

NEPC Review: Inside the Mississippi Marathon (Progressive Policy Institute, April 2026)



Reviewed by:

**Gregory Camilli
Rutgers University**

July 2026

National Education Policy Center

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

Acknowledgements

NEPC Staff

Faith Boninger
Publications Manager

Elaine Duggan
Production Design

Alex Molnar
NEPC Director

Kevin Welner
NEPC Director

Suggested Citation: Camilli, G. (2026). *NEPC Review: Inside the Mississippi marathon*. Boulder, CO: National Education Policy Center. Retrieved [date], from <https://nepc.colorado.edu/review/mississippi>

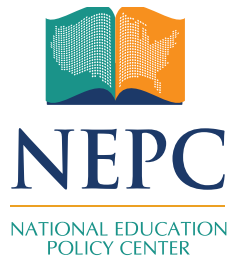
Funding: This review was made possible in part by funding from the Great Lakes Center for Education Research and Practice.



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

This publication is provided free of cost to NEPC's readers, who may make non-commercial use of it as long as NEPC and its author(s) are credited as the source. For inquiries about commercial use, please contact NEPC at nepc@colorado.edu.

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: Inside the Mississippi Marathon (Progressive Policy Institute, April 2026)

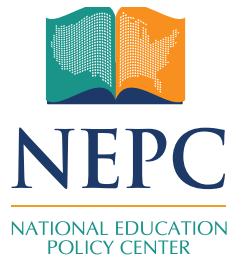
Reviewed by:

Gregory Camilli
Rutgers University

July 2026

Summary

A recent report from the Progressive Policy Institute, *Inside the Mississippi Marathon*, presents an insider account of how the K–3 education system in Mississippi evolved over roughly two decades. The report layers multiple policy framings around the state’s 2013 Literacy-Based Promotion Act (LBPA), including third-grade retention, accountability, instructional coaching, ongoing assessment, and the so-called science of reading. The report explicitly rejects the “Mississippi Miracle” framing in favor of a two-decade “marathon,” offering a narrative explanation of work that went into the state’s reforms. This review focuses on just a slice of the report: its description of the state’s performance in reading on the National Assessment of Educational Progress (NAEP), and, to a lesser degree, the policy implementation. Mississippi’s NAEP performance has indeed improved, and while this shift appears to be accelerating relative to public school performance, the gap’s closure partly results from a downward national trend in public school performance. The report also identifies four drivers of improvement in NAEP performance including standards and testing, accountability, teacher training in the science of reading, and instructional coaching; however, the extent of test preparation in third grade has not been generally recognized in academic critiques or press coverage about Mississippi. The review also examines the linkages between the LBPA and state performance in NAEP reading with respect to policy implementation, methodology, instruction, and statistical selection.



NEPC Review: Inside the Mississippi Marathon (Progressive Policy Institute, April 2026)

Reviewed by:

Gregory Camilli
Rutgers University

July 2026

I. Introduction

The *Mississippi Marathon*, released by the Progressive Policy Institute (PPI), is Rachel Canter's account of how the K–3 education system in Mississippi evolved over roughly two decades.¹ Canter is the current Director of Education Policy at PPI, and was formerly the director of Mississippi First, an education nonprofit in Mississippi. The report provides an insider's view of an effort to address persistent educational challenges in Mississippi. The report discusses multiple policy framings and initiatives (e.g., the science of reading and third-grade retention) within the state's Literacy-Based Promotion Act (LBPA), which originally took effect on April 18, 2013. The LBPA was intended to address early reading proficiency by aligning standards, teacher preparation, coaching, accountability, and intensive screening.² As a policy narrative, the report lays blame evenly across a wide swath of what Canter calls "pro-status quo" views for the previous lack of progress in reading instruction, not sparing the whip for any stripe of partisanship, ideology, or conspiracy theory.³

The "Mississippi Miracle" has taken on a highly visible role in the thinking of some journalists and lawmakers.⁴ The *Mississippi Marathon*, however, explicitly rejects this framing, asserting instead that "Mississippi's progress is neither a miracle nor a myth, as some skeptics have insisted; it's been a two-decade marathon." The report contributes to an understanding of this progress, adding policy context and implementation to discussions about how to increase child literacy. However, the report is not a research study and it does not appear to be intended as one; indeed, empirical studies—whether positive or negative—are not mentioned. There is a substantial body of such research on the LPBA examining the empirical evidence of its effectiveness. Consequently, this review turns to additional material, and readers should be

aware that I examine the effectiveness of the LBPA in a context that is broader than the *Mississippi Marathon* report itself.

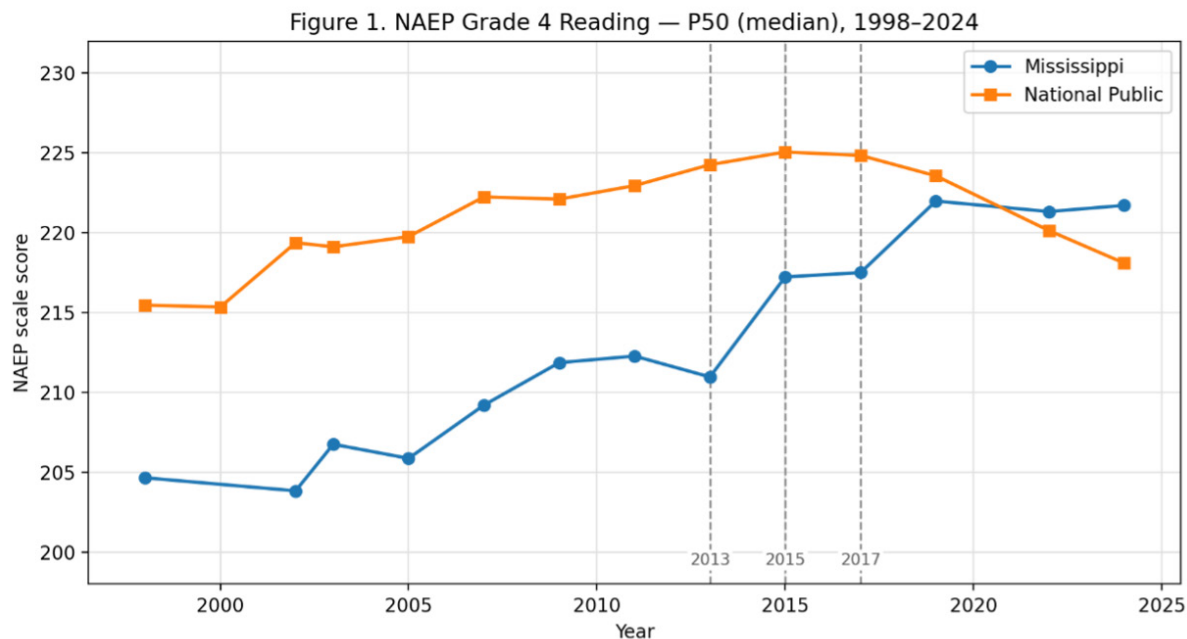
Rather than directly assessing the strengths and weaknesses of the report, this review uses it as a starting point or a hub for an empirical examination of the key claims regarding the LBPA—made both in this report and elsewhere—about the extent and causes of Mississippi’s improvement on the fourth-grade NAEP reading assessment. I focus on the success of LBPA, not the report per se. Accordingly, this review’s format and content vary somewhat from typical NEPC reviews.

II. Findings and Conclusions of the Report

The graphic most commonly cited as evidence of Mississippi’s progress is the state’s Grade 4 NAEP reading performance from 2002–2024, compared with those of all public schools, nationally. In 1998 the state’s median was about 11 points below the National Public median; by 2024 it was marginally above.⁵ This constitutes the popular headline framing of the “Mississippi Miracle.” This trend is shown in Figure 1. Two aspects are important to recognize in the plot: The median score (P50, for short) in Mississippi is going up *and* the national median score is trending down. This state of affairs is a double-edged sword for interpretation because two macro trends are evident, suggesting that students in Mississippi appear to have resisted the national trend. This is a good thing, but some would argue that it is not immediately clear what either of these long-term trends represents or (as discussed later in this review) why Mississippi students have developed immunity to the national trend.

As others have noted,⁶ the positive trends for Mississippi were apparent prior to recently enacted education laws and regulations, and in particular, the Literacy-Based Promotion Act (LBPA).⁷ The state’s pre-reform trend of improvement from 2002 to 2011 is largely unchanged through to 2024; the national trend, however, reversed. Looking again at the figure, we see a noticeable dip in scores the year that LBPA passed (2013). Then, roughly two-thirds of the state-national closure happened between 2013 and 2019, and the remainder happened between 2019 and 2024. I note later in this review that Mississippi’s NAEP performance in mathematics (not shown here) has experienced a similar trend and involves similar challenges of interpretation. Despite these nuances, the headline chart is a big deal in the area of early literacy.⁸

Figure 1. NAEP Grade 4 Reading—P50 (Median), 1998–2024. Mississippi vs. National Publics



Source: Graphs were prepared with data downloaded from the NAEP Data Explorer (National Center for Education Statistics, 2024).

Although the first year of LBPA implementation was technically AY 2013-14, the PPI report notes:

No state department of education, though, can implement a complex policy at scale in just a few months, so many of the law’s requirements were phased in between fall 2013 and spring 2015, when the law required the third-grade retention policy to begin.⁹

In fact, full implementation of the various pillars of the LBPA arguably took place in 2017-2019.

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

No research argument is given for Mississippi’s improvement other than four time-series charts based on NAEP data. These charts visually depict the longitudinal performance for fourth and eighth graders in NAEP reading and mathematics. However, this absence of causal analysis is not a criticism. The report is not a research study, and the target audience for the report consists of practitioners and policy-makers. In Section VI, I will address the rationale for the claim of effectiveness of Literacy-Based Promotion Act (LBPA) implementation.

IV. The Report's Use of Research Literature

The PPI report also presents no research literature on effectiveness of the LBPA. Again, it is not presented or intended as a research study, and the target audience for the report consists of practitioners and policymakers. So this is not a criticism. In examining the extent and possible causes of Mississippi's improvement on the fourth-grade NAEP reading assessment, I discuss findings from key empirical studies in Section VI.

V. Review of the Report's Methods: Considering the Literacy-Based Promotion Act (LBPA) and How to Determine Its Impact

Since the PPI report does not itself address research methods, this review differs from other NEPC reviews in that my core focus is instead on the larger empirical issues underlying the report. In this section I first describe the nature of the LBPA "treatment" to provide context for understanding its effectiveness. I then describe and explain two methods that have been used to quantitatively estimate its impact by drawing together separate streams of analysis from the research literature.

Implementation of the Literacy-Based Promotion Act (LBPA)

The LBPA was not a single intervention but a phased multi-component policy package whose elements reached Mississippi K–3 students at different times and intensities depending on which school they attended. Five core statewide components of the Act applied to all (roughly 728) Mississippi public elementary schools containing Grade 3:

1. A universal screening test was administered three times per year in Grades K–3, beginning fall 2014.
2. A parental notification and individual reading plan requirement were triggered by screening results.
3. A retention gate at the end of third grade was first administered in spring 2015 and tightened by amendment in 2016.
4. A "science of reading" licensure requirement for newly credentialed pre-K–3 teachers was phased in through workforce turnover.
5. An A–F accountability system was implemented in 2013–14 by relying on a formula that included reading proficiency as one of six ELA proficiency cells.

In addition to these universal elements, the Mississippi Department of Education (MDE) selected schools for literacy support based primarily on historically low third-grade reading performance, with priority given to the lowest-performing schools statewide on the state reading assessment.¹⁰ Out of approximately 700 Mississippi public elementary schools containing Grade 3, this literacy support program reached 50 (~7%) in 2013–14, 87 (~12%) in 2014–15, and 126 (~18%) in 2015–16. In these schools, a literacy coach (hired, trained, and monitored by the state) was assigned to support teachers. The state also required designated schools to use a reading curriculum that the state had reviewed and approved as consistent with the “science of reading” (SOR).¹¹ The coaching role was to model lessons, observe and provide feedback, co-lead “Language Essentials for Teachers of Reading and Spelling” (LETRS) modules, and facilitate data meetings. LETRS, an eight-module science-of-reading professional development program delivered over two years, was required statewide for all K–3 teachers and school principals.

The LBPA thus created a coach-mediated intervention in designated schools, in which coaches had no direct involvement in student instruction. The law’s K-3 scope described the coaching roles for changing reading instruction across the entire K-3 span.¹² In particular, coaches were experienced teachers who were apparently each reassigned to one or more of the designated schools. When considering the number of K-3 teachers in designated schools, number of coaches, workday length, and time available net of administrative duty and travel, the coaching/coached time may equate to approximately 15-35 minutes per week per teacher.¹³

As noted above, a substantial majority of Mississippi Grade 4 students primarily attended non-designated schools during their K–3 years (those attending approximately 93% of schools in 2013–14, 88% of schools in 2014–15, and 82% of schools in 2015–16). In non-designated schools, LBPA exposure consisted of a modest state-wide bundle of supports including retention screening, possibly an Individual Reading Plan (IPA), and student instruction from a mix of grandfathered and recently hired SOR-licensed teachers.

Research Design

Figure 1 shows what happened to Grade 4 NAEP reading scores in Mississippi, but this does not address the question, “What would have happened to Mississippi students absent the LBPA?” To make a decent guess, one might calculate a simple post-LBPA minus pre-LBPA difference, assuming the pre-intervention trend would stay unchanged. A comparative design using a control group is usually preferred to pre-post differences by researchers to rule out potentially confounding variables that change in between pre- and post-testing, such as the increasing use of smartphones and social media. Yet there is a way to use the time series and comparative methods simultaneously. Suppose you want to know whether a policy caused an im-

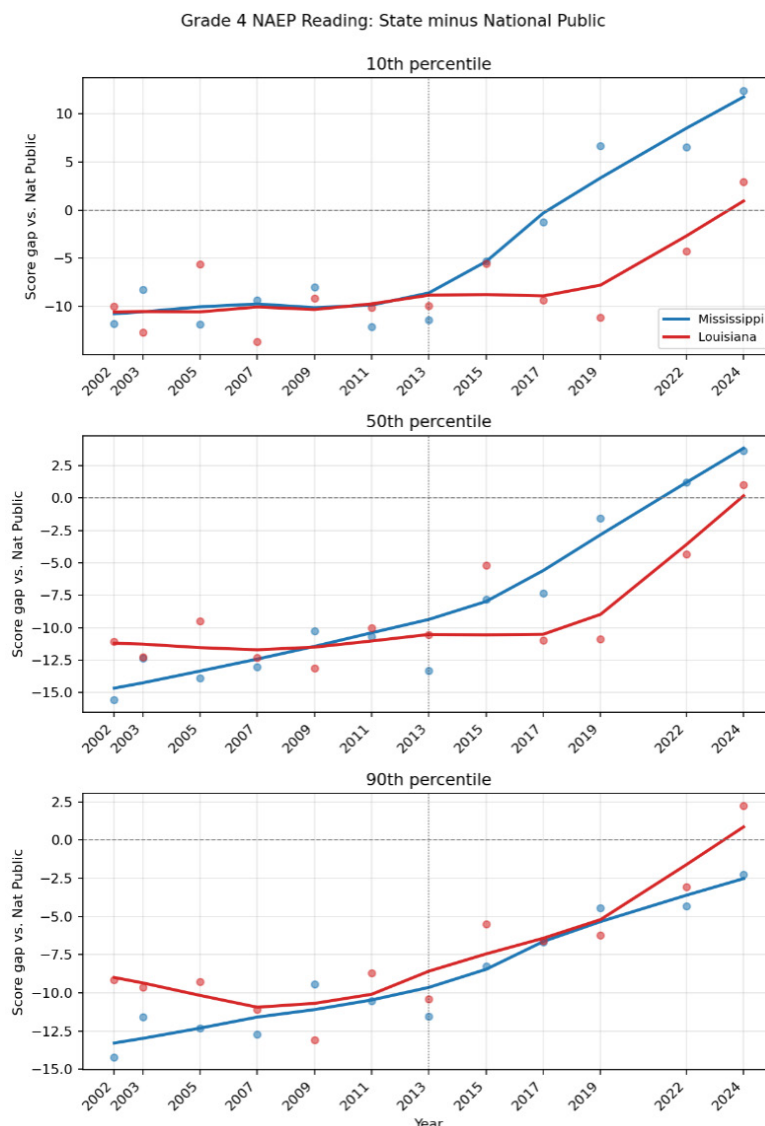
provement, by comparing Mississippi, which implemented the LBPA policy, to one or more similar states that did not. All these states have pre- and post-policy NAEP scores. The idea is to subtract the before-after change in the control state(s) from the Mississippi before-after change. What remains offers a better guess at the policy effect, net of whatever was already happening to both groups over time. This well-known method is called *differences-in-differences* (DID, for short), and it is usually preferred in program evaluation to time series or group comparison alone, when large-scale experiments cannot be conducted.

In Section VI, I illustrate this method using one state, Louisiana, as a comparison because it is regionally and demographically similar, with no third-grade gate (i.e., retention) filtering the fourth-grade NAEP pool over the entire 2002–2024 window (although Louisiana has recently adopted a Grade 3 retention law [Act 422] that went into effect in the 2024–2025 school year).¹⁴ The two states (Mississippi and Louisiana) have similar racial composition (Black 47% vs. 42%, and White 42% each), though Louisiana has roughly double the Mississippi Hispanic student population.¹⁵ These two states also have the highest levels of child poverty in the country.

VI. Review of the Validity of the Findings and Conclusions: Estimating the Impacts of the Literacy- Based Promotion Act (LBPA)

In this section, I will discuss empirical estimates of effectiveness and then describe three factors that might have contributed to the overall estimates. The factors are selection bias, instruction, and test preparation.

Figure 2. NAEP Grade 4 Reading—10th (P10), 50th (Median; P50), and 90th (P90) Percentiles. Mississippi and Louisiana vs. National Public



Source: Graphs were prepared with data downloaded from the NAEP Data Explorer (National Center for Education Statistics, 2024). Trendlines use LOWESS-smoothed (frac = .50) values.

Findings

In two 2026 studies, Wainer and Robinson used a time series approach to compare Mississippi's 2015-2017 pre-post LBPA improvement in NAEP reading, resulting in an estimate of about one scale score point improvement for 2015-2017.¹⁶ The year 2015 was chosen by as the baseline because the retention gate first applied to third graders in 2014-15. Extending the post intervention out another two years (from 2017 to 2019) would have shown an improvement of about five points in total. In

contrast, using the DID approach, Spencer used 2013 as the baseline year and obtained about a five-point gain when averaging across the 2015, 2017, and 2019 NAEP administrations. Comparing only 2019 performance to its baseline in 2013, a gain of about 8 ½ points was estimated.¹⁷ Note that these two results (pre-post and DID) are incommensurate because in the first, only the retention gate is considered to be the intervention, and in the second, all components of the LPBA constitute the intervention. Westall and Cummings conducted a related evaluation of early literacy policies, but Mississippi is not explicitly mentioned.¹⁸

Consider the time series plots for Mississippi and Louisiana given in Figure 2 for fourth graders at the 10th, 50th, and 90th percentiles (P10, P50, and P90) in NAEP reading. The dotted horizontal line (at 0.0) represents the national public percentile, and the smoothed blue and red lines represent differences from the national percentile. As shown in Table 1, I obtained the informal DID estimates for 2015-2017 and 2013-2019 using Louisiana and the comparison group. The largest effects are evident at P10 and the smallest at P90. The effects are also larger for 2013-2019 (possibly because of full implementation) than those for 2015-2017. The latter might be viewed as resulting from partial implementation of the LPBA.

Table 1. Mississippi Grade 4 NAEP Reading Change. Differences-in-Differences (DID) Estimates With Louisiana as Comparison

Percentile	2015-2017	2013-2019
P10	+5.0	+10.6
P50	+2.7	+5.3
P90	+0.8	+0.9
Average	+2.5	+4.8

Source: Estimates based on data downloaded from the NAEP Data Explorer (National Center for Education Statistics, 2024). The DID estimator nets out the Louisiana trend. “Average” refers to the NAEP average scale score. Values are NAEP scale-score points based on LOWESS-smoothed (frac = .50) trends.

The last row of Table 1 represents the effects for all students. However, this is more of a descriptive representation than an econometric estimate like Spencer¹⁹ and Mahmud²⁰ provided.

Selection Bias

This brief review of outcomes seems to indicate a rough consensus on the effectiveness of LBPA. That is not the case, however. The ballpark agreement is about the

increased scores, not about the cause of that increase. In the evaluation of LBPA, three complicating issues arise with respect to selection bias: retention, exclusion, and accommodation. Each of these mechanisms may have affected the magnitude of the LBPA effect of NAEP scores.

These mechanisms operate on different populations at different times; all can distort the impression of the effectiveness of an educational intervention. Collectively, they highlight the issue of selection bias.²¹ *Retention* in Grade 3 removes a student from being sampled for the fourth-grade NAEP assessment. In this case, retained students appear in the following year's fourth-grade sample frame. In contrast, NAEP *exclusion* policies remove from the tested cohort an already-sampled student with a disability (SD) or English Learner (EL) who is in Grade 4. For the latter, a NAEP score is never generated. *Accommodation* may keep a sampled SD or EL student in the tested cohort under modified administration conditions (extended time, separate setting, breaks, or scribing); these students contribute scores, but under conditions that differ from the standard administration, raising questions about score comparability.

Under the LBPA, a student can be held back in Grade 3 only once.²² The policy doesn't remove anyone from NAEP permanently; it just delays them by a year (if they remain in the district). A student retained in Grade 3 simply shows up in the Grade 4 sample one year later than they would have otherwise, albeit with an additional year of learning. But that delay still changes who is in the Grade 4 sample each year. In the first affected year, the lowest-scoring students are held back in Grade 3, so they are missing from Grade 4 altogether. Their absence pushes the measured Grade 4 scores upward. A year later, those students move up into Grade 4. These students likely tend to push NAEP scores downward (they are lower-scoring students), but the push down may be weaker than the push up due to the extra year of learning. This asymmetry results in selection bias, which Spencer²³ estimated to be about one NAEP point. In contrast, Wainer, Grabovsky, and Robinson reported a selection bias effect of about 9.2 points, but they considered neither Mississippi's pre-existing retention rate nor the likely counteracting effect of retained third graders entering fourth grade the ensuing year.²⁴

Unfortunately, these kinds of differences cannot be resolved by voting. Selection bias estimates are highly sensitive to counterfactual assumptions, and listening to experts wrangle on this issue is not fun. That said, I conducted an independent analysis, which provided an estimate of 1.07 points as an independent estimate of selection bias.²⁵ If one looks at the glass as half-full, a selection effect of one point does not cancel the effects obtained by Spencer or those shown in Table 1. Mahmud (2025) estimated the additional pure retention policy effect (that is, the threat of intervention without instructional change) also at roughly one point on the NAEP Grade 4 reading scale. Assuming an effect of five to eight points, factors unrelated

to literacy may have caused about 1/4 to 1/3 the total gain while the remainder may represent the effect of LBPA components. Studies by Spencer²⁶ and Mahmud²⁷ also suggest that instructional effects may have increased post-2017.

Instruction

It could be argued that the “threat of retention” was the primary treatment for the majority of Mississippi students, because students in non-designated schools received only distal instructional support in the form of teacher training. This threat itself may have caused principals, teachers and students to devote more instructional time and effort to literacy. In contrast, instructional support for teachers was arguably modest in LBPA-designated schools and less so in non-designated schools. Moreover, relatively few students in designated schools contributed to NAEP scores, and they likely had little influence on the overall trends shown in Figures 1 and 2. Relatedly, recall that the intervention did not include supplemental contact with students (e.g., tutoring). In short, the mechanism by which most of the students received instruction in reading is not sufficient to explain improved performance on NAEP, and neither is the selection effect.

A press release from the National Assessment Governing Board (NAGB) suggests that in 2015, the MDE revised its state assessment to align more closely with the NAEP assessment. The MDE “looked at the NAEP frameworks—the blueprints for the content and design of each assessment—to ensure that the new state assessment mirrored expectations on The Nation’s Report Card.”²⁸ As suggested by this text, two documents are available from NAGB:²⁹ It is unclear how exactly MDE used these documents, especially the NAEP item specifications (i.e., the test blueprint), to revise the state’s assessment.³⁰ Aligning state assessments with NAEP certainly could have increased NAEP performance, but it could also have led to score inflation.³¹

Test Preparation

Among the statewide policy changes in instruction was the use of testing as an instructional tool.³² Table 2 shows that the degree of local assessment administered prior to the Grade 4 NAEP increased substantially by 2016 or 2017. Such assessment is a feature of the approach used by *ExcelinEd*, the Florida organization that the report identifies as having played a key role in developing the LBPA.³³ While *ExcelinEd* emphasizes fewer, more effective assessments, NAEP is also prominently positioned in the organization’s message of educational reform.³⁴

Table 2. Mississippi Assessments and Frequency Administered Before 4th-Grade NAEP

Name of Assessment	Inception Date	Frequency
LBPA universal literacy screener (locally chosen in 2013–14; state-procured STAR Early Literacy from March 2014)	2013–14 AY (LBPA screening requirement)	Up to 3×/year in grades K–3; cumulatively up to about 12 administrations by winter of Grade 4.
Kindergarten Readiness Assessment (STAR Early Literacy)	August–September 2014	2× total: at kindergarten entry (fall) and exit (spring).
State-adopted dyslexia screener (Miss. Code § 37-173-15)	2013 statute; standardized statewide screener requirement effective July 1, 2017	2× total: spring of kindergarten and fall of Grade 1.
3rd-grade reading gate (STAR Early Literacy in 2014–15 and 2015–16; reading portion of the state ELA assessment — MAP, later MAAP — from 2016–17 onward)	Spring 2015	1× in spring of Grade 3, plus up to 2 retakes (1–3 administrations).
State assessment, grades 3–8 ELA and Math (PARCC for spring 2015 only; MAP, later renamed MAAP, from spring 2016)	Spring 2015 (PARCC); Spring 2016 (MAP/MAAP)	1× in spring of Grade 3 prior to Grade 4 NAEP.
District-chosen interim/benchmark assessments (NWEA MAP Growth, i-Ready, DIBELS, FAST, Renaissance Star, etc.)	Varies by district	Typically 2–3×/year, grades K through fall of Grade 4. Not state-mandated.

Sources: Mississippi Department of Education; Miss. Code § 37-173-15; Sun Herald (May 2017); PPI, Inside the Mississippi Marathon (implementation timeline, App. D).

The story becomes more interesting if one also considers fourth-grade NAEP mathematics, because the improved performance is roughly in line with the content emphasis in the Common Core: There is strong performance in the NAEP strands of Number and Operations, but much less improvement in Geometry.³⁵ Moreover, the improved performance of Mississippi on the NAEP assessment is accompanied by considerable test preparation. This provides some corroboration of a positive effect of Mississippi policies beyond selection bias, but it also turns the spotlight toward the intensity of test preparation as a potential mechanism for reaching across the curriculum and student populations.

Bass and Canter, in the report *Understanding District and State Testing*, rounded out the picture of assessment in four districts in Mississippi.³⁶ The report appears to indicate that Grade 3 was the most state-tested grade. In 2014-2015, third graders took four state tests in eight administrations, for a total of over 11 hours of state testing that year—more than any other grade. In 2017-2018 their state testing dropped to four tests across six administrations, for 9.5 hours of total state testing.³⁷ On the district side, testing was optional, but in the four studied districts, Grade 3 benchmarking ranged from five to 12 administrations each year.³⁸ This was in addition to the state tests.

Two of the four studied districts, Mannequin and Overton, administered more math tests than ELA tests, but the Bass and Canter report treats this as a classification artifact rather than a mathematics focus. The report attributes the greater number of math tests directly to the K–3 reading screener being counted as a state test rather than a district test; it asserts that the difference was not a result of a special focus on mathematics. The MDE paid for and required STAR Reading as the K–3 literacy screener, but bundled it with the mathematics assessment. This made the Grade 3 mathematics benchmarking a free rider on the reading mandate rather than a separate mandate.

Bass and Canter’s Appendix D includes a model Grade 3 page that makes the symmetry concrete, showing numerous assessments in both mathematics and reading.³⁹ In fact, their 2018 report’s main finding is that low-performing districts over-relied on “test prep”: “We were alarmed to find that teachers in both of the low-performing school districts in our sample stop teaching content at the end of the third nine weeks and begin daily, intensive test prep” (p. 23). The net picture for Grade 3 is straightforward: State mathematics testing amounted to one annual test (PARCC Math, then MAAP Math), while district testing varied by local choice but was considerable, consisting of STAR Math anywhere from three to nine times depending on the district, plus a local mathematics benchmark two to four times in two of the four districts. The latter testing was optional but available due to the bundled mathematics module.

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

Consensus on the effectiveness of an educational intervention is often elusive because multiple influences on outcome measures are difficult to distinguish. This is not a fatal criticism, but rather a well-known challenge on which substantial progress has been made. That said, some methodological problems are difficult to resolve, and they extend beyond choice of a comparison group. Despite such difficulty, it remains critical to identify plausible levers that were, or could be, operational at scale.

Policymakers may legitimately feel as though they can't wait for scientists to broadly agree, especially because evaluators themselves are playing a game of catch-up.

Inside the Mississippi Marathon is a valuable insider account of two decades of policy. An important strength is its detailed reconstruction of which reforms passed, in what order, and how they survived political resistance.⁴⁰ It then shifts gears to the narrative of what succeeded from a policy perspective. The report did not seek to answer questions foremost in the minds of quantitative analysts, but rather viewed progress through a policy lens. I have attempted in this review to provide some perspective on both policy implementation and methodology.

Inside the Mississippi Marathon makes a particular set of empirical claims about performance in reading: that Mississippi's change in NAEP performance is the visible signature of a sustained, layered reform package, and that the resulting closure of the gap with National Public NAEP scores reflects what Mississippi fourth graders have actually learned. Four previous challenges to this claim are examined in this review. First, as Figure 1 shows, the convergence between Mississippi and national public-school performance results in part from deceleration throughout the rest of the U.S. Yet the difference-in-difference approach indicates that netting out longitudinal factors, a relatively strong effect remains. Second, while statistical biases may exist, they appear to be relatively small. Third, attribution of the Grade 4 NAEP reading scores to the Literacy-Based Promotion Act (LBPA) is sometimes challenged by the argument that NAEP mathematics gains mirror gains in reading at Grade 4, yet the core intervention did not include mathematics. As demonstrated above, this may be true with regard to some elements of the LBPA (Science of Reading and grade retention, for instance), but it is not true in terms of interim assessment and summative testing. Finally, critics also point to the lack of consistency with Mississippi's less impressive Grade 8 NAEP scores, which fade some in math and even more in reading. This finding is also evident in the state assessment scores available from the Stanford Education Data Archive (SEDA) at Stanford University.⁴¹ Fade has also been observed in other well-known interventions.⁴² This is partly because learning is indisputably affected by factors other than the intervention, such as student culture, academic support, and family resources. This is not a disqualifying criticism of the LBPA.

The *Mississippi Marathon* report attributes the change in NAEP reading scores to a model for change based, in order, on four pillars: (1) standards, testing, and accountability; (2) consequences for poor performance (school ratings, district take-over authority); (3) evidence-informed instructional policy (here, Science of Reading); and (4) state-level support for implementation (coaching, screeners, training, and curriculum control). Anyone attempting to understand the effects of the LPBA must come to grips with the nature of this package of policy initiatives, and statistical analysis will not sort out the qualitative issues involved. In my opinion, there is

substantial evidence of improved NAEP reading scores, but that is only the “what” half of the story. The “why” half requires understanding how the four policy pillars gave rise to increased reading scores. The third and fourth pillars are typically considered “policy implementation” and consist of factors such as coaching, instruction, and teacher training. Yet they most likely had modest impact on NAEP scores due to the relatively small number of students in designated schools and the even smaller number selected for the NAEP sample. While licensure in the science of reading may have had some impact, the similar pattern of longitudinal improvement in both mathematics and reading suggests a lever that is not literacy-based. The lever may be hiding in plain sight.

Mississippi Marathon may understate the effect on NAEP performance of the first two pillars: standards, testing, and accountability; and consequences for poor performance. By 2018, these two policy levers resulted in a substantial increase in the frequency of both formative and high-stakes assessment in reading and mathematics, as well as increasing the alignment of the state assessment to NAEP. In fact, the degree of test preparation in Grade 3 is surprising, and it is largely undocumented beyond Bass and Canter’s 2018 report. I will not attempt to sketch the vast literature on the outcomes and efficacy of test preparation, which is a notoriously controversial topic, other than to say that it varies highly in quality and effectiveness. In the right circumstances, testing—including summative, formative and interim, and classroom—is an essential instructional support. In other circumstances it may lead to increases in test scores without enlarging students’ capabilities—their freedom to do and be things they otherwise could not.⁴³ In any case, a coherent curriculum framework is necessary to guide how test-score information becomes meaningful and actionable for helping students to develop adaptable skills and durable conceptual understandings.

Notes and References

- 1 Canter, R. (2026, April). *Inside the Mississippi marathon*. Progressive Policy Institute. Retrieved May 25, 2026, from <https://www.progressivepolicy.org/inside-the-mississippi-marathon/>
- 2 Mississippi Literacy-Based Promotion Act. (2013, amended 2016). Miss. Code §37-177-1 et seq.
- 3 Canter, R. (2026, April). *Inside the Mississippi marathon* (p. 3). Progressive Policy Institute. Retrieved April 12, 2026, from <https://www.progressivepolicy.org/inside-the-mississippi-marathon/>
- 4 Kristof, N. (2026, February 9). These three red states are the best hope in schooling. *New York Times*. Retrieved June 22, 2026, from <https://www.nytimes.com/2026/02/09/opinion/red-states-good-schools.html>
- 5 Graphs were prepared with data downloaded from the NAEP Data Explorer (National Center for Education Statistics, 2024). I used medians (P50) and two other percentiles (P10 and P90) rather than averages for visual inspection.

National Center for Education Statistics (2024). NAEP Data Explorer; NAEP Inclusion Policy; A Closer Look at Exclusion and Accommodations as Related to Assessment Results; 2024 NAEP Mathematics Assessment: Results at Grades 4 and 8 for the Nation, States, and Districts. U.S. Department of Education.
- 6 For example:

Thomas, P L. (2019, December 6; updated September 27, 2023). Mississippi miracle, mirage, or political lie?: 2019 NAEP reading scores prompt questions, not answers [blog post]. *Radical Scholarship*. Retrieved May 18, 2026, from <https://radicalscholarship.com/2019/12/06/mississippi-miracle-or-mirage-2019-naep-reading-scores-prompt-questions-not-answers/>
- 7 Mississippi Literacy-Based Promotion Act. (2013, amended 2016). Miss. Code §37-177-1 et seq.
- 8 Mississippi Monitor (2025, January 30). *Mississippi 4th graders no. 1 in the nation for NAEP gains over time*. Retrieved June 8, 2026, from <https://www.themississippimonitor.com/mississippi-4th-graders-no-1-in-the-nation-for-naep-gains-over-time/>
- 9 Canter, R. (2026, April). *Inside the Mississippi marathon* (p. 53). Progressive Policy Institute. Retrieved April 12, 2026, from <https://www.progressivepolicy.org/inside-the-mississippi-marathon/>
- 10 Canter, R. (2026, April). *Inside the Mississippi Marathon* (pp. 24, 51-52). Progressive Policy Institute. Retrieved April 12, 2026, from <https://www.progressivepolicy.org/inside-the-mississippi-marathon/>
- 11 I could not determine whether there was a single curriculum, whether schools or districts used their own curricula, or other alternatives.
- 12 Mississippi Literacy-Based Promotion Act. (2013, amended 2016). Miss. Code §37-177-1 et seq.
- 13 This estimate is my back-of-the-envelope calculation.
- 14 Louisiana's test-based retention has shifted grades over time. From 1999–2017, students at Grades 4 and 8 failing the Louisiana Educational Assessment Program (LEAP) repeated the

grade. The legislature ended this mandatory retention in 2017. From 2023–present Act 422 created Louisiana’s first Grade-3 reading gate. Retention is based on a screening test (DIBELS), with multiple attempts and good-cause exemptions, modeled after the LBPA. This policy was implemented 2024-25.

- 15 National Center for Education Statistics. (2024). *Digest of education statistics 2023 (Table 203.70: Percentage distribution of enrollment in public elementary and secondary schools, by race/ethnicity and state or jurisdiction: Fall 2012, fall 2021, and fall 2022)*. U.S. Department of Education. Retrieved June 24, 2026, from https://nces.ed.gov/programs/digest/d23/tables/dt23_203.70.asp
- 16 Robinson, D.H. & Wainer, H. (2026, May 6). The Mississippi miracle: Smoke and mirrors vs. reality. *Skeptic*. Retrieved June 29, 2026, from <https://www.skeptic.com/article/mississippi-miracle-smoke-and-mirrors-vs-reality/>

Wainer, H., Grabovsky, I., & Robinson, D.H. (2026, January). On education miracles in general (and those in Mississippi in particular). *Significance*, 23(1), 32–35.
- 17 Spencer, N. (2024). Comprehensive early literacy policy and the “Mississippi Miracle.” *Economics of Education Review*, 103, 102598.
- 18 Westall, J. & Cummings, A. (2023). *The effects of early literacy policies on student achievement* (EdWorkingPaper No. 23-788). Annenberg Institute at Brown University). Retrieved July 1, 2026, from <https://edworkingpapers.com/ai23-788>
- 19 Spencer, N. (2024). Comprehensive early literacy policy and the “Mississippi Miracle.” *Economics of Education Review*, 103, 102598.
- 20 Mahmud, S. (2025). *Does state-mandated third-grade reading retention policy improve achievement? Evidence from a staggered-adoption difference-in-differences design* (EdWorkingPaper No. 25-1352). Annenberg Institute at Brown University. Retrieved June 8, 2026, from <https://doi.org/10.26300/boem-9a84>
- 21 Collins, T. (2019, December 18; author’s update August 5, 2022). *Mississippi rising? A partial explanation for its NAEP improvement is that it holds students back*. Thomas B. Fordham Institute. Retrieved June 30, 2026, from <https://fordhaminstitute.org/national/commentary/mississippi-rising-partial-explanation-its-naep-improvement-it-holds-students>
- 22 Mississippi Literacy-Based Promotion Act. (2013, amended 2016). Miss. Code §37-177-1 et seq. Retrieved June 24, 2026, from <https://law.justia.com/codes/mississippi/title-37/chapter-177/>
- 23 Spencer, N. (2024). Comprehensive early literacy policy and the “Mississippi Miracle.” *Economics of Education Review*, 103, 102598.
- 24 Wainer, H., Grabovsky, I., & Robinson, D.H. (2026, January). On education miracles in general (and those in Mississippi in particular). *Significance*, 23(1), 32–35.
- 25 I used the Mississippi 2013 NAEP Grade-4 reading distribution to conduct an analysis of selection bias. This NAEP administration occurred prior to implementation of LBPA. My simulation assumed a pre-existing retention rate of 3.3% already present in the 2013 NAEP data for Mississippi, and that an additional 2.4% of students would have been retained as a result of the LBPA. Rather than removing students at a hard cut, I modeled retention probabilistically: Each student’s probability of

retention was proportional to a Beta-Binomial (49, 1.5, 13) weight on their score bin, concentrating removal in the lower tail while allowing some low-scoring students to escape retention. The weights were scaled so that 2.4% of students were removed in expectation, and I computed the resulting selection bias as the difference between the post-removal and original means, averaged over Monte Carlo replications. The estimate is 1.07 points (SD 0.04). The students removed under this model average roughly 165 on the NAEP scale, well above the 112 mean of a hard bottom cut, which is why the estimate falls below the 2.36 points implied by strict truncation

Mississippi Literacy-Based Promotion Act. (2013, amended 2016). Miss. Code §37-177-1 et seq.

- 26 Spencer, N. (2024). Comprehensive early literacy policy and the “Mississippi Miracle.” *Economics of Education Review*, 103, 102598.
- 27 Mahmud, S. (2025). *Does state-mandated third-grade reading retention policy improve achievement? Evidence from a staggered-adoption difference-in-differences design* (EdWorkingPaper No. 25-1352). Annenberg Institute at Brown University. Retrieved June 8, 2026, from <https://doi.org/10.26300/boem-9a84>
- 28 National Assessment Governing Board (2018). Mississippi: NAEP gains follow state’s efforts to improve student achievement. Retrieved July 1, 2022, from <https://www.nagb.gov/content/dam/nagb/en/documents/newsroom/press-releases/2018/release-20181105-mississippi-narrative.pdf>
- 29 As specified by NAGB, “The 2026 NAEP Reading Framework describes the content and design of the 2026 NAEP Reading Assessment; it is intended for a general audience. A second document, the Assessment and Item Specifications for the 2026 NAEP Reading Framework, serves as the ‘test blueprint’ with information about passage selection, item development and other aspects of test development; it is intended for a more technical audience, including NCES and the contractors who will develop the NAEP Reading Assessment.”

National Assessment Governing Board. (2021). *Reading framework for the 2026 National Assessment of Educational Progress* (p. 1). Retrieved July 1, 2026, from <https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/reading/2026-reading-framework/NAEP-Reading-Framework-for-Board-Action.pdf>
- 30 National Assessment Governing Board. (2021). *Reading assessment and item specifications for the 2026 National Assessment of Educational Progress*. Washington, DC. Retrieved June 30, 2026, from <https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/reading/naep-2026-reading-specifications.pdf>
- 31 A skeptic might observe that revisions of multiple states’ assessments based on the NAEP framework and item specifications could also open a backdoor to a de facto national curriculum, whereas “The initial design of the NAEP program was developed carefully to protect the rights of the states. It virtually precluded using NAEP as a lever for policy changes in American education: NAEP sampled students by age, not by grade; NAEP reported results at the regional level of aggregation, not by states or districts; and NAEP participation was strictly voluntary.”

Bourque, M.L. (2009). *A History of NAEP achievement levels: Issues, implementation, and impact 1989–2009* (p. 7). Paper commissioned for the 20th Anniversary of the National Assessment Governing Board. Retrieved July 1, 2026, from <https://www.nagb.gov/content/dam/nagb/en/documents/what-we-do/quarterly-board-meeting-materials/2018-11/09-history-and-context-of-naep-achievement-levels.pdf>

- 32 Mississippi modeled its 2013 Literacy-Based Promotion Act (LBPA) on Florida’s 2002 retention reform, with the policy transfer occurring as the LBPA was drafted. The first documented contact between Mississippi and Florida policy infrastructures was 2009, when Canter met Patricia Levesque, Executive Director of ExcelinEd (Canter, 2026, Footnote 78). Levesque contacted Canter to say ExcelinEd was “interested in Mississippi” (p. 37). In spring 2012, ExcelinEd embedded its accountability expert Christy Hovanetz in the state accountability task force, on loan to MDE (p. 37). Canter describes the event as the pivot point at which “a real reform agenda seemed within the realm of possibility” (p. 37). The LBPA passed nine months later.

Canter, R. (2026, April). *Inside the Mississippi marathon* (p. 37). Progressive Policy Institute. Retrieved April 12, 2026, from <https://www.progressivepolicy.org/inside-the-mississippi-marathon/>

Florida Office of Program Policy Analysis and Government Accountability (2006). *Third grade retention policy leading to better student performance statewide* (Report No. 06-66).

Mississippi Literacy-Based Promotion Act. (2013, amended 2016). Miss. Code §37-177-1 et seq.

- 33 ExcelinEd. (2026, January 15). *Introducing the 2026 ExcelinEd education policy playbook*. ExcelinEd. Retrieved June 8, 2026, from <https://excelined.org/2026/01/15/foundations-for-a-brighter-future-introducing-the-2026-education-policy-playbook/>
- 34 In a 2020 statement published on the ExcelinEd website, ExcelinEd CEO criticized the National Center for Educational Statistics’s decision to put the 2021 National Assessment for Educational Progress (NAEP) test on hold because of the pandemic. She wrote that NAEP “is the only measure of student academic performance that is comparable from state to state and over time. It is an important external check on state data and gives us a better understanding of trends in student performance.”
- Levesque, P. (2020, November 25). 2021 Nation’s report card on hold [blog post]. *ExcelinEd*. Retrieved June 8, 2026, from <https://excelined.org/2020/11/25/2021-nations-report-card-on-hold/>
- Levesque has also written that “[NAEP] is the only measure of student academic performance that is comparable from state to state” and serves as “an important external check on state data.”
- Levesque, P. (2014, June). Student success depends on testing and accountability [blog post]. *ExcelinEd*. Retrieved June 24, 2026, from <https://www.excelined.org/edfly-blog/student-success-depends-on-testing-and-accountability/>
- 35 Camilli, G. (2021). The 2013-15 decline in NAEP mathematics: What it teaches us about NAEP and the Common core. *Measurement Interdisciplinary Research and Perspectives*, 19(4), 236–245. Retrieved June 22, 2026, from <https://doi.org/10.1080/15366367.2021.1873062>
- 36 Bass, A. & Canter, R. (2018, March). *Understanding district and state testing in Mississippi*. Jackson, MS: Mississippi First. Retrieved June 23, 2026, from <https://www.mississippifirst.org/reports/understanding-district-and-state-testing-in-mississippi/>
- 37 MKAS2 ×1, STAR Reading ×3, PARCC ELA ×2, PARCC Math ×2, where MKAS2 is the assessment system used to meet the requirements of the LBPA.

Bass, A. & Canter, R. (2018, March). *Understanding district and state testing in Mississippi*

(pp. 58-63). Jackson, MS: Mississippi First. Retrieved June 23, 2026, from <https://www.mississippifirst.org/reports/understanding-district-and-state-testing-in-mississippi/>

- 38 The Overton district (low-performing) ran 12 math administrations (STAR Math ×9, i-Ready Math ×3) within 29 total Grade 3 tests; Hillside (high-performing) ran STAR Math ×6 within 23 total—its most-tested grade, with no i-Ready or NWEA math until Grade 4; Mannequin (high-performing) ran five (STAR Math ×3, District Math ×2) within 15 total; and Sunset (low-performing) ran seven (STAR Math ×3, District Math ×4) within 27 total, but its four 180-minute District Math sittings alone consumed 12 hours.

Bass, A. & Canter, R. (2018, March). *Understanding district and state testing in Mississippi* (pp. 58-63). Jackson, MS: Mississippi First. Retrieved June 23, 2026, from <https://www.mississippifirst.org/reports/understanding-district-and-state-testing-in-mississippi/>

- 39 STAR Reading ×6, STAR Math ×6, plus one MAAP ELA, one MAAP Math, and one MKAS2, with reading administrations slightly exceeding mathematics.

Bass, A. & Canter, R. (2018, March). *Understanding district and state testing in Mississippi* (p. 64). Jackson, MS: Mississippi First. Retrieved June 23, 2026, from <https://www.mississippifirst.org/reports/understanding-district-and-state-testing-in-mississippi/>

- 40 The report notes, “Pundits have dubbed our story ‘the Mississippi miracle,’ as if it must have taken divine intervention for us to do what so many others are failing to do—improve education for kids of all races, incomes, and achievement levels.”

Canter, R. (2026, April). *Inside the Mississippi marathon* (p. 3). Progressive Policy Institute. Retrieved April 12, 2026, from <https://www.progressivepolicy.org/inside-the-mississippi-marathon/>

- 41 Reardon, S.F., Ho, A.D., Shear, B.R., Fahle, E.M., Kalogrides, D., & Saliba, J. (2026). *Stanford Education Data Archive* (Version 6.0). Stanford Digital Repository. Retrieved June 21, 2026, from <https://purl.stanford.edu/xh833nn4025>

- 42 Observers may want to keep in mind that neither the Tennessee STAR (class-size reduction) reform nor the Florida test-based retention interventions produced test-score gains that survived four years under a same-age comparison; in both cases the early achievement effects faded to statistical insignificance. STAR’s effects on test scores became statistically insignificant by Grade 8 (Krueger & Whitmore, 2001); Florida’s retention gains were statistically insignificant after six years (Schwerdt, West, & Winters, 2017). Yet both resulted in durable effects on at least some later non-test outcomes—STAR on college attendance and non-cognitive measures (Chetty et al., 2011) and Florida on high-school GPA and remedial-course-taking (Schwerdt, West, & Winters, 2017).

Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D.W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *Quarterly Journal of Economics*, 126(4), 1593–1660.

Krueger, A.B. & Whitmore, D.M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *Economic Journal*, 111(468), 1–28.

Schwerdt, G., West, M.R., & Winters, M.A. (2017). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. *Journal of Public Economics*,

152, 154–169.

43 Sen, A. (1999). *Development as freedom*. Alfred A. Knopf.