Activate Research, Inc.
www.activateresearch.net

Trends in U.S. High School Students' Average Reading Scores on NCES Large-Scale Assessments: 1998–2022

Executive Summary

Introduction

This consolidated review of National Center for Education Statistics (NCES) high school student assessment data presents trends in average test scores across multiple NCES studies. This report extends our analyses of U.S. students' grade 4 and grade 8 average test scores on math, reading, and science assessments between 1998 and 2024. As reported in that study, for both grade 4 and grade 8 average math and reading scores, generally, we observe score increases from about 2000 through the first half of the 2010s. Scores largely hold steady or decrease in the second half of the 2010s, and most scores in the 2020s are either lower than or not measurably different from the scores from about 2000.

In this analysis, we report U.S. high school students' average test scores in reading[1] between 1998 and 2022. These findings document student performance on NCES-sponsored large-scale assessments[2] during the last quarter century, a period marked at the beginning by an increased focus on accountability ushered in by the passage of the No Child Left Behind Act (NCLB) in 2002 and at the end by the aftereffects of the COVID-19 pandemic.

---

[1] Our team intended to conduct the same analyses for high school students that we conducted for students in elementary and middle school (e.g., math, reading, and science). However, due to a lack of comparable math and science data for high school students, we were only able to present findings for reading scores.

[2] Assessments include the National Assessment of Educational Progress Long-Term Trend ([NAEP-LTT], 17-year-old students), NAEP (students in grade 12), and the Program for International Student Assessment (PISA) study (15-year-old-students).

The figure presented here uses publicly available data from national and international assessments. Our analyses build on work conducted by NCES and its contracting partners.

Key findings

1. We observe that U.S. high schoolers' average reading scores generally decreased in the first half of the 2000s. Thereafter, no consistent pattern emerges.
2. NCES high school math and science assessment data come from data collection years that are not comparable to each other. Therefore, we lack sufficient comparable data to analyze and report key findings for high school students in these subjects.

Implications for research and practice

Our analysis highlights the difficulty in comparing high school scores across large-scale assessments because current assessments rely on different data collection schedules and assess students of different ages. Over the last 24 years, the federal government has placed a smaller emphasis on gathering trend data on the performance of high school students than it has on gathering similar data on elementary and middle school students. Unfortunately, limited data points for high school students hinder accurate trend analysis, particularly for math and science and, to a lesser extent, reading. Recent plans to reduce the frequency of the National Assessment of Educational Progress Long-Term Trend (NAEP-LTT)[3] assessment suggest that federal efforts to evaluate high school student performance will become even more limited in future years.

Average U.S. Reading Scores

This section provides key findings relating to our analysis of the change in average reading scores for U.S. high school students on NAEP-LTT, NAEP, and the Program for International Student Assessment (PISA) between 1998 and 2022 (see figure 1 for additional details).

We observe that U.S. high schoolers' average NAEP-LLT and NAEP reading scores generally decreased in the first half of the 2000s. Thereafter, no consistent pattern emerges.

- Compared to 1999 (average scale score = 288), the average NAEP-LTT reading score for 17-year-old students was lower in 2004. Average reading scores in 2008 and 2012 were not measurably different than scores in 1999; the average reading scores in 2008 and 2012 were higher than the average reading score in 2004.

[3] The NAEP assessment calendar is available at https://nces.ed.gov/nationsreportcard/about/calendar.aspx.

- Compared to 1998 (average scale score = 290), the average NAEP reading score for grade 12 students was lower in all subsequent NAEP assessments administered between 2002 and 2019. In addition, the average score in 2019 was lower than the average reading scores in 2009, 2013, and 2015.
- Compared to 2000 (average scale score = 504), the average PISA reading score for 15-year-old students was not measurably different in any subsequent PISA assessment administered between 2003 and 2022.

Figure 1. Percentage change in average reading scale scores for National Assessment of Educational Progress Long-Term Trend (NAEP-LTT, age 17), NAEP (grade 12), and Program for International Student Assessment (PISA, age 15), relative to earliest presented year: 1998 through 2022

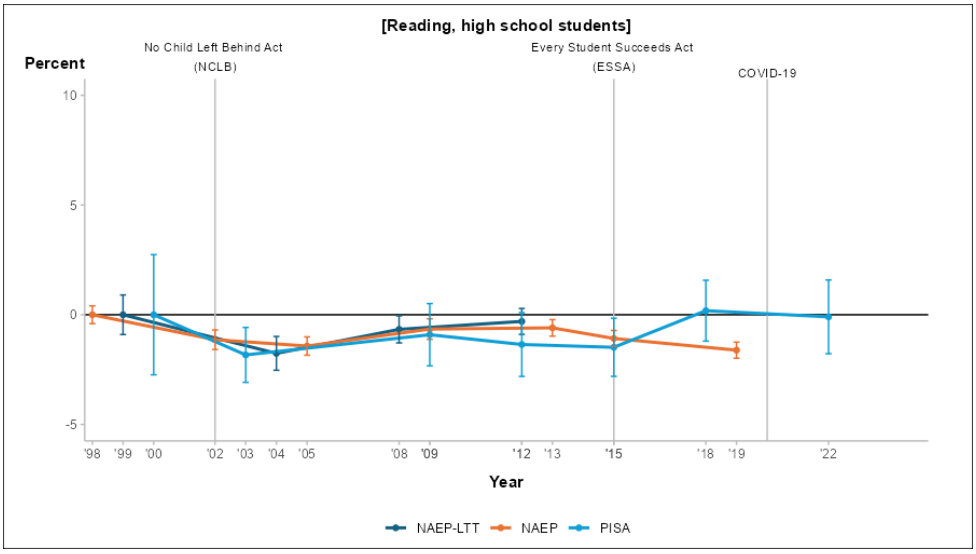

FIGURE READS: The NAEP 2002 average score was approximately 1 percent lower than the NAEP 1998 average score. The PISA 2022 average score was not measurably different from the PISA 2000 average score (as evidenced by the confidence interval stretching across the zero line).

NOTE: For each survey, each value is the percentage change between the respective scale score and the overall average scale score for the first data collection that is presented in the figure. The first year presented for NAEP-LTT is 1999; for NAEP, it is 1998; and for PISA, it is 2000. The figure displays confidence intervals (CIs). The upper bound of each interval is marked by "┬", while the lower bound of each interval is marked by "┴". The center of each interval is the estimated group average. The figure does not directly present information on statistical significance and some apparent differences may not be statistically significant. See the "Notes and Sources" section below for additional details.

Notes and Sources

Data sources

We produced our findings from publicly available NCES data explorers for the following assessments:

- NAEP-LTT reading assessments, 1999 through 2012
- NAEP reading assessments, 1998 through 2019
- PISA assessments, 2000 through 2022

The NCES NAEP Data Explorer (NDE) is available at https://www.nationsreportcard.gov/ndecore/landing.

The NCES International Data Explorer (IDE) is available at https://nces.ed.gov/surveys/international/ide/.

The data tables containing all estimates used in this report are available at https://www.activateresearch.net/our-work/trends-us-reading-scores-nces-assessments-report-home.

Measures

The scales for the NAEP-LTT and NAEP assessments range from 0 to 500. The scale for the PISA assessment ranges from 0 to 1000. The scales were developed independently for each assessment program and are designed for within-assessment-program comparisons over time. To compare across assessments, our team created a common metric to measure change in performance over time. We calculated percentage change values to allow us to plot assessments with different scales on a common axis. We converted each scale score to a percentage change by subtracting the average score for the first year of the assessment presented in the figure from the respective scale score value, dividing that difference by the average score for the first year of the assessment presented in the figure, and then multiplying that quotient by 100. This calculation yields the percent change relative to the study's first selected year's average score.

Variance estimation and statistical testing

The findings in this report are based on analyses of samples of students rather than entire populations of students. We used *t*-tests to determine statistical differences between pairs of average scale scores. The NDE and IDE each provide *t*-tests for comparisons of year-to-year estimates, which our team verified as part of our quality control procedures. To

account for sampling error, both the data explorers and our own calculations incorporated standard errors in the *t*-tests used for the bivariate comparisons. We assessed statistical significance using two-tailed tests with an alpha level of .05.

All differences that we describe in the text are statistically significant, unless we say the scores "were not measurably different." We use the term "not measurably different" to mean there is not enough information available in the collected data to determine which of two values is larger given the selected statistical test; "not measurably different" does not mean that the two values are equivalent.

We investigated trends over time by fitting ordinary least squares (OLS) regression models to the set of average scale scores for each assessment. We note that, ideally, we would have investigated trends using raw data, rather than population-level averages. The IDE and NDE allow for conducting many types of regression analyses, but they do not allow for regressing average scores over time. Unfortunately, recent staff terminations and program disruptions at NCES and its restricted-use data licensing program prevented us from accessing the data files for NAEP-LTT and NAEP that would allow for an examination of trends using regression analysis on the full samples. For the OLS regression models, we modeled the average scale scores as the dependent variable, with time (the number of years since the first assessment) as the independent variable.

The figure above includes 95 percent confidence intervals (CIs). A CI provides a measure of uncertainty around an estimate and can display a visual approximation of whether two estimates are statistically significantly different from each other. We calculated each interval as the estimated group average plus/minus 1.96 times the standard error of the average. In the figures, the upper bound of each confidence interval is designated with the "┬" symbol, while the lower bound of each confidence interval is designated with the "┴" symbol. Where the CIs overlap, statistical tests need to be performed to determine whether the apparent difference is statistically significant. In this report, estimates with a CI that includes zero are not measurably different from the corresponding assessment's earliest selected year in the figure.

Additional notes

Results for NAEP-LTT 1999 are from the original assessment format and results for 2004 through 2012 are from the revised assessment format. See https://nces.ed.gov/nationsreportcard/ltt/bridge_study.aspx for details. Results for NAEP 1998 through 2015 are from a paper-and-pencil-based assessment and results for 2019 are from a digitally based assessment. See https://nces.ed.gov/nationsreportcard/dba for details.

NAEP-LTT assesses students at age 17, with most of these students enrolled in grade 11. NAEP assesses students in grade 12; PISA assesses students at age 15, with most of these students enrolled in grade 10.

Related Publications

NAEP-LTT

U.S. Department of Education, National Center for Education Statistics (2013). *The Nation's Report Card: Trends in Academic Progress 2012* (NCES 2013-456). Washington, D.C. Retrieved May 19, 2025 from https://nces.ed.gov/nationsreportcard/subject/publications/main2012/pdf/2013456.pdf

NAEP

U.S. Department of Education, National Center for Education Statistics (2019). *2019 NAEP Mathematics and Reading Assessments: Highlighted Results at Grade 12 for the Nation* (NCES 2020-090). Washington, DC. Retrieved May 19, 2025 from https://nces.ed.gov/use-work/resource-library/report/statistical-analysis-report/2019-naep-mathematics-and-reading-assessments-highlighted-results-grade-12-nation.

PISA

Berger, M. (2023). *Highlights of PISA 2022 U.S. Results* (NCES 2023-115). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved May 19, 2025 from https://ies.ed.gov/use-work/resource-library/report/first-look-ed-tab/highlights-pisa-2022-u-s-results.

Authorship

Steven Bahr: Formal analysis (supporting); Validation (supporting). **Nancy Collins**: Formal analysis (lead); Methodology (equal). **Kimberly Curtis**: Writing – original draft and editing (equal). **Jane Hall**: Writing – original draft and editing (equal). **Robbie Kaplan**: Project coordination (supporting); Web publication (lead). **Mark Low**: Validation (lead); Methodology (equal). **Kathleen Mulvaney-Panjwani**: Validation (supporting). **Tim Oltman**: Data curation (lead); Visualization (lead). **Melissa Patton**: Formal analysis (supporting); Validation (supporting).

About Activate

Activate Research, Inc. is a woman-owned small business that provides expert social science consulting services and works with government agencies and private-sector clients to conduct rigorous research and data analysis. Its program evaluation, quantitative and qualitative data collection, and technical assistance services help clients make informed decisions and meet their strategic goals. Activate delivers high-quality solutions tailored to clients' unique needs by emphasizing technical accuracy, compliance with standards, and a commitment to impactful, actionable insights.