

TL;DR: There is real educational improvement behind this story—but the rankings make the improvement look more sweeping than the evidence supports. Y

our suspicion is justified: accurately calculated test scores can still create a misleading impression of a state's overall education system.

The strongest qualifications are that **rankings can rise when competitors decline; third-grade retention complicates fourth-grade comparisons; and early reading gains have not consistently translated into stronger performance in later years.** A particularly relevant 2026 study examines that last issue directly.

1. The underlying NAEP scores are credible. The interpretation needs more caution.

NAEP uses a common national assessment and representative samples. A state cannot raise its NAEP score simply by making its own state test easier or lowering its passing threshold. Its reading assessment includes comprehension of literary and informational texts, so these results measure more than children's ability to sound out words. nagb.gov

But these are **statistical estimates**, with sampling uncertainty. A numerical increase is not necessarily a statistically significant increase, and neighboring positions in a ranking may not represent meaningful differences.

An important qualification missing from the article: the National Assessment Governing Board reported **no statistically significant state reading-score increases in either fourth or eighth grade between 2022 and 2024.** Louisiana did, however, significantly exceed its **2019** fourth-grade reading performance. Those statements concern different comparison periods and are compatible. nagb.gov

That makes the article's broad description of sustained post-pandemic NAEP gains too imprecise. We need to specify the subject, grade, starting year, and statistical significance.

 2. Mississippi’s ranking improved much more recently than its actual reading score.

Here is the clearest example:

Year	Mississippi fourth-grade NAEP reading score	What the comparison shows
2013	209	Low starting point
2019	219	Substantial improvement before the pandemic
2022	217	Pandemic-era setback
2024	219	Return to approximately the 2019 level

Scores are rounded scale scores, not percentages correct. Historical federal data and the 2024 state report establish this pattern. The 2022–2024 increase was not statistically significant. nces.ed.gov

Mississippi moved from approximately **29th in 2019 to ninth in 2024**, while its rounded score stayed at 219. The national decline therefore accounts for much of that more recent ranking improvement. Its substantial earlier improvement—from 2013 to 2019—should still receive credit. sciencedirect.com

Maintaining achievement while other states lose ground is valuable. But it is a different accomplishment from continuing to produce large absolute gains.

■ 3. Holding students back can affect the comparison—but does not explain away the gains.

Mississippi's third-grade promotion policy creates a legitimate measurement concern because fourth-grade NAEP tests children **by grade, rather than at a fixed age**.

There are two possible effects:

- Children retained in third grade are absent from the fourth-grade group they otherwise would have joined.
- When they reach fourth grade, they have had an additional year of schooling and maturation.

Consequently, a higher fourth-grade average does not automatically mean children of the same age learned more in the same amount of time.

However, the simple claim that Mississippi permanently removes its weakest readers from testing is wrong. Retained children ordinarily advance and become eligible for fourth-grade testing later. Researcher Noah Spencer also points out that Mississippi's initial 2015 fourth-grade gains preceded exposure of that tested group to the retention requirement. These facts weaken the claim that the entire improvement is a statistical artifact. statmodeling.stat.columbia.edu

A Boston University study found substantial sixth-grade reading benefits for students affected by Mississippi's retention policy near the promotion threshold. But its own limitations matter: it evaluated **retention plus additional instructional support**, and it could not compare test outcomes at the same age. It therefore does not establish that repeating a grade alone was responsible. wheelockpolicycenter.org

Louisiana provides an especially useful check: its new third-grade reading retention requirement first applied in **2024–2025**, after the 2024 NAEP assessment. That requirement cannot explain its earlier fourth-grade NAEP gains. doe.louisiana.gov

My assessment: retention deserves scrutiny, but dismissing the Southern gains as merely “keeping weak students out of the test” goes beyond the evidence.

■ 4. The strongest concern is whether early gains last.

This is where your question about overall educational quality is most compelling.

In 2024, Mississippi's **eighth-grade reading score was 253, below the national public-school average of 257**. It was not significantly different from Mississippi's 2022 score—or its 1998 score. That is a much less impressive picture than the fourth-grade ranking. nces.ed.gov

This comparison does not prove the early reforms failed: NAEP does not follow the same individual children, and pandemic disruption complicates comparisons across years. Nevertheless, it clearly limits claims of a comprehensive educational transformation.

More directly relevant is Noah Spencer's **2026 working paper, “Long-Run Effects of Early Literacy Policy: Evidence from Mississippi.”** He previously published research supporting the reforms' effects on fourth-grade achievement. His follow-up reports little evidence of effects on:

- ACT scores.
- ACT participation, used as an indirect indicator of dropout.
- Students' postsecondary plans.
- Advanced Placement participation and performance.

He reexamined the early results and still found large fourth-grade effects. His interpretation is that those initial effects faded, suggesting a need for sustained support in later grades. **This is preliminary working-paper evidence; I verified the abstract, rather than independently auditing its full analysis.** papers.ssrn.com

That is remarkably close to the possibility you raised: **real improvement on an early educational measure, without demonstrated improvement across the later outcomes we ultimately care about.** It does not establish that there were no lasting benefits whatsoever.

■ 5. The article combines different kinds of rankings and evidence.

Three distinctions matter:

Measure	What it tells us	What it does not establish
Achievement level	How students perform now	How much schools improved their performance
Growth or recovery	How much scores changed over a period	Whether students now perform at a high level
Demographically adjusted performance	How results compare with those predicted for similar students	A definitive ranking of overall school quality

For example, the article’s Tennessee claim—second in math growth and fourth in reading from 2022–2025—comes from the **Education Scorecard**, which combines state assessments with national assessment data. Tennessee’s education department specifies comparison pools of **38 states for math and 35 for reading**, rather than all 50. That is encouraging evidence, but it is not the same statistic as a nationwide NAEP achievement ranking. [tn.gov](https://www.tn.gov)

Demographic adjustments can make comparisons fairer, especially for high-poverty states. But the Urban Institute explicitly acknowledges limitations in income measures and special-education classifications. Its initial 2024 adjustment also used 2022 demographic adjustments, assuming population characteristics had not changed substantially. These are useful analytical estimates, not an unqualified verdict on which state provides the best education. [Urban Institute](https://www.urbaninstitute.org)

Student exclusions, participation, and public-school enrollment changes also warrant examination. NAEP documents inclusion of students with disabilities and English learners. **I did not find evidence sufficient to attribute these states’ gains to improper exclusions**, and I would not make that accusation without a detailed sample audit. [NAEP](https://www.naep.gov)

 6. The policy recommendations are more plausible than the article's certainty about causation.

There is evidence supporting comprehensive literacy reform. The 2026 Education Scorecard identifies reading recovery in Mississippi, Louisiana, Tennessee, D.C., and several other jurisdictions implementing such reforms. But it also identifies reform states that have **not** experienced a turnaround. The researchers call for further evaluation of specific components, including literacy coaches and retention. [Education Scorecard](#)

That supports studying and adapting successful practices. It does not establish that the author's entire package—accountability, political leadership, charter policy, curriculum, and tutoring—caused the gains, or that opposite policies caused every struggling state's decline.

Similarly, lower spending alongside better results does not establish that spending is unimportant. It raises questions about how resources were used, what services cost locally, and which investments made a difference.

The supplied article is an opinion essay by a former Tennessee education commissioner who now leads Accelerate. That perspective is relevant to reading its confident policy prescriptions; it does not invalidate the evidence. Opinion | These states show how...

For PSA's purposes, I would treat these states as promising sources of practices to investigate, while requiring evidence beyond fourth-grade rankings.

The most useful checks are whether gains **persist into adolescence, reach the lowest-performing students, improve comprehension and writing, and lead to stronger high-school and postsecondary outcomes.**

On those broader tests, the story remains substantially less settled than this article suggests.