentile Score Deviation from 2012



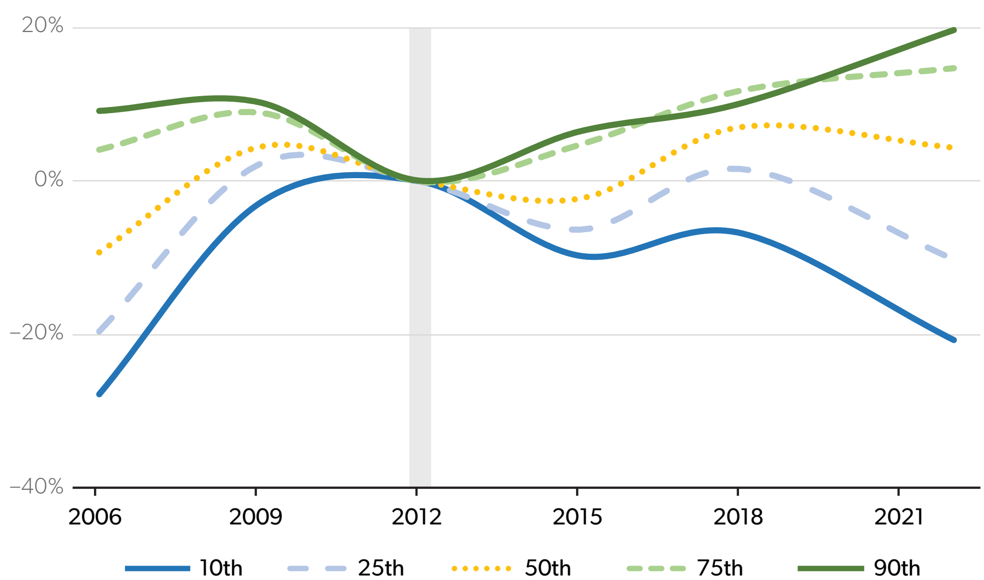
Source: PISA, Age 15 Math, 2003–22. See Appendix A for details.

Figure B16. PISA Age 15 Reading Percentile Score Deviation from 2012



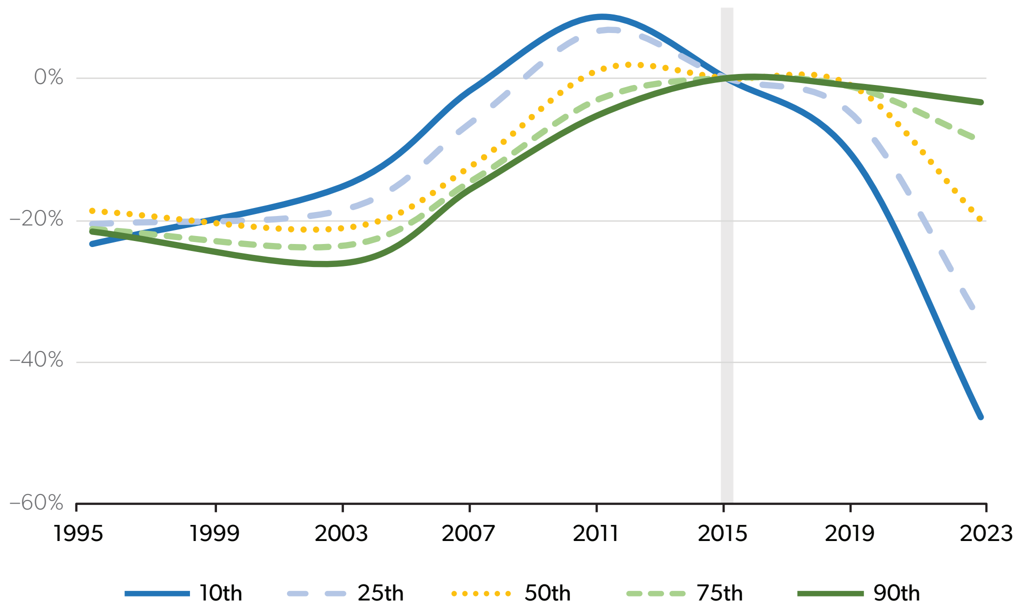
Source: PISA, Age 15 Reading, 2000–22. See Appendix A for details.

Figure B17. PISA Age 15 Science Percentile Score Deviation from 2012



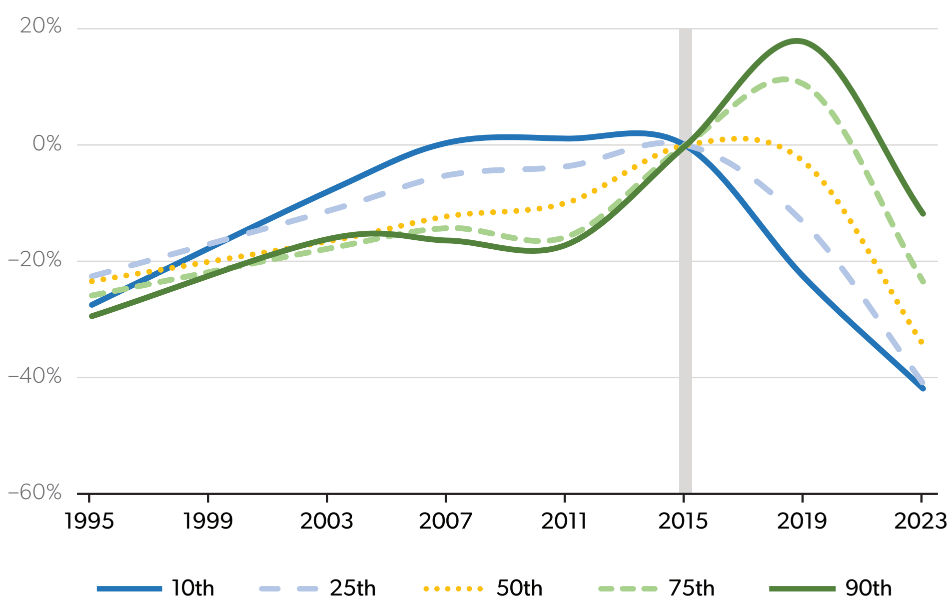
Source: PISA, Age 15 Science, 2006–22. See Appendix A for details.

Figure B18. TIMSS Fourth-Grade Math Percentile Score Deviation from 2015



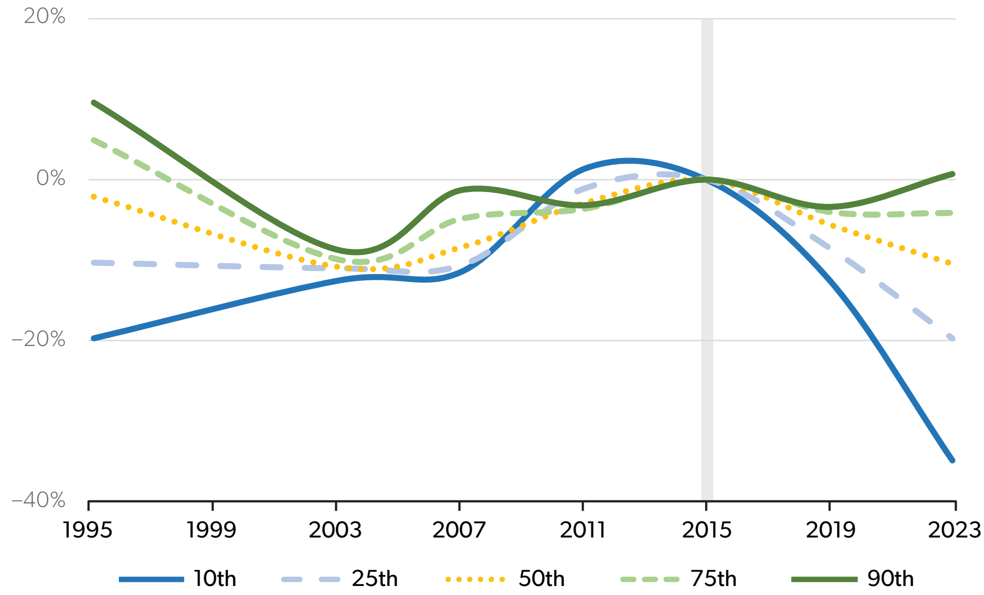
Source: TIMSS, Fourth-Grade Math, 1995–2023. See Appendix A for details.

Figure B19. TIMSS Eighth-Grade Math Percentile Score Deviation from 2015



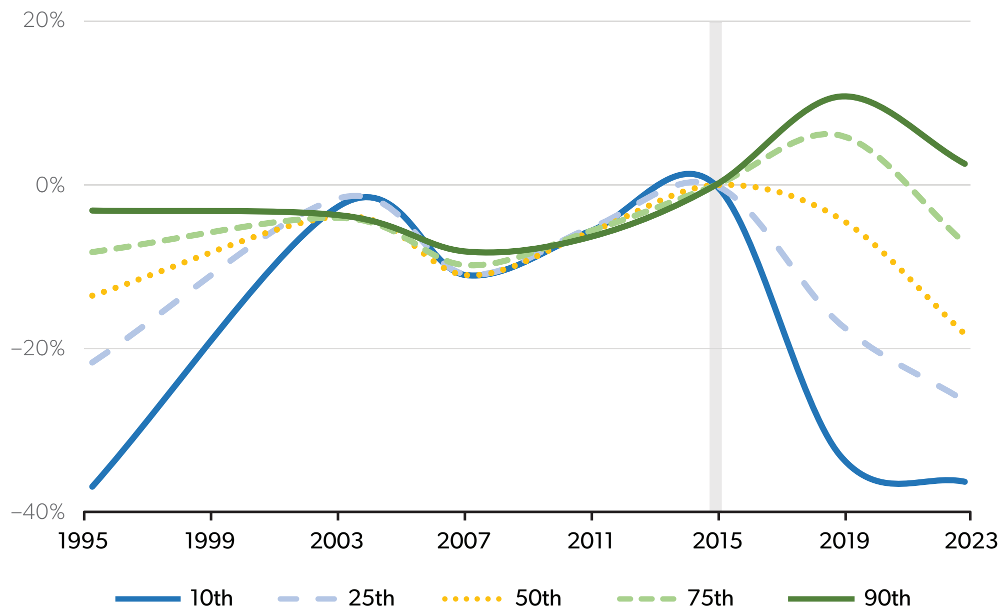
Source: TIMSS, Eighth-Grade Math, 1995–2023. See Appendix A for details.

Figure B20. TIMSS Fourth-Grade Science Percentile Score Deviation from 2015



Source: TIMSS, Fourth-Grade Science, 1995–2023. See Appendix A for details.

Figure B21. TIMSS Eighth-Grade Science Percentile Score Deviation from 2015



Source: TIMSS, Eighth-Grade Science, 1995–2023. See Appendix A for details.

Notes

1. In general, each score is subtracted from that assessment's high score between 2012 and 2015. However, for some assessments groups (fourth-grade NAEP and all TIMSS assessments), a common high score year was used even though an assessment in each group had another slightly higher point between 2012 and 2015. The differences in these patterns are minor.

2. Chad Aldeman, "Interactive: See How Student Achievement Gaps Are Growing in Your State," *The 74*, February 26, 2024, https://www.the74million.org/article/interactive-see-how-student-achievement-gaps-are-growing-in-your-state/.

3. This is an unweighted average of the gaps across countries and, though a rough estimation, usefully illustrates the US as an outlier. Nat Malkus, "A Decade of PISA Scores Shows Alarming Achievement Gap Growth," *AEIdeas*, December 28, 2023, https://www.aei.org/education/a-decade-of-pisa-scores-shows-alarming-achievement-gap-growth-2/.

4. Further, since the US test score apogee was in 2011 on TIMSS, but not in all other countries, I compared the maximum gap growth among all other countries between any test administrations from 2011, 2015, 2019, and 2023, and the US still had the largest gap growth compared with all other countries for any time gap in that period.

5. England should also be added to this list of countries with higher achievement gap growth than the US in eighth-grade science, if one counts any possible margin between 2011 and 2023. In the case of England, the estimated achievement gap growth between 2015 and 2023 was larger than the growth gap in the US between 2011 and 2023.

6. The PIAAC is a large-scale international survey conducted by the OECD that evaluates and compares the skills and competencies of adults (ages 16 to 65) across countries, including the US. PIAAC assessments differ from students' academic-oriented assessments, but they deal with adult skills in terms of literacy and numeracy that are somewhat analogous to student math and reading tests. PIAAC results are available for many nations from 2012 and 2023, and the US had a 2017 round of PIAAC assessments. I extend these PIAAC scores with equated scores from two additional adult assessments: for literacy and numeracy from the ALL in 2003–08 and for literacy from the IALS in 1994–98.

7. These are simple averages across nationally representative test scores. They are not weighted but are only used to communicate average declines.

8. Joshua Bleiberg, "Does the Common Core Have a Common Effect? An Exploration of Effects on Academically Vulnerable Students," *AERA Open* 7, no. 1 (2021): 1–18, https://files.eric.ed.gov/fulltext/EJ1323929.pdf; and Benjamin W. Arold and M. Danish Shakeel, "The Unintended Effects of the Common Core State Standards on Non-Targeted Subjects," Working Paper No. 21-418 (Annenberg Institute at Brown University, June 2021), https://edworkingpapers.com/ai21-418.

9. Jean M. Twenge, "Increases in Depression, Self-Harm, and Suicide Among U.S. Adolescents After 2012 and Links to Technology Use: Possible Mechanisms," *Psychiatric Research and Clinical Practice* 2, no. 1 (2020): 19–25, https://pubmed.ncbi.nlm.nih.gov/36101887/; Pedro Antonio Sánchez-Miguel et al., "School and Non-School Day Screen Time Profiles and Their Differences in Health and Educational Indicators in Adolescents," *Scandinavian Journal of Medicine & Science Sports* 32, no. 11 (2022): 1668–81, https://pubmed.ncbi.nlm.nih.gov/35856173/; and Amanda Lenhart, *Teens, Social Media & Technology Overview 2015*, Pew Research Center, April 9, 2015, https://www.pewresearch.org/internet/2015/04/09/teens-social-media-technology-2015/.

10. Sheri Madigan et al., "Assessment of Changes in Child and Adolescent Screen Time During the COVID-19 Pandemic: A Systematic Review and Meta-Analysis," *JAMA Pediatrics* 176, no. 12 (2022): 1188–98, https://jamanetwork.com/journals/jamapediatrics/fullarticle/2798256; Mireia Adelantado-Renau et al., "Association Between Screen Media Use and Academic Performance Among Children and Adolescents: A Systematic Review and Meta-Analysis," *JAMA Pediatrics* 173, no. 11 (2019): 1058–67, https://pmc.ncbi.nlm.nih.gov/articles/PMC6764013/; Sudheer Kumar Muppalla, "Effects of Excessive Screen Time on Child Development: An Updated Review and Strategies for Management," *Cureus* 15, no. 6 (2023), https://pmc.ncbi.nlm.nih.gov/articles/PMC10353947/; and Victoria Rideout et al., *The Common Sense Census: Media Use by Tweens and Teens, 2021*, Common Sense, March 9, 2022, https://www.commonsensemedia.org/research/the-common-sense-census-media-use-by-tweens-and-teens-2021.

11. Monica Anderson et al., *Teens, Social Media and Technology 2023*, Pew Research Center, December 11, 2023, https://www.pewresearch.org/internet/2023/12/11/teens-social-media-and-technology-2023/.

12. Madigan et al., "Assessment of Changes in Child and Adolescent Screen Time During the COVID-19 Pandemic."

13. Reading Zone, "Survey Reveals Marked Decline in Reading for Pleasure in 2023," press release, September 5, 2023, https://www.readingzone.com/news/survey-reveals-marked-decline-in-reading-for-pleasure-in-2023/; and Nation's Report Card, "Scores Decline Again for 13-Year-Old Students in Reading and Mathematics," US Department of Education, Institute of Education Sciences, National Center for Education Statistics, https://www.nationsreportcard.gov/highlights/ltt/2023/.

14. National Endowment for the Arts, *Arts Participation Patterns in 2022: Highlights from the Survey of Public Participation in the Arts*, October 2023, https://www.arts.gov/impact/research/publications/arts-participation-patterns-2022-highlights-survey-public-participation-arts; Katharina Buchholz, "This Chart Shows How Reading for Pleasure Is Declining in the US," World Economic Forum, April 28, 2022, https://www.weforum.org/stories/2022/04/reading-pleasure-america-covid19/; and American Academy of Arts & Sciences, "Time Spent Reading," https://www.amacad.org/humanities-indicators/public-life/time-spent-reading.

15. Nation's Report Card, "Scores Decline Again for 13-Year-Old Students in Reading and Mathematics."

16. US Department of Education, Institute of Education Sciences, National Center for Education Statistics, Main Data Explorer, https://www.nationsreportcard.gov/ndecore/landing; US Department of Education, Institute of Education Sciences, National Center for Education Statistics, Data Explorer for the Long-Term Trend Assessments, https://www.nationsreportcard.gov/ndecore/xplore/ltt; and US Department of Education, Institute of Education Sciences, National Center for Education Statistics, International Data Explorer, https://nces.ed.gov/surveys/international/ide/.