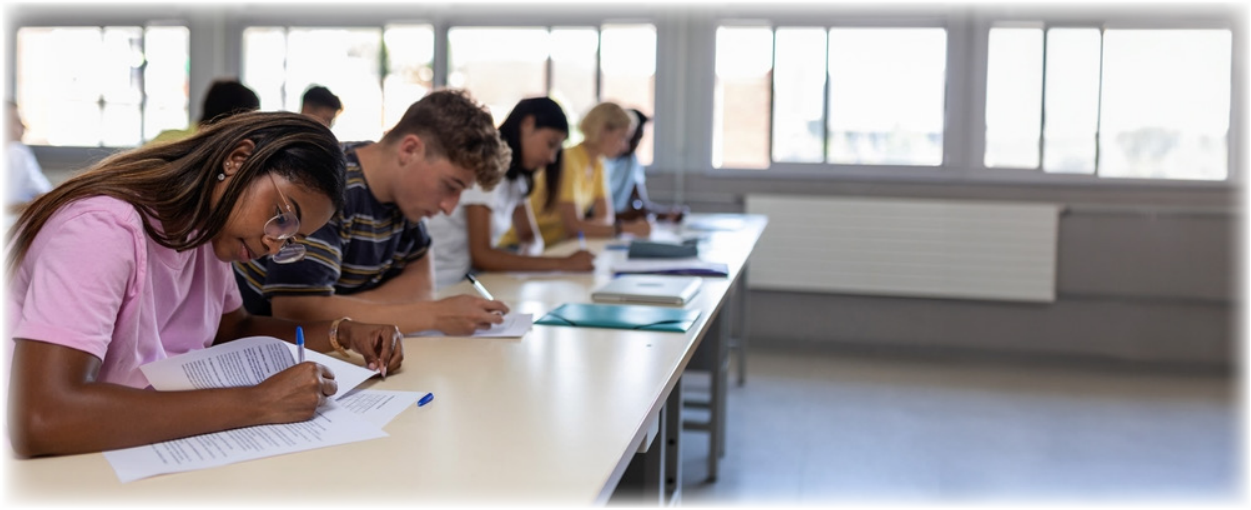
NEPC Review: Testing Theories of Why: Four Keys to Interpreting US Student Achievement Trends (American Enterprise Institute, January 2025)



Xavier Lorenzo/Shutterstock.com

Reviewed by:

Chris Domaleski

The National Center for the Improvement of Educational Assessment

April 2025

National Education Policy Center

School of Education
University of Colorado Boulder
nepc.colorado.edu

Acknowledgements

NEPC Staff

Faith Boninger
Publications Manager

Patricia Hinchey
Academic Editor

Elaine Duggan
Production Design

Alex Molnar
NEPC Director

Kevin Welner
NEPC Director

Suggested Citation: Domaleski, C. (2025). *NEPC Review: Testing theories of why: Four keys to interpreting US student achievement trends*. Boulder, CO: National Education Policy Center. Retrieved [date] from https://nepc.colorado.edu/review/achievement

Funding: This review was made possible in part by funding from the Great Lakes Center for Educational Research and Practice.

The National Education Policy Center (NEPC), a university research center housed at the University of Colorado Boulder School of Education, sponsors research, produces policy briefs, and publishes expert third-party reviews of think tank reports. NEPC publications are written in accessible language and are intended for a broad audience that includes academic experts, policymakers, the media, and the general public. Our mission is to provide high-quality information in support of democratic deliberation about education policy. We are guided by the belief that the democratic governance of public education is strengthened when policies are based on sound evidence and support a multiracial society that is inclusive, kind, and just. Visit us at: http://nepc.colorado.edu

NEPC Review: Testing Theories of Why: Four Keys to Interpreting US Student Achievement Trends (American Enterprise Institute, January 2025)

Reviewed by:

Chris Domaleski

The National Center for the Improvement of Educational Assessment

April 2025

Summary

A recent American Enterprise Institute report examines long-term patterns in student performance using data from national and international assessments. It identifies four major trends: (1) student performance peaked in the early 2010s before declining, (2) lower-performing students have seen the sharpest declines, (3) achievement gaps in the U.S. are widening more than in other countries, and (4) similar declines appear in adult literacy and numeracy scores. The report points to these trends to challenge implicit theories, such as the view that recent performance declines are solely a pandemic effect or that the effects are uniform for all students. Yet, while the report does effectively highlight these key trends, its analytical approach raises concerns. It overlooks important data sources that could provide a fuller picture. Additionally, it does not account for sources of uncertainty or provide guidance to detect differences that matter. Finally, the report's effort to "pressure test" possible explanations is underdeveloped—its theories lack depth and fail to engage with established research on causal inference. While the report is a useful starting point for discussions on U.S. student achievement, it falls short of delivering a framework for understanding *why* these trends exist.

NEPC
NATIONAL EDUCATION POLICY CENTER

NEPC Review: Testing Theories of Why: Four Keys to Interpreting US Student Achievement Trends (American Enterprise Institute, January 2025)

Reviewed by:

Chris Domaleski

The National Center for the Improvement of Educational Assessment

April 2025

I. Introduction

Concerns about academic achievement in the United States have intensified, especially after the educational disruptions caused by the COVID-19 pandemic. Despite the widespread use of standardized testing, obtaining clear and meaningful insights from the limited number of nationally comparable assessments remains challenging. This difficulty is exacerbated by the complexities inherent in interpreting differences among assessments over time, making it hard to identify common trends. It is even more difficult to develop compelling theories to explain the source of these trends.

It is this challenge that Nat Markus at the American Enterprise Institute seeks to address in *Testing Theories of Why: Four Keys to Interpreting US Student Achievement Trends*.[1] The report primarily relies on data from the National Assessment of Educational Progress (NAEP)—both the main NAEP and the long-term trend (LTT) NAEP.[2] It also draws on data from international assessments to include the Program for the International Assessment of Adult Competencies (PIAAC), the Program for International Student Assessment (PISA), and the Trends in International Mathematics and Science Study (TIMMS). The report focuses on trends in these assessments from approximately 1990 to 2022, emphasizing relative performance in comparison to a perceived peak at or around 2013, to shape an overarching narrative about student achievement in the US. The narrative is positioned as a challenge to some implicit theories, such as the view that recent performance declines are solely a pandemic effect or that the effects are uniform for all students.

The report's primary purpose is to establish criteria against which the credibility of potential theories about the influences on student achievement can be tested. Plausible "theories of why," the report posits, must account for the achievement patterns described in the report.

The timing of this report's release, just before the 2024 NAEP results, adds to its significance. The NAEP findings, widely seen as alarming, confirm that student performance across all grades and subjects remains below pre-pandemic levels, with achievement gaps continuing to widen.[3] Given these concerning trends, understanding both the factors influencing student achievement and the reasons behind them is more critical than ever.

II. Findings and Conclusions of the Report

The report is divided into two main sections. The first, which makes up the majority of the document, presents longitudinal assessment data to support four claims about national trends in student achievement. The second section briefly examines potential explanations for these trends, as the report attempts to "pressure test" three explanations.

The claims about student achievement trends are straightforward. First, student performance peaked in the early 2010s before beginning a steady decline. Second, these declines were largely due to sharper drops from lower percentiles, meaning lower-performing students experienced the greatest setbacks. Third, the widening achievement gaps are more pronounced in the US than in other countries. Finally, performance declines occur for adult learners as well as students, based on results from the PIAAC, which assesses examinees aged 16-65 in literacy and numeracy. The report asserts, "the similarities that adult PIAAC score trajectories have with student assessments are striking."[4] Specifically, performance peaked around 2012 and has since declined, with the largest drops occurring among lower-performing examinees, contributing to the US having the largest achievement gaps.

In the final section, the report attempts to "try out" some theories that might explain the patterns observed. The reader is cautioned in advance that no single explanation is sufficient, and theories are simply presented as "exercises in identifying plausible 'whys' behind these troubling trends."[5] The first theory suggests that national crises, such as the Great Recession and the COVID-19 pandemic, contributed to learning setbacks, particularly for lower-performing students. However, the report finds that theory inadequate because the achievement decline does not fully align with the timeline for the recession or the pandemic. The second theory points to education policy changes, particularly the transition from No Child Left Behind (NCLB) to the Every Student Succeeds Act (ESSA) and later to the implementation of Common Core Standards. This progression is cast as a weakening of federal accountability measures, and the report alludes to adverse effects of the Common Core for lower-performing students—but notes that this explanation does not account for similar declines in adult scores. The third theory focuses on cultural shifts, including increased screen time and declining engagement in recreational reading. While this theory may align with the timing of achievement declines, it fails to explain the more pronounced US gaps. The report concludes that each theory provides valuable insights, but no single explanation fully accounts for the performance trends.

III. The Report's Rationale for Its Findings and Conclusions

The report primarily focuses on analyzing national trends in student achievement based on four key claims. The supporting rationale comes from descriptive analyses of average and percentile scores from national and international assessments spanning from the 1990s to the present. The report asserts, "Together, these assessments provide a comprehensive view of US educational and skill trends across different populations over the past two decades."[6] Ultimately, the report aims to demonstrate that the consistency, direction, and magnitude of the findings provide strong evidence.

As noted, the report does not offer a rationale to account for the findings. In fact, the report stipulates that adequate explanations are difficult to construct and defend, as illustrated by the exercise of critiquing three possible hypotheses. Rather than offering a conclusive answer, the report aims to encourage further investigation into the factors that may explain the four key achievement trends identified.

IV. The Report's Use of Research Literature

Other than citations for data sources, the report omits references to or discussion of other studies that could corroborate or complicate its claims about assessment trends. To some extent, this is understandable because the trends seem evident based on the high-level descriptive data provided. However, citing studies that explored different assessments, grades, and/or content areas as well as studies that employed different units of analyses (states or district, for example) would strengthen the claims.

One prominent source of information about recent US achievement trends not referenced is the Standard Education Data Archive (SEDA), which provides comparable national data by aggregating assessments from thousands of school districts and placing them on a common scale.[7] Studies based on the SEDA data, such as the annual Education Recovery Scorecard (ERS), do not contradict the high-level national trends noted in this report, but they do provide a more detailed view of variability at the district and state level; they also point to areas of modest gains in recent years.[8] Another notable omission is any discussion of data from the well-known ACT and SAT assessments, which do not fully comport with the overall patterns in the report. ACT data has been generally more stable over the report's time period.[9] And, while interpretation of SAT data is complicated by changes to the test, its trends prior to 2017 are also relatively stable with a modest peak in the early to mid-2000s.[10]

More importantly, the report lacks a serious discussion supported by the literature to address the characteristics of credible "theories of why." The report seems to adopt a posture of "just asking questions," in an attempt to stave off responsibility for creating a credible explanation of the findings. While this may have been a deliberate choice to keep the report's scope manageable, it highlights the need to define the hallmarks of good explanations rather than to simply illustrate weak ones. That is, the report could draw from research literature to outline key attributes of strong causal investigations. For example, strong re-

search should: establish that changes in one variable are related to changes in another (*covariation*); demonstrate that the reported cause occurred before the effect (*temporal precedence*); rule out alternative explanations; account for context and interaction effects; and, emphasize the role of experimental design and replication.[11] Including these foundational elements would strengthen the report's ability to guide further inquiry.

V. Review of the Report's Methods

Because data are drawn from several different national and international assessments with different scales, the report uses standard deviation units to measure how much each year's average score differs in comparison to the 2013 high point average. Displays are further detailed by percentile (10th, 25th, 50th, 75th, and 90th) to show patterns across higher and lower levels of achievement.

While methodical choices for displaying performance patterns are generally reasonable, the report omits any analyses or discussion to address the extent to which differences are influenced by error. All assessments have some degree of error related to measurement and sampling. For this reason, NAEP results are typically reported with error bands or are flagged to note whether values are statistically different, but estimates of uncertainty are not acknowledged in this report. For example, a review of the NAEP website reveals that between 2007 and 2017, there was not a statistically significant difference in NAEP average scale scores in fourth grade reading.[12] This fact alone neither invalidates the methods nor obviates the overall claims about performance trends, but it does point out that there are exceptions to the patterns that may be noteworthy, especially given the report's focus on examining all performance differences as a deviation from 2013.

On a related note, when the report switches from graphing average assessment scores to graphing how those scores differ from those in 2013, the vertical or y axis shrinks considerably, making patterns look much more pronounced. This is because differences are expressed in standard deviation units and then reported and graphed in percentages. While this metric standardizes the scale across assessments, its interpretation may not be intuitively clear. For example, the average Grade 4 NAEP math score in 2003 is 235; it peaks in 2013 at 242—a difference of seven points. But when that difference is expressed and plotted as a percentage difference in standard deviation units, its value is somewhere in the range of -20% to -30% (the precise value is difficult to obtain because results are shown in graph format only). The magnitude (seven points) is unchanged, but the unit and scale used to report on it (-20% to -30%) seems much larger and may confound interpretation.

In short, the methodology offers no information regarding errors and also makes patterns appear unusually pronounced—making it difficult to reliably identify differences that matter.

VI. Review of the Validity of the Findings and Conclusions

The report's claims about longitudinal achievement trends are credible, if incomplete and potentially overstated. The credibility is bolstered by the inclusion of data from four different well-established national and international assessment programs. The data analysis methods are defensible and are provided with sufficient detail to make a convincing case for the four proposed trends.

However, there are some noteworthy challenges to the validity of these findings. First, the report omits some sources of evidence that could complicate the narrative, such as data from SAT or ACT, or analyses based on state and district data, such as those compiled by SEDA. Additionally, the report neither addresses sources of error nor provides any guidance to help the reader identify differences that matter.

Regarding the three "theories of why" included in the report, no claim about their validity is necessary, as the report itself acknowledges that each is inadequate. These theories are presented as an exercise in evaluating plausible explanations, but the analysis is incomplete because it does not incorporate established research on making sound causal inferences. Drawing from this literature could have improved the rigor of the theoretical exploration and provided a stronger foundation for understanding the observed trends.

VII. Usefulness of the Report for Guidance of Policy and Practice

For policymakers and practitioners seeking to explore national and international achievement trends, this report serves as a starting point. It draws on multiple decades of data from various sources to present an imperfect but generally persuasive case for four overarching patterns in student achievement.

However, the report falls far short of its claim to be *testing theories of why*. Without a more robust approach to causal inquiry, the report remains a descriptive exercise rather than a substantive exploration of the factors driving student achievement trends.

Notes and References

1 Malkus, N. (2025, January). *Testing theories of why: Four keys to interpreting US student achievement trends*. Washington, DC: American Enterprise Institute. Retrieved February 22, 2025, from https://www.aei.org/research-products/report/testing-theories-of-why-four-keys-to-interpreting-us-student-achievement-trends/

2 U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP). *About the NAEP long-term trend assessment*. Retrieved February 22, 2025, from https://www.nationsreportcard.gov/ltt/about/reporting/?age=9

Note: the main NAEP is administered every two years in grades 4, 8, and 12; NAEP Long-Term Trend (LTT) is administered approximately every four years to 9, 13, and 17 year olds.

3 U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP). *10 takeaways from the newly released 2024 NAEP results*. Retrieved February 22, 2025, from https://www.nagb.gov/powered-by-naep/the-2024-nations-report-card/10-takeaways-from-2024-naep-results.html

4 Malkus, N. (2025, January). *Testing theories of why: Four keys to interpreting US student achievement trends* (p. 10). Washington, DC: American Enterprise Institute. Retrieved February 22, 2025, from https://www.aei.org/research-products/report/testing-theories-of-why-four-keys-to-interpreting-us-student-achievement-trends/

5 Malkus, N. (2025, January). *Testing theories of why: Four keys to interpreting US student achievement trends* (p. 14). Washington, DC: American Enterprise Institute. Retrieved February 22, 2025, from https://www.aei.org/research-products/report/testing-theories-of-why-four-keys-to-interpreting-us-student-achievement-trends/

6 Malkus, N. (2025, January). *Testing theories of why: Four keys to interpreting US student achievement trends* (p. 14). Washington, DC: American Enterprise Institute. Retrieved February 22, 2025, from https://www.aei.org/research-products/report/testing-theories-of-why-four-keys-to-interpreting-us-student-achievement-trends/

7 Reardon, S.F., Fahle, E.M., Ho, A.D., Shear, B.R., Saliba, J., Min, J., Shim, J., & Kalogrides, D. (2025). Stanford Education Data Archive (Version SEDA 2024). Retrieved February 23, 2025, from https://purl.stanford.edu/pt329xg7054

8 Fahle, E., Kane, T.J., Patterson, T., Reardon, S.F., & Staiger, D.O. (2025). *Pivoting from pandemic recovery to long-term reform: A district-level analysis*. Education Recovery Scorecard. Retrieved February 23, 2025, from https://educationrecoveryscorecard.org/wp-content/uploads/2025/02/Pivoting-from-Pandemic-Recovery-to-Long-Term-Reform-A-District-Level-Analysis.pdf

9 National Center for Education Statistics. (2024). U.S. Department of Education, Institute of Education Sciences. Retrieved February 23, 2025, from https://nces.ed.gov/programs/digest/d23/tables/dt23_226.50.asp

10 National Center for Education Statistics. (2024). U.S. Department of Education, Institute of Education Sciences. Retrieved February 23, 2025, from https://nces.ed.gov/programs/digest/d19/tables/dt19_226.20.asp

11 Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

12 National Center for Education Statistics. (2024). NAEP reading: National trends and student skills. The Nation's Report Card. Retrieved February 28, 2025, from https://www.nationsreportcard.gov/reports/reading/2024/g4_8/national-trends/?grade=4