



# 11. Standards-Based Reform

---

Michael J. Petrilli

## Executive Summary

“Standards-based reform” in the heyday of the education reform movement was the strategy of setting statewide standards, measuring student performance against those standards, and then holding schools accountable for the results. It was at the heart of the No Child Left Behind Act (NCLB) and dominated education policy from the 1990s into the 2010s.

Did it succeed as an overall strategy? Were there individual components that were particularly effective? We know that student achievement improved markedly in the late 1990s and early 2000s—the very time that states were starting to put standards, tests, and “consequential accountability” into place. Some of the gains can be directly attributed to those policies, but the improvement was likely driven by other factors too.

On the flip side, when student achievement plateaued and even started to decline in the 2010s, it’s plausible that the tapering-off was related to the softening of school-level accountability. But hard evidence is scant, and it’s difficult to know for sure. We unfortunately have limited information about what exactly schools did to get those better results in the earlier era.

For policymakers, though, there are some clear lessons. Capacity at the school and district levels is critical. These reforms were driven by the assumption that schools knew what to do to get better and were not doing it. We learned they needed significantly more support to improve. We also need broader measures of success. NCLB set the same outcome goals for all students, but multiple achievement measures can better meet the needs of students and families. Reforms that change the day-to-day work of schools take a long time to implement, so policymakers need to be patient and stay the course. They also need to invest in research. There is more we can learn about what works.

- The No Child Left Behind era showed some successes, but how and why they occurred is unclear.

- One lesson learned is the need for multiple achievement measures instead of identical goals.
- Policy changes should result from more research—and more patience.

• • •

“Standards-based reform” in the heyday of the education reform movement was a bit like the title of a recent film: *Everything Everywhere All at Once*. The strategy of setting statewide standards, measuring student performance against those standards, and then holding schools accountable for the results was at the heart of the No Child Left Behind Act (NCLB) and dominated education policy for most of the “long NCLB period” from the 1990s into the 2010s. To many observers, standards-based reform *was* education reform, and so the question about whether standards-based reform worked is equivalent to asking whether education reform worked.

Answering that question is only possible if we define what’s in and what’s out: What counts under the umbrella of standards-based reform? Did it succeed as an overall strategy? Were there individual components that were particularly effective?

In this chapter, we will work our way through these and related questions, but readers should beware that the results will not be entirely satisfying. Get ready for a lot of shrugging.

We know, for example, that student achievement improved markedly in the late 1990s and early 2000s—the very time that states were starting to put standards, tests, and “consequential accountability” into place. Some of the gains can be directly attributed to those policies. But the improvement was likely driven by other factors, too, some of which had very little to do with education policy or even schools, such as the plummeting child poverty rate at the time.

On the flip side, when student achievement plateaued and even started to decline in the 2010s, it’s plausible that the tapering off was related to the softening of school-level accountability, as NCLB lost steam and eventually gave way to the Every Student Succeeds Act and the Common Core State Standards. But hard evidence is scant, and it’s difficult to know for sure, especially because—again—so much else was going on at the same time. That included the aftermath of the Great Recession (and its budget cuts) as well as the advent of smartphones and social media, which may have depressed student achievement just as they boosted teenage anxiety and depression.

And while we know that standards, testing, and especially accountability drove some of the improvements in student outcomes in the 1990s and 2000s, especially in math, we unfortunately have limited information about exactly what schools did to get those better results. For the most part, the “black box” that is the typical K-12 classroom stayed shut.

Here's the good news: despite all these uncertainties, there's still much we can learn from the era of standards-based reform—both for future efforts to use standards, assessments, and accountability to improve outcomes and for education reform writ large.

## A SHORT HISTORY OF STANDARDS-BASED REFORM

---

The NCLB Act locked into place a specific version of standards-based reform, one that incorporated a mishmash of ideas that had been floating around since the 1980s and arguably since the 1960s. Think of it like a dish at a fusion restaurant, reflecting a novel combination of flavors and culinary lineages—not always with a satisfying outcome.

One might even say that this version of standards-based reform was incoherent—which is ironic, given that coherence was arguably the number-one goal of the original progenitors of the idea. In a series of articles and books in the late 1980s, scholars Jennifer O'Day and Marshall Smith argued for what they called “systemic reform.” Their key insight was that the multiple layers of governance baked into the US education system as well as myriad conflicting policies emanating from the many cooks in the K-12 kitchen were pulling educators in too many directions. What we needed was to fix the system as a whole, to think comprehensively and coherently and thereby get everyone rowing in the same direction in pursuit of stronger and more equitable student outcomes.<sup>1</sup>

To do so, we needed to get serious about “alignment.” We should start with a clear set of desired outcomes, also known as standards, delineating what we expect students to know and be able to do—at the end of high school but also at key milestones along the way. Those curricular standards would set forth both the content of what kids needed to learn and the level at which they needed to learn it. Regular assessments would help practitioners and policymakers understand whether kids were on track to meet expectations and ready to progress to the next grade level and, ultimately, high school graduation. This approach would allow for the assessment of student performance against common expectations and criteria rather than measuring students against one another (norm-referenced evaluation and rankings) to determine academic achievement. But perhaps most importantly, all the other key pieces of the education apparatus needed to be aligned to the standards as well—especially teacher preparation, professional development, instructional materials, and funding systems.

O'Day and Smith didn't say much about “accountability” as we would later come to talk about it—consequences that would accrue to educators, especially for poor student performance. Instead, their focus was primarily on coherence, alignment, and building “capacity” in the system to improve teaching and learning.

Systemic reform was popular with traditional education groups.<sup>2</sup> It spoke to the frustration of classroom teachers as well as principals and superintendents, without directly threatening the

political power of key constituencies, especially teachers' unions. They welcomed the additional help envisioned by scholars such as O'Day and Smith—and the additional money.

But this approach was hardly the only school improvement game in town. Other ideas were gaining prominence in the late 1980s and early 1990s, too, ideas promulgated by governors, economists, political scientists, and business leaders. To oversimplify a bit, they coalesced around the “reinventing government” frame<sup>3</sup>—namely that to reform a broken system like K-12 education, leaders needed to embrace a “tight-loose” strategy: tight about the results to be accomplished and loose about how people closer to the problem might get there. This was how business titans of the time steered their organizations, especially as the economy was shifting to knowledge work. To get the best results, people on the front lines had to have the autonomy to make decisions and solve problems themselves in real time rather than take orders from the top. They should be rewarded when they improved productivity accordingly. But if they failed to generate the desired results, unpleasantness might be expected to follow. They might even lose their jobs.

This struck a chord among some education scholars as well. As far back as 1966's *Coleman Report*, we knew about the disconnect between education inputs and outcomes. If we wanted better results, it made sense to focus on the latter.<sup>4</sup> Furthermore, many of the reforms embraced in the wake of 1983's *A Nation at Risk* report tried to tweak inputs such as teacher salaries, course requirements, and days in the school year. In an era of stagnant achievement and widening achievement gaps, none of that seemed to be working. It was time, many thought, for something else.

By the early 1990s, the tight-loose frame was a big driver behind the charter schools movement and the notion of “accountability for results” for public schools writ large. Lamar Alexander, who was governor of Tennessee before becoming US secretary of education under George H. W. Bush, was apt to talk about “an old-fashioned horse trade”: greater autonomy for schools and educators in return for greater accountability for improved student outcomes.<sup>5</sup> And it wasn't just Republican governors who embraced this model; several Democratic ones did, too, especially southern governors such as Jim Hunt (North Carolina), Richard Riley (South Carolina), and Bill Clinton (Arkansas). It helped that the Progressive Policy Institute—a think tank for the New Dems—supported this approach enthusiastically.

This version of standards-based reform had some overlap with O'Day and Smith's systemic reform, especially when it came to the centrality of academic standards. But it put greater emphasis on the measurement of achievement against those standards—in other words, high-stakes testing—and especially on accountability measures connected to results. This reflected the thinking of both economists and political scientists, who thought that the right incentives might allow local schools and school systems to break through the political barriers to change. With enough pressure from on high, schools might finally put the needs of kids first rather than follow the lead of adult interest groups, especially unions. They would remove ineffective teachers from the classroom, for example, ditch misguided curricula, and untie the hands of principals. The assumption was that the major barrier to improvement was not

incoherence or the lack of capacity per se, but small-p politics and, especially, union politics. Getting the incentives right by tying real accountability to results could take a sledgehammer to the political status quo in communities nationwide.<sup>6</sup>

This made sense to some key actors on the political left as well, especially the Education Trust and other civil rights organizations. They bought into this version of standards-based reform but with an important twist: doing right by kids would be defined primarily as doing right by kids *who had been mistreated by the education system*. That meant Black, Hispanic, and low-income students especially. These reformers wanted to counterbalance the political power of the unions but also that of affluent parents and other actors who tended to steer resources to the children and families who needed them the least. They wanted to use top-down accountability to redirect money, qualified teachers, and attention to the highest-poverty schools and the most disadvantaged kids.<sup>7</sup>

These various flavors of standards-based reform were all in the mix in the 1990s, with many public discussions in particular about the wisdom of a strategy focused on “capacity building” versus one that stressed “accountability for results.”<sup>8</sup> The enactment of NCLB settled the debate; the accountability hawks won. Capacity building would mostly be put on the shelf in favor of a muscular, federally driven effort to hold schools accountable, especially for the achievement of the groups that most concerned civil rights leaders.

## ENTER NO CHILD LEFT BEHIND

---

The No Child Left Behind Act of 2001, the Bush-era reauthorization of the Elementary and Secondary Education Act, was the law of the land for an entire generation of students. The kids who entered kindergarten in the fall of 2002, nine months after then president George W. Bush put his signature on NCLB, were seniors in high school in December 2015 when then president Barack Obama signed into law the Every Student Succeeds Act (its reauthorized successor).

That’s not to say that the same policy was set in stone for those thirteen years. For the first half of its life, federal officials implemented it rather faithfully, but the second half came with major policy shifts driven by regulatory actions and what might be termed “strategic non-enforcement.” Let’s take a brief trip down memory lane.

“NCLB-classic”—which was the 2001 reauthorization of the 1965 Elementary and Secondary Education Act—centered on the three-legged stool of standards, tests, and accountability. But those three elements were not treated with the same level of prescription. States had complete control over their standards—both in terms of the content to be included and in terms of the level of performance that would be considered good enough. Not so when it came to the tests—those had to be given annually to students in grades three through eight in reading and in math, plus once in high school, plus three times in science. And the assessments had to meet a variety of technical requirements.

But where NCLB's designers really got prescriptive was around accountability requirements. They created a measure called adequate yearly progress, which judged schools against statewide targets for performance and decreed that subgroups of students—the major racial groups plus low-income kids, students with disabilities, and English learners—would need to hit those targets as well. If schools failed to achieve any of their goals in a given year, they would face a cascade of sanctions that grew more severe with each unsuccessful year. Students would have the right to attend other public schools in their same district and, eventually, to receive “supplemental education services” (i.e., free tutoring) from private providers. Districts were charged with intervening in low-performing schools with ever-increasing intensity.

NCLB had a plethora of other provisions, from mandating that schools hire only “highly qualified teachers” to bringing “scientifically based reading instruction” (now called the science of reading) to the nation's schools. Some of these other pieces could be considered capacity-building efforts. But overwhelmingly, NCLB was about accountability for results. It assumed that with enough pressure, schools and districts would cut through the Gordian knot that was holding them back in order to raise the achievement of students, especially those from marginalized groups. That was the theory. And as we'll get to in a moment, it partly worked.

But it also soon became clear that many schools and systems didn't know what to do in response to the accountability pressure—or couldn't steel themselves to make the requisite changes in long-established practices and structures. Some educators narrowed the curriculum, significantly expanding the time spent on math and reading at the expense of other subjects. Stories filled the nation's newspapers about schools teaching to the test, canceling recess, even ignoring lice outbreaks, all because of the accountability pressures of NCLB. In perhaps the most notable education scandal, teachers and principals in the Atlanta Public Schools district were found to have cheated on state-administered tests by providing students with the correct answers to questions and even changing students' answers and modifying test sheets to ensure higher scores.

## NCLB EVOLVES

---

As with most federal statutes, Congress was supposed to update NCLB after a few years. A reauthorization push in 2007 came close to doing so and would have made the law even tougher, but it fell apart under fierce opposition from teachers' unions and other education advocacy groups. So the law lumbered on even as it became clearer to its strongest supporters, including then education secretary Margaret Spellings, that parts of it were becoming unworkable.

One of the major issues was that an increasing number of schools were failing to meet NCLB's adequate yearly progress provisions. If tens of thousands of schools were deemed subpar, then the sting and stigma were lost, as was much of the motivation to do something to fix it. In

particular, the law's focus on achievement rather than progress over time was snaring virtually all high-poverty schools in its trap, given the enduring relationship between test scores and kids' socioeconomic backgrounds. Now that annual tests were in place, and states had, with federal money and support, built more sophisticated data systems, it was technically feasible to measure individual students' progress from one year to the next. Such measures were much fairer to schools whose students arrived several years below grade level. But these growth models weren't contemplated back in 2001, so they weren't allowed under the law.

Through a series of regulatory actions, Spellings (under George W. Bush) and Arne Duncan (under Obama) allowed states to make critical changes to their implementation of NCLB to address these concerns. They allowed growth models provided the models still expected students to hit "proficiency" within a few years. They loosened rules around supplemental services so that school districts could provide tutoring themselves rather than outsource it to private providers. The cascade of sanctions was replaced with a menu of intervention options and funded generously through the School Improvement Grants program—all meant to encourage "school turnarounds." An Obama-era waiver program allowed states even greater flexibility to tinker with their accountability targets in return for commitments to embrace other reforms the administration supported.

Meanwhile, states were working to address another key issue with NCLB: its encouragement of low-level academic standards and much-too-easy-to-pass tests. Because the law required states to set targets that would result in virtually all students reaching the "proficient" level by 2014, it incentivized states to set the proficiency bar very low. This, in turn, may have encouraged educators to engage in low-level instruction, with teaching to the test and "drill and kill" methods. It also provided parents with misleading information, as states told most parents that their children were "proficient" in reading and math, even if they were actually several years below grade level and nowhere near on track for college or a decent-paying career. In Tennessee, for example, the state reported that 90 percent of students were "proficient" in fourth-grade reading in 2009 while the National Assessment of Educational Progress (NAEP) had the number at 28 percent.<sup>9</sup> Advocates came to call this the "honesty gap."

Under the leadership of the National Governors Association and the Council of Chief State School Officers, states started collaborating on a set of common standards for English language arts and math—what would eventually become the Common Core State Standards. The hope was that, by working together and providing political cover to one another, the states would finally set the bar suitably high—at a level that indicated that high school graduates were truly ready for college or career and that would encourage teachers to aim for higher-level teaching. It would certainly be hard for the effort to result in worse standards than what most states had in place. Multiple reviews of state standards over the years from the American Federation of Teachers, Achieve, and the Thomas B. Fordham Institute found that they were generally vague, poorly written, and lacking in the type of curricular content that "systemic reformers" had envisioned so many years before.<sup>10</sup> It wasn't surprising, then, that so many educators reported teaching to the test. The tests became the true standards, and they were perceived to be of low quality too.

The Common Core standards were adopted by more than forty states in 2010 and 2011, changing the very foundation of NCLB’s architecture. No longer were states aiming to get low-achieving students to basic literacy and numeracy; now the goal was to get everyone to college and career readiness. But that shift was largely overlooked at the time, drowned out by a fierce political backlash to the Common Core. It mostly came from the right, as the newly emerging conservative populist movement seized on Obama’s involvement in encouraging the adoption of the standards (through his Race to the Top [RttT] initiative). Nonetheless, by 2015, more than a dozen states were using new assessments tied to the standards (largely paid for through RttT funds), and even today, most states still use the Common Core standards or close facsimiles.

## SO DID STANDARDS-BASED REFORM WORK DURING THE NCLB ERA?

---

As mentioned before, judging the success or failure of such a sprawling reform effort is hard to do. Thankfully, scholars Dan Goldhaber and Michael DeArmond of the CALDER Center at the American Institutes for Research offered a wonderful overview of the research literature in a recent report for the US Chamber of Commerce, *Looking Back to Look Forward: Quantitative and Qualitative Reviews of the Past 20 Years of K-12 Education Assessment and Accountability Policy*.<sup>11</sup> I strongly encourage readers to review their findings; allow me to summarize them here.

First, it’s clear that student achievement in the United States improved dramatically from the mid to late 1990s until the early 2010s—especially in math, especially at the elementary and middle school levels, and especially for the most marginalized student groups. Pointing to studies by M. Danish Shakeel, Paul Peterson, Eric Hanushek, Ayesha Hashim, Sean Reardon, and others, Goldhaber and DeArmond conclude that “the long-term gains on the NAEP reveal a decades-long narrowing of test score achievement gaps between underserved groups (e.g., students of color, lower achieving students) and more advantaged groups (e.g., White students, higher achieving students).”<sup>12</sup>

My own analysis of NAEP trends from that time period focused on the impressive gains made by the nation’s low-income, Black, and Hispanic students, especially at the lower levels of achievement.<sup>13</sup> The proportion of Black fourth-graders scoring at the “below basic” level on the NAEP reading exam, for example, dropped from more than two-thirds in 1992 to less than half in 2015. Likewise, the percentage of Hispanic eighth-graders scoring “below basic” in math dropped from two-thirds in 1990 to 40 percent in 2015. Those numbers were still much too high, but the improvement over time was breathtaking.

Nor was it just student achievement. High school graduation rates shot up as well, climbing fifteen points on average from the mid-1990s until today. We saw major improvements in college completion, too, with the percentage of Black and Hispanic young adults with four-year

degrees climbing from 15 percent and 9 percent, respectively, in 1995 to 23 percent and 21 percent by 2017. Some analysts have argued that these improvements might reflect a softening of graduation standards, but rigorous studies have found that a significant proportion of the gains were real.<sup>14</sup>

Alas, the progress in test scores stalled in the early to mid-2010s, and achievement even declined in some subjects and grade levels in the late 2010s, before the pandemic wiped out decades of gains. As Goldhaber and DeArmond explain, this has led some analysts to argue that the rise and fall of test-based accountability can explain the rise and fall of student achievement.

That's possible, but NAEP's design makes it hard to know for sure. What scholars can do is compare states with various policies (and policy implementation timelines) to try to link the adoption of standards-based reform to changes in student achievement. That's exactly what a series of studies did in the 2000s, including ones by Martin Carnoy and Susanna Loeb, another by Eric Hanushek and Margaret Raymond, and a seminal paper by Tom Dee and Brian Jacob.<sup>15</sup> The latter compared states that adopted "consequential accountability" in the late 1990s to those that adopted it in the early 2000s, once NCLB mandated them to do so. Dee and Jacob found large impacts of those policies on math achievement (an effect size in the neighborhood of half a year of learning), with even greater effects for the lowest-achieving students as well as Black, Hispanic, and low-income kids. The impacts on reading and science were null.

Another study, by Manyee Wong, Thomas D. Cook, and Peter M. Steiner, used Catholic schools as a control group and found more evidence that accountability policies raised achievement in math in the public schools.<sup>16</sup> Other research, also reviewed by Goldhaber and DeArmond, looked at the impact of NCLB on the so-called bubble kids—the students who were closest to the proficiency line or the schools most at risk of sanctions. Most studies found the largest gains for such students and schools, for better or worse.<sup>17</sup>

A brand-new study, by Ozkan Eren, David N. Figlio, Naci H. Mocan, and Orgul Ozturk, found that accountability policies had an impact on more than just test scores. "Our findings indicate that a school's receipt of a lower accountability rating, at the bottom end of the ratings distribution, decreases adult criminal involvement. Accountability pressures also reduce the propensity of students' reliance on social welfare programs in adulthood and these effects persist at least until when individuals reach their early 30s."<sup>18</sup>

Circumstantial evidence from individual states also points to a big impact from consequential accountability. Massachusetts, which combined standards-based reform with an enormous increase in spending in its 1993 Education Reform Act, saw student achievement skyrocket in the late 1990s and early 2000s—the much-remarked "Massachusetts miracle." Fourth-grade reading scores increased by nineteen points from 1998 through 2007—the equivalent of about two grade levels. Eighth-grade math scores jumped thirty-one points from 2000 to 2009. With its high-quality academic standards, intensive supports for teachers, lavish funding, and new high school graduation exam for students, the Bay State showed what was possible.

Nor was Massachusetts alone. Other states made significant progress, too, including Texas and North Carolina in the 1990s, Florida in the late 1990s and early 2000s, Mississippi in the 2010s, and the District of Columbia throughout the entire reform period.

What we can say, then, is that NCLB-style accountability worked, at least for a while and at least in math. Nationally, it didn't make an impact in reading, even though reading achievement was improving during the NCLB era (including in states like Massachusetts and Mississippi). We also aren't sure if achievement plateaued in the 2010s because accountability necessarily stopped working or because *accountability* stopped.

It doesn't help that we don't have much evidence about the mechanisms that might have driven the gains Dee and Jacob (and others) found. Did schools improve their approach to teaching mathematics? Did they make more time for intensive interventions such as tutoring, especially for their lowest-performing kids? Did they work harder or smarter to support teachers and get their best folks where they were needed most? Why did accountability lead to gains in math but not in reading?

We only have a few studies on how these policies might have changed classroom practice. As mentioned above, it was widely perceived that schools—especially elementary schools, where the schedule is more flexible—narrowed the curriculum and spent more time on math and reading and less time on social studies and science. Several teacher surveys showed this to be the case.<sup>19</sup> (Perhaps that's one reason standards-based reform failed to move the needle on reading achievement, given the growing evidence linking content knowledge in subjects like social studies to improvements in reading comprehension.<sup>20</sup>) The improvement of scores for bubble kids indicates that schools and teachers may have shifted their attention to kids near the proficiency line. And teaching to the test was also thought to be pervasive; some teacher surveys, for example, found that instruction became more teacher centered and focused on basic skills.<sup>21</sup>

Alas, studying policy implementation all the way into the classroom is difficult and expensive. So save from surveying teachers about their practice—which is better than nothing but not terribly reliable—not much else was done.<sup>22,23</sup> As a result, when it comes to changes that standards-based reform might have brought to the classroom, we have more questions than answers.

## **SCHOOL IMPROVEMENT, SCHOOL CHOICE, AND SCHOOL CLOSURE**

---

In 2009, the Obama administration successfully lobbied Congress to allocate \$3.5 billion (eventually growing to \$7 billion) into the Title I School Improvement Grants program. This sum was directed primarily to the 5 percent of schools in each state with the lowest academic achievement. The federal government instructed districts to select from four intervention

options, from replacing the principal to closing the school entirely. Most selected the least onerous option, and perhaps for that reason, a federal evaluation of the effort found no impacts on test scores, high school graduation, or college enrollment.<sup>24</sup>

However, as Goldhaber and DeArmond explain, some local and state studies did find positive impacts arising from the SIG initiative. California's implementation was particularly well studied by scholars including Thomas Dee, Susanna Loeb, Min Sun, Emily K. Penner, and Katharine O. Strunk.<sup>25</sup> Both statewide and in particular cities, the results were generally positive, with improvements in both reading and math. This may be because California required its lowest-performing schools to implement more intensive interventions. It also focused a great deal of money—up to \$1.5 million—on each school and gave the school lots of help in spending it well.

Though not addressed by Goldhaber and DeArmond, another place to look for lessons on accountability is the school choice movement. In particular, we can compare the relative success of charter schools with private school choice, given that the former operates under a strict accountability regime while the latter, in most states, does not. A growing body of research, including a new study from CREDO at Stanford University, shows charter school students outpacing their traditional public school peers both on test scores and on long-term outcomes such as college completion. That is especially the case for urban charter schools and for Black and Hispanic students.<sup>26</sup>

Private school choice programs, on the other hand, have been markedly less effective in boosting student outcomes, at least as judged by test scores. Recent studies of large-scale voucher programs in Ohio, Indiana, and Louisiana all show voucher recipients trailing their public school peers on test score growth, sometimes quite significantly.<sup>27</sup> To be sure, another set of voucher studies finds positive long-term impacts on measures such as high school graduation and college enrollment.<sup>28</sup> But the negative findings on achievement are still worrying and might reflect the lack of consequential accountability baked into these programs.

In the charter schools sector, authorizers are empowered to close low-performing or financially unsustainable schools, and they do so with regularity. This is real accountability, and the threat of closure very likely contributes to—perhaps even causes much of—the charter achievement advantage.

What's less clear, once again, are the exact mechanisms. Does the threat of school closure encourage charter schools to improve? Perhaps—and a series of studies from the Fordham Institute and others have found that charter schools tend to embrace a variety of practices associated with improved achievement, from higher teacher expectations to greater teacher diversity to firmer policies around student discipline.<sup>29</sup> On the other hand, it's surely the case that school closures themselves automatically improve the performance of the charter sector, as the worst schools disappear, shifting the bell curve of achievement to the right. Whatever the reason, it's clear that accountability plays a key role in the relative success of charter schools.

## UNRESOLVED TENSIONS IN STANDARDS-BASED REFORM

---

### ACCOUNTABILITY VERSUS CAPACITY BUILDING

The most fateful decision in the history of standards-based reform might have been the move—cemented by NCLB—to place accountability at the heart of the strategy while largely neglecting capacity building; in other words, to assume that the only problem was the lack of will rather than skill. As Robert Pondiscio argues in chapter 5 of this series, that decision was particularly critical when it came to the issue of curriculum. Even those of us who believe in the importance of standards understand that they don't teach themselves, nor do they provide day-to-day guidance to teachers on how to instruct students in an effective, engaging, evidence-based way.

Yet only in recent years have reformers embraced curriculum as a key lever for school improvement, with foundations and even states investing in building high-quality instructional materials and organizations such as EdReports judging them for alignment with rigorous standards. Imagine how much more progress we might have made had we embarked on these efforts twenty years earlier!

Yet that would have been hard to do, since back then states were just developing their standards, and they differed dramatically from one another even as most were of low quality. Only with the creation of the Common Core State Standards was there an opportunity to build a truly national marketplace for curricular materials, which is exactly what has happened in recent years. As high-quality products like Core Knowledge Language Arts and Eureka Math gain market share, we might be returning to the capacity-building effort we ditched so many decades ago. Perhaps fixing teacher preparation and professional development can come next.

It's become clear that states need to show leadership around curriculum and instruction rather than sit back and hope districts make the right decisions on their own. States that have done so over the past twenty-five years—including, at various times, Massachusetts, Tennessee, and Mississippi—have seen improvements in achievement (though, of course, correlation does not equal causation).

### IS THE WHOLE GREATER THAN THE SUM OF ITS PARTS?

As with so much else about this topic, it's hard to know whether there were particular components of standards-based reform that made a bigger difference than others. As explained earlier, seminal studies found that it was “consequential accountability” that led to test score gains in the late 1990s and early 2000s—which meant *some* sort of system to classify schools and *some* legitimate threat that *something* might happen to those deemed low-performing. My vague language is intentional. State policies, especially pre-NCLB, varied greatly, and yet scholars still detected an impact on achievement. We can say, then, that the threat of rating schools as poor and potentially taking action was enough to move the needle—at least when these policies were first introduced.

It's likely, though, that when accountability systems were discovered to be mostly bark and no bite—because state officials were loath to follow through and actually shutter schools—these impacts faded. That brought us to a new stage, when the federal government spent billions of dollars through the School Improvement Grants program to turn around low-performing schools. This was a helping-hand approach rather than tough love, and as discussed earlier, it mostly didn't work.

Nor can we make strong claims about the standards and assessments that are at the heart of standards-based reform. Scholars have failed to detect any difference in achievement in states that had low standards versus high ones or weak tests versus strong ones. As they say, the absence of evidence is not the evidence of absence. It's hard to believe that the quality of standards and assessments does not matter; rather, it's more likely that to drive positive change, demanding expectations and tests must be connected to sophisticated school rating systems; meaningful accountability for results; and capacity-building efforts, like the introduction of high-quality curricular materials, to help students succeed.

The lesson for standards-based reform—and many other reforms as well—is that policy-makers can't view components as items on an à la carte menu. In order to drive improvements, it's all or nothing. Especially in the push for "systemic," coherent reform, the effort is only as strong as its weakest link. If the question is which is most important (standards, assessments, school ratings, consequences, turnaround efforts, or capacity building, especially around curriculum), the correct answer is "all of the above."

## **COMMON STANDARDS VERSUS STUDENT VARIATION**

Other key issues that reformers often swept under the rug were (1) the inevitable conflict between the desire to set a single, high standard for achievement and the undeniable reality that kids come into school with widely varying levels of readiness and may need varying amounts of support and time to reach standard; and (2) that schools and school systems in the United States have historically underserved and undersupported students experiencing poverty and students with lower socioeconomic status.

The standards-based reform movement succeeded in promoting the idea that "all students can learn" and that we must reject the "soft bigotry of low expectations." These are powerful and necessary maxims. But they rub up against the lived experience of educators, who must cope with the reality of classrooms of students who can be as many as seven grade levels apart on the first day of school.<sup>30</sup>

Slogans about "holding schools accountable for results" elide critical questions over the details. Results for which students? All of them? Including the ones who start the school year way above or way below grade level? The embrace of "growth models" in the late NCLB period and under ESSA helped to circle this square. By focusing on progress from one school

year to the next, accountability systems could give schools credit for helping all of their students make gains, no matter where they started on the achievement spectrum.

NCLB had an answer to this question, implicit though it may have been: the sharp focus of NCLB was on helping the lowest-achieving students—who tended to be Black, Hispanic, or low-income, or students with disabilities, or those still learning English—reach basic standards. And as discussed earlier, this focus worked for a time (again mostly in math) as those were the precise groups whose achievement rose the most during the 1990s and 2000s and who were much more likely to graduate from high school in the 2010s. But did this hyper-focus unintentionally incentivize the success and growth of some students over others? And was getting these students to a baseline level of proficiency setting them up for post-secondary success?

### **TESTS AS ACCOUNTABILITY METRICS VERSUS INSTRUCTIONAL TOOLS**

Another key conflict throughout the standards-based reform era was the role of testing. To put it mildly, “high-stakes tests” were not (and are not) popular—with the general public, parents, and especially educators—even though “accountability” in education polls quite well.<sup>31</sup>

The pushback to testing has been significant. Some of that stemmed from how schools responded to the tests—as discussed earlier, by “teaching to the test” or narrowing the curriculum. Some of it related to the Obama-era push to tie teacher evaluations to test scores. Some of it focused on the tests themselves. Making kids sit for annual assessments from grades three through eight ate up precious instructional time. But since the results didn’t come back until months later—even until the next school year—they weren’t of much help to educators. They weren’t “instructionally useful.” Thus, most school districts opted to give students additional standardized tests, such as NWEA’s Measures of Academic Progress, in order to receive real-time information about how students were doing. One study found students spending as many as twenty-five hours a year sitting for tests.<sup>32</sup>

In recent years, some advocates and assessment providers have called for testing systems that can produce both accountability data and instructionally useful information for educators. That’s an understandable impulse, but trade-offs are unavoidable. Some approaches would assess students three times a year, for example—so-called through-year assessments—which might *increase* the testing load and encourage schools to adopt a curriculum closely aligned with the scope and sequence of the tests, for better or worse. Assessments that return results immediately, meanwhile, are by definition not graded by humans, and (so far at least) they can’t test the same higher-order skills that the better state assessments today can. This might encourage a return to low-level teaching of the skill-and-drill variety.

A key issue going forward is whether states will pursue these more instructionally useful assessment systems or simply acknowledge that we need a variety of tests, some to guide instruction and others to generate accountability data, as unpopular as the latter may be.

## LESSONS FOR THE FUTURE

---

What can tomorrow’s policymakers learn from our experience with standards, assessments, and accountability?

- Be clear-eyed about capacity in the system. Some of us wrongly assumed that incentives were the only big problem—that once we put pressure on schools to improve, they would figure out how to help their students meet standards. What standards-based reform revealed, however, was how little capacity existed in many schools. Educators didn’t know how to boost achievement, or they only knew how to do this for some kids in their schools. They didn’t know what curricula to use. And accountability wasn’t generally strong enough to overcome the political incentives operating in the system, especially union politics. Reformers can’t wish realities like these away. Fixing perverse incentives is necessary but not sufficient; capacity building is needed too. And that means states need to take a more muscular role around issues like curriculum and teacher preparation than some of us once imagined.
- Be wary of any reform that is about “all” students (or all schools). Yes, all kids need to learn to read, write, and do math, and virtually all students can reach basic standards. But not all kids need to (or can be) college ready. Reforms that don’t come to terms with the huge variability in kids’ readiness levels, cognitive abilities, and prior achievements will lose popular support and will flounder.
- Don’t take success for granted! Especially in the wake of the awful COVID-19 pandemic and its disastrous impact on our schools, it’s hard not to romanticize the period in the late 1990s and early 2000s when achievement was skyrocketing. What we wouldn’t give to have those test score gains back! Yet the education debate at the time wasn’t full of celebration and confidence, but angst about things not moving quickly enough. What we need to remember is that education happens slowly, year by year, and we need to make sure that policy leaders stay on course over a long period of time. We should fight the urge to look for the “next big thing.” At the current moment, for example, there’s much enthusiasm about universal education savings accounts as new and exciting, in contrast to charter schools, which feel old and dated to some. Yet based on their strong track record, slowly but surely continuing to expand high-quality charter schools may be the best approach to improving student outcomes and expanding parental options. Policymakers, advocates, and philanthropists need to get better at finishing what we started.
- Scholars need new ways to study policy change all the way to the classroom. Thanks in part to the data produced by standards-based reforms, the field of education research has improved markedly in recent decades. Experimental and quasi-experimental designs are much more common, and every day brings important new findings about interventions and their impact on student outcomes. Yet as this chapter demonstrates, we still struggle to follow policy changes all the way down to the classroom. But

that doesn't have to be a given. It's now technically and financially feasible to put cameras and microphones in classrooms nationwide to collect detailed information about teaching and learning. Breakthroughs in artificial intelligence will soon allow us to analyze such data to gain insights about curriculum implementation, effective instructional strategies, grouping practices, student discipline, and much else. The question is whether we will have the political will to make this vision a reality while ensuring safeguards for teacher and student privacy.

The conventional wisdom in some quarters is that standards-based reform in general, and NCLB in particular, didn't work. That conventional wisdom is incorrect. These policies deserve some of the credit for the historically large achievement gains of the 1990s and 2000s and the equally impressive improvements in the high school graduation and college completion rates of more recent years.

But this approach to reform will work much better if it is combined with efforts to boost the knowledge, skills, and confidence of educators on the front lines. Providing high-quality instructional materials is arguably the best way to do that, and it's an effort that states have finally embarked upon. This is still no panacea; the Gordian knot hasn't been sliced through, nor have teachers' unions disappeared, nor have we solved the riddle of how to get fourteen thousand school districts to embrace smart policies and practices. Systemic dysfunction remains. But a recommitment to accountability for results, along with a focus on making classroom instruction more coherent, effective, and equitable, could yield stronger results in the years ahead.

## HESI PRACTITIONER COUNCIL RESPONSES

---

*Essays in this series were reviewed by members of the Hoover Education Success Initiative (HESI) Practitioner Council. For more information about the Practitioner Council and HESI, visit us online at [hoover.org/hesi](http://hoover.org/hesi).*

Mike Petrilli is partly right about the lessons from the past decades of standards-based reform. More research is certainly needed, but we must be smart about how we evaluate the impact of these reforms. In the multilayered, locally controlled US system of K-12 education, trends on state NAEP scores will never yield clear answers, because the action is at the local level.

A recent report from the Council of the Great City Schools (CGCS) points to a very productive way forward. It asks the right research question: what accounts for effective districtwide instructional improvement initiatives, and what role, if any, did state standards play? To answer that question, the CGCS examined the instructional improvement strategies in large districts and compared those that produced increased achievement with those that did not. This work focused on school districts—where key implementation actions occur—and looked

at strategies, behaviors, and tools that formed each district’s improvement strategy. That’s one excellent way to analyze the impact of standards-based reforms.

That approach does not yield a simple succeed/fail conclusion on any one reform initiative. But it does tell us where, how, and why standards are making a difference. And it underscores the need to systematically build the capacity of local schools and districts to dramatically improve the coherence of curriculum and instructional materials; professional learning; and assessments, accountability, and continuous improvement efforts.

—Mike Cohen, former president of Achieve

Petrilli does an excellent job of summarizing the education reform efforts from the late 1980s to the present. There are several points he made that resonate with me and with the experiences I have had leading reform at the district and state levels.

First, the “tight-loose” strategy was doomed to failure. It assumes people would know what to do, what strategies to employ, and how to improve student achievement. In my experience, there simply are not enough leaders who have been trained to do this work in order to effect change for a diverse population of students. Petrilli’s statement, “States need to show leadership around curriculum and instruction rather than sit back and hope districts make the right decisions on their own,” is correct. If people knew what to do, they would be doing it.

Second, there is a need for “coherence, alignment, and building ‘capacity’ in the system to improve teaching and learning.” That needs to be coupled with accountability with a capital A. In my opinion, accountability drives the behaviors you want to see from educators. The old adage “What gets measured gets done” is true. I wholeheartedly agree with Petrilli as he summarizes his article stating, “A recommitment to accountability for results, along with a focus on making classroom instruction more coherent, effective, and equitable, could yield stronger results in the years ahead.” Students’ futures are at stake, and we owe them nothing less.

—Dr. Carey M. Wright, former state superintendent of education for Mississippi

## NOTES

1. See especially Marshall S. Smith and Jennifer O’Day, “Quality and Equality in American Education: Systemic Problems, Systemic Solutions,” in *The Dynamics of Opportunity in America: Evidence and Perspectives*, ed. Irwin Kirsch and Henry Braun (Cham, Switzerland: Springer International Publishing, 2016), 297–358, [https://link.springer.com/chapter/10.1007/978-3-319-25991-8\\_9](https://link.springer.com/chapter/10.1007/978-3-319-25991-8_9); Consortium for Policy Research in Education, “Putting the Pieces Together: Systemic School Reform,” *CPRE Policy Briefs* (1991): 1–10, [https://www.cpre.org/sites/default/files/policybrief/847\\_rb06.pdf](https://www.cpre.org/sites/default/files/policybrief/847_rb06.pdf).
2. See, for example, American Federation of Teachers, “Achieving the Goals of Standards-Based Reform,” last modified 2002, accessed June 26, 2023, <https://www.aft.org/resolution/achieving-goals-standards-based-reform>.

3. David Osborne and Ted Gaebler, *Reinventing Government: How the Entrepreneurial Spirit Is Transforming the Public Sector* (New York: The Penguin Group, 1993).
4. Eric A. Hanushek, "What Matters for Student Achievement," *Education Next* 16, no. 2 (January 13, 2016): 18-26, <https://www.educationnext.org/what-matters-for-student-achievement>.
5. National Governors Association, "Time for Results: The Governors' 1991 Report on Education" (1986).
6. Chester E. Finn Jr., *We Must Take Charge! Our Schools and Our Future* (New York: Free Press, 1991); David Osborne, *Reinventing America's Schools: Creating a 21st Century Education System* (New York: Bloomsbury, 2017).
7. Kati Haycock, "Good Teaching Matters . . . A Lot," *OAH Magazine of History* 13, no. 1 (Fall 1998): 61-63, <http://www.jstor.org/stable/25163257>.
8. See, for example, Achieve, Inc., "1999 National Education Summit Overview" (1999), <https://www.achieve.org/publications/1999-summit-briefing-book>.
9. Paul E. Peterson and Carlos Xabel Lastra-Anadón, "State Standards Rise in Reading, Fall in Math," *Education Next* 10, no. 4 (May 20, 2010), <https://www.educationnext.org/state-standards-rising-in-reading-but-not-in-math>.
10. Sheila Byrd Carmichael, Gabrielle Martino, Kathleen Porter-Magee, and W. Stephen Wilson, "The State of State Standards—and the Common Core—in 2010," Thomas B. Fordham Institute (July 2010), <https://fordhaminstitute.org/national/research/state-state-standards-and-common-core-2010>; American Federation of Teachers, "Assessing Implementation of the Common Core Standards" (2013), [https://www.aft.org/sites/default/files/assessingccss\\_0.pdf](https://www.aft.org/sites/default/files/assessingccss_0.pdf).
11. Dan Goldhaber and Michael DeArmond, "What Does Empirical Research Say about Federal Policy from NCLB to ESSA?," in *Looking Back to Look Forward: Quantitative and Qualitative Reviews of the Past 20 Years of K-12 Education Assessment and Accountability Policy*, US Chamber of Commerce Foundation, March 28, 2023: 1-35, [https://www.uschamberfoundation.org/sites/default/files/USCCF\\_FutureOfDataInK12EducationResearchReport2.pdf](https://www.uschamberfoundation.org/sites/default/files/USCCF_FutureOfDataInK12EducationResearchReport2.pdf).
12. Goldhaber and DeArmond, "What Does Empirical Research Say," 8. The research referenced is as follows: Eric A. Hanushek, Paul E. Peterson, Laura M. Talpey, and Ludger Woessmann, "Long-Run Trends in the US SES-Achievement Gap," National Bureau of Economic Research Working Papers, no. 26764 (2020), <https://doi.org/10.3386/w26764>; Shirin A. Hashim, Thomas J. Kane, Thomas Kelley-Kemple, Mary E. Laski, and Douglas O. Staiger, "Have Income-Based Achievement Gaps Widened or Narrowed?," National Bureau of Economic Research Working Papers, no. 27714 (2020), <https://doi.org/10.3386/w27714>; M. Danish Shakeel and Paul E. Peterson, "A Half Century of Progress in US Student Achievement: Agency and Flynn Effects; Ethnic and SES Differences" (December 3, 2021), <http://dx.doi.org/10.2139/ssrn.3800305>; Sean F. Reardon, "The Widening Academic Achievement Gap between the Rich and the Poor: New Evidence and Possible Explanations," in *Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances*, ed. Greg J. Duncan and Richard J. Murnane (New York: Russell Sage Foundation, 2011): 91-116.
13. Michael J. Petrilli, "Fewer Children Left Behind: Lessons from the Dramatic Achievement Gains of the 1990s and 2000s," Thomas B. Fordham Institute (2019), <https://fordhaminstitute.org/national/research/fewer-children-left-behind>.
14. Douglas N. Harris, Lihan Liu, Nathan Barrett, and Ruoxi Li, "Is the Rise of High School Graduation Rates Real? High-Stakes School Accountability and Strategic Behavior," Annenberg Institute at Brown University, EdWorkingPapers (March 2020), <https://doi.org/10.26300/5jh6-0526>.
15. Martin Carnoy and Susanna Loeb, "Does External Accountability Affect Student Outcomes? A Cross-State Analysis," *Educational Evaluation and Policy Analysis* 24, no. 4 (2002): 305-31, <https://doi.org/10.3102/01623737024004305>; Thomas S. Dee and Brian Jacob, "The Impact of No Child Left Behind on Student Achievement," *Journal of Policy Analysis and Management* 30,

no. 3 (2011): 418–46, [www.jstor.org/stable/23018959](http://www.jstor.org/stable/23018959); Eric A. Hanushek and Margaret E. Raymond, “Does School Accountability Lead to Improved Student Performance?,” *Journal of Policy Analysis and Management* 24, no. 2 (2005): 297–327, <https://doi.org/10.1002/pam.20091>.

16. Manyee Wong, Thomas D. Cook, and Peter M. Steiner, “Adding Design Elements to Improve Time Series Designs: No Child Left Behind as an Example of Causal Pattern-Matching,” *Journal of Research on Educational Effectiveness* 8, no. 2 (2015): 245–79, <https://www.tandfonline.com/doi/full/10.1080/19345747.2013.878011>.

17. Thomas Ahn and Jacob Vigdor, “The Impact of No Child Left Behind’s Accountability Sanctions on School Performance: Regression Discontinuity Evidence from North Carolina,” National Bureau of Economic Research Working Papers, no. 20511 (2014), <https://doi.org/10.3386/w20511>; Dale Ballou and Matthew G. Springer, “Has NCLB Encouraged Educational Triage? Accountability and the Distribution of Achievement Gains,” *Education Finance and Policy* 12, no. 1 (2017): 77–106, <https://eric.ed.gov/?id=EJ1125573>; John M. Krieg, “Which Students Are Left Behind? The Racial Impacts of the No Child Left Behind Act,” *Economics of Education Review* 30, no. 4 (2011): 654–64, <https://doi.org/10.1016/j.econedurev.2011.02.004>; Derek Neal and Diane Whitmore Schanzenbach, “Left Behind by Design: Proficiency Counts and Test-Based Accountability,” *Review of Economics and Statistics* 92, no. 2 (2010): 263–83, <https://doi.org/10.1162/rest.2010.12318>; Randall Reback, “Teaching to the Rating: School Accountability and the Distribution of Student Achievement,” *Journal of Public Economics* 92, no. 5 (2008): 1394–415, <https://doi.org/10.1016/j.jpubeco.2007.05.003>; Matthew G. Springer, “The Influence of an NCLB Accountability Plan on the Distribution of Student Test Score Gains,” *Economics of Education Review* 27, no. 5 (2003): 556–63, <https://doi.org/10.1016/j.econedurev.2007.06.004>.

18. Ozkan Eren, David N. Figlio, Naci H. Mocan, and Orgul Ozturk, “School Accountability, Long-Run Criminal Activity, and Self-Sufficiency,” National Bureau of Economic Research (2023), <https://www.nber.org/papers/w31556>.

19. Laura S. Hamilton, Brian M. Stecher, Julie A. Marsh, Jennifer Sloan McCombs, Abby Robyn, Jennifer Lin Russell, Scott Naftel, and Heather Barney, “Standards-Based Accountability under No Child Left Behind: Experiences of Teachers and Administrators in Three States,” RAND Corporation (2007); Jennifer McMurrer, “Choices, Changes, and Challenges: Curriculum and Instruction in the NCLB Era,” Center on Education Policy (2007); Martin West, “Testing, Learning, and Teaching: The Effects of Test-Based Accountability on Student Achievement and Instructional Time in Core Academic Subjects,” in *Beyond the Basics: Are Math, Reading, and Science Sufficient for a 21st Century Education?*, ed. Chester E. Finn Jr. and Diane Ravitch (Washington, DC: Thomas B. Fordham Institute, 2007): 45–62, [http://edexcellencemedia.net/publications/2007/200707\\_beyondthebasics/Beyond\\_The\\_Basics\\_West.pdf](http://edexcellencemedia.net/publications/2007/200707_beyondthebasics/Beyond_The_Basics_West.pdf).

20. Adam Tyner and Sarah Kabourek, “Social Studies Instruction and Reading Comprehension: Evidence from the Early Childhood Longitudinal Study,” Thomas B. Fordham Institute (2020), <https://fordhaminstitute.org/national/resources/social-studies-instruction-and-reading-comprehension>.

21. See, for example, Joseph Pedulla, Lisa Abrams, George Madaus, Michael Russell, Miguel Ramos, and Jing Miao, “Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers” (2003), <https://www.bc.edu/research/nbetpp/statements/nbr2.pdf>.

22. For an excellent overview of survey data linking accountability policies with instructional practice, see Jane Hannaway and Laura Hamilton, “Performance-Based Accountability Policies: Implications for School and Classroom Practices,” Urban Institute, [https://webarchive.urban.org/UploadedPDF/411779\\_accountability\\_policies.pdf](https://webarchive.urban.org/UploadedPDF/411779_accountability_policies.pdf).

23. Among the rare high-quality studies that followed accountability policies all the way to the classroom were John B. Diamond, “Accountability Policy, School Organization, and Classroom Practice: Partial Recoupling and Educational Opportunity,” *Education and Urban Society* 44,

- no. 2 (2012): 151–82, [https://www.johnbdiamond.com/uploads/6/5/0/7/65073833/\\_4\\_diamond2012.pdf](https://www.johnbdiamond.com/uploads/6/5/0/7/65073833/_4_diamond2012.pdf); and Jennifer Booher-Jennings, “Below the Bubble: ‘Educational Triage’ and the Texas Accountability System,” *American Educational Research Journal* 42, no. 2 (2005): 231–68, <https://journals.sagepub.com/doi/10.3102/00028312042002231>.
24. Lisa Dragoset, Jaime Thomas, Mariesa Herrmann, John Deke, Susanne James-Burdumy, Cheryl Graczewski, Andrea Boyle, Rachel Upton, Courtney Tanenbaum, and Jessica Giffin, “School Improvement Grants: Implementation and Effectiveness,” US Department of Education, NCEE 2017-4013 (2017), <https://files.eric.ed.gov/fulltext/ED572215.pdf>.
25. Thomas Dee, “School Turnarounds: Evidence from the 2009 Stimulus,” National Bureau of Economic Research Working Papers, no. 17990 (2012), <https://doi.org/10.3386/w17990>; Min Sun, Emily K. Penner, and Susanna Loeb, “Resource- and Approach-Driven Multidimensional Change: Three-Year Effects of School Improvement Grants,” *American Educational Research Journal* 54, no. 4 (2017): 607–43, <https://doi.org/10.3102/0002831217695790>; Min Sun, Alec I. Kennedy, and Susanna Loeb, “The Longitudinal Effects of School Improvement Grants,” *Educational Evaluation and Policy Analysis* 43, no. 4 (2021): 647–67, <https://doi.org/10.3102/01623737211012440>; Katharine O. Strunk, Julie A. Marsh, Ayesha K. Hashim, Susan Bush-Mecenas, and Tracey Weinstein, “The Impact of Turnaround Reform on Student Outcomes: Evidence and Insights from the Los Angeles Unified School District,” *Education Finance and Policy* 11, no. 3 (2016): 251–82, [https://doi.org/10.1162/EDFP\\_a\\_00188](https://doi.org/10.1162/EDFP_a_00188).
26. David Griffith and Michael J. Petrilli, “The Case for Urban Charter Schooling,” *National Affairs* (Fall 2020), <https://www.nationalaffairs.com/publications/detail/the-case-for-urban-charter-schooling>; Margaret E. Raymond, James L. Woodworth, Won Fy Lee, and Sally Bacher, “As a Matter of Fact: The National Charter School Study III 2023,” Center for Research on Education Outcomes, Stanford University (2023), <https://ncss3.stanford.edu/wp-content/uploads/2023/06/Credo-NCSS3-Report.pdf>.
27. On Ohio, see David Figlio and Krzysztof Karbownik, “Evaluation of Ohio’s EdChoice Scholarship Program: Selection, Competition, and Performance Effects,” Thomas B. Fordham Institute (July 2016), <https://eric.ed.gov/?id=ED575666>; on Indiana, see R. Joseph Waddington and Mark Berends, “Impact of the Indiana Choice Scholarship Program: Achievement Effects for Students in Upper Elementary and Middle School,” *Journal of Policy Analysis and Management* 37, no. 4 (2018): 783–808, <https://doi.org/10.1002/pam.22086>; on Louisiana, see Atila Abdulkadiroglu, Parag A. Pathak, and Christopher R. Walters, “School Vouchers and Student Achievement: Evidence from the Louisiana Scholarship Program,” National Bureau of Economic Research Working Papers, no. 21839 (December 2015), [https://www.nber.org/system/files/working\\_papers/w21839/revisions/w21839.rev1.pdf](https://www.nber.org/system/files/working_papers/w21839/revisions/w21839.rev1.pdf).
28. Patrick J. Wolf, Brian Kisida, Babette Gutmann, Michael Puma, Nada Eissa, and Lou Rizzo, “School Vouchers and Student Outcomes: Experimental Evidence from Washington, DC,” *Journal of Policy Analysis and Management* 32, no. 2 (2013): 246–70, <http://www.jstor.org/stable/42001528>.
29. Joshua D. Angrist, Parag A. Pathak, and Christopher R. Walters, “Explaining Charter School Effectiveness,” National Bureau of Economic Research Working Papers, no. 17332 (August 2011), [https://www.nber.org/system/files/working\\_papers/w17332/w17332.pdf](https://www.nber.org/system/files/working_papers/w17332/w17332.pdf); Seth Gershenson, “The Power of Expectations in District and Charter Schools,” Thomas B. Fordham Institute (November 2022), <https://fordhaminstitute.org/national/research/the-power-of-expectations-district-charter>; Seth Gershenson, “Student-Teacher Race Match in Charter and Traditional Public Schools,” Thomas B. Fordham Institute (June 2019), <https://fordhaminstitute.org/national/research/student-teacher-race-match-charter-and-traditional-public-schools>.
30. Scott J. Peters, Karen Rambo-Hernandez, Matthew C. Makel, Michael S. Matthews, and Jonathan A. Plucker, “Should Millions of Students Take a Gap Year? Large Numbers of Students

Start the School Year above Grade Level," *Gifted Child Quarterly* 61, no. 3 (2017): 229–38, <https://doi.org/10.1177/0016986217701834>.

31. Christopher Berry and William G. Howell, "Accountability Lost," *Education Next* 8, no. 1 (October 31, 2009), <https://www.educationnext.org/accountability-lost>.

32. Ray Hart, Michael Casserly, Renata Uzzell, Moses Palacios, Amanda Corcoran, and Liz Spurgeon, "Student Testing in America's Great City Schools: An Inventory and Preliminary Analysis," Council of the Great City Schools (October 2015), <https://www.cgcs.org/cms/lib/DC00001581/Centricity/Domain/87/Testing%20Report.pdf>.





The publisher has made this work available under a Creative Commons Attribution-NonCommercial license 4.0. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc/4.0>.

Copyright © 2023 by the Board of Trustees of the Leland Stanford Junior University

The views expressed in this essay are entirely those of the author and do not necessarily reflect the views of the staff, officers, or Board of Overseers of the Hoover Institution.

28 27 26 25 24 23      7 6 5 4 3 2 1

## ABOUT THE AUTHOR

---



### **MICHAEL J. PETRILLI**

Michael J. Petrilli is president of the Thomas B. Fordham Institute, research fellow at the Hoover Institution, executive editor of *Education Next*, and host of the *Education Gadfly Show* podcast. He is the author of *The Diverse Schools Dilemma*, editor of *Education for Upward Mobility*, and coeditor of *How to Educate an American* and *Follow the Science to School*.

---

### **A Nation at Risk + 40**

The modern school-reform movement in the United States was set in motion by the release of the report *A Nation at Risk* in 1983. Countless education policy changes at the local, state, and national levels came as a result. *A Nation at Risk + 40* is a research initiative designed to better understand the impact of these efforts. Each author in this series has gone deep in a key area of school reform, exploring the following questions: *What kinds of reforms have been attempted and why? What is the evidence of their impact? What are the lessons for today's education policymakers?* As the nation's schools work to recover from the effects of the COVID-19 pandemic, this series not only describes the education-reform journey of the past forty years, it also provides timely and research-driven guidance for the future.

*The Hoover Institution gratefully acknowledges Allan B. and Kathy Hubbard, the Daniels Fund, the William and Flora Hewlett Foundation, and the Koret Foundation for their generous support of the Hoover Education Success Initiative and this publication.*

**Hoover Institution, Stanford University**  
434 Galvez Mall  
Stanford, CA 94305-6003  
650-723-1754

**Hoover Institution in Washington**  
1399 New York Avenue NW, Suite 500  
Washington, DC 20005  
202-760-3200

