

A PUBLICATION OF THE HOOVER INSTITUTION

A Nation At Risk:

*A Review
of Progress in
US Public Education*

40

Hoover Institution

November 2023



The Hoover Institution gratefully acknowledges the following individuals and foundations for their generous support of the HOOVER EDUCATION SUCCESS INITIATIVE and this publication:

Allan B. and Kathy Hubbard
Daniels Fund
William and Flora Hewlett Foundation
Koret Foundation

CONTENTS

Introduction: *A Nation at Risk* at Forty **1**
Stephen L. Bowen

1.
The Imperative for High-Quality Pre-K **19**
Deborah Stipek

2.
A Survey of Whole-Child School Reforms **41**
Maria D. Fitzpatrick

3. Strategically Prioritizing Teacher Quality and Quantity Can Reinvigorate
the Teacher Pipeline **61**
Michael Hansen

4.
School Staffing and Teacher Quality **93**
Thomas S. Dee

5.
The Case for Curriculum: Since *A Nation at Risk*, Education Reform Efforts
Have Mostly Stopped at the Classroom Door **109**
Robert Pondiscio

6.
The World Is a Lab: Innovation in Schools after *A Nation at Risk* **125**
Eric Bettinger

7.
Unfulfilled Promise: The Forty-Year Shift from Print to Digital and
Why It Failed to Transform Learning **145**
Tom Vander Ark

8.	Fixing Schools through Finance	161
	<i>Eric A. Hanushek</i>	
9.	From the One Best System to Student-Centered Systems: Lessons from a Half Century of K-12 Governance Reform	181
	<i>Michael T. Hartney</i>	
10.	School Choice Policies in the United States: Retrospect and Prospect	207
	<i>John D. Singleton</i>	
11.	Standards-Based Reform	225
	<i>Michael J. Petrilli</i>	
12.	Lessons from Newark: The Lineage of Modern School Reform and Where We Go Next	247
	<i>Cami Anderson</i>	
	Conclusion: Times Have Changed. The School System? Not So Much.	281
	<i>Margaret E. Raymond</i>	
	About the Authors	295



Introduction

A Nation at Risk *at Forty*

Stephen L. Bowen

Executive Summary

Forty years ago, the release of *A Nation at Risk* led to what we know today as the modern school reform movement. With its calls for increased academic rigor, more productive use of instructional time, more effective teaching, and more impactful leadership, *A Nation at Risk* set in motion policy and practice changes at every level of the education system. The last four decades of school reform can be divided into four eras:

1983–1989: The States Respond The states were the first to respond to the recommendations of *A Nation at Risk*. It is estimated that states launched three thousand reform measures in this era, including changes to graduation requirements, teacher certification and training, and use of instructional time.

1989–2002: Standards and Systems Once it became clear that a series of marginal, disconnected reforms were insufficient, more systemic approaches were attempted. States adopted academic standards, expanded standardized testing, and began tracking school and student performance. Charter schools and other choice options were launched in this era as well.

2002–2015: The NCLB Era With the passage of the No Child Left Behind Act, the federal government took the lead role in education reform, requiring states to take a series of actions related to standards, assessment, accountability, and school improvement. States and districts made some progress in this era, but some elements of the law proved unworkable.

2015–2023: ESSA and COVID-19 The Every Student Succeeds Act replaced No Child Left Behind, moving some reform authority back to the states. School reform efforts were largely halted by the COVID-19 pandemic, which closed schools and had significant impacts on schools and students.

As the nation's schools face challenges in their recovery from COVID-19, there are lessons to be learned from the past forty years of reform, which are explored in the essays in this collection.



By the time then president Ronald Reagan entered the State Dining Room a little after 4 p.m. on April 26, 1983, his day had already been busy. He was scheduled to deliver a speech to Congress on Central America the following day, and he later noted in his diary that he had been “scrambling all day” to prepare.¹ He had started his morning with a briefing by his national security team but was pulled away from speech preparation by several phone calls and a visit from the NATO secretary general, with whom he discussed nuclear missiles in Europe. After lunch, Reagan made still more phone calls, discussed job training programs with the governor of New Hampshire, and met with members of the House Intelligence Committee to discuss Nicaragua.² Indeed, when Reagan finally did step to the microphone, several minutes behind schedule, he lightheartedly blamed his tardiness on “members of the Congress.”³

Reagan was in the State Dining Room to formally receive a report by the National Commission on Excellence in Education, a bipartisan eighteen-member task force empaneled by then education secretary Terrel Bell. Bell tasked the commission with examining “the quality of learning and teaching in the nation’s schools,” a review and analysis made necessary, Bell had said in his instructions to the panel, by “the widespread public perception that something is seriously remiss in our educational system.” After eighteen months of work, including a series of convenings and public hearings as well as the commissioning of more than forty research papers, the commission had produced a report as notable for its brevity—thirty-six pages plus an appendix—as it was for its alarming tone, which was well reflected by its attention-grabbing title: *A Nation at Risk: The Imperative for Educational Reform*.⁴

In his brief remarks in receipt of the report, Reagan thanked the members of the commission and noted they had found that “our educational system is in the grip of a crisis caused by low standards, lack of purpose, ineffective use of resources, and a failure to challenge students to push performance to the boundaries of individual ability.” To Bell’s chagrin, Reagan spent little time describing the commission’s recommendations, preferring instead to highlight elements of his own education reform agenda, which included “tuition tax credits, vouchers, education savings accounts, voluntary school prayer,” and, again to Bell’s dismay, “abolishing the [US] Department of Education.”⁵

After thanking the commission again, Reagan closed his remarks with a characteristically self-deprecating joke about his own 1932 degree from Eureka College being “honorary” and took his leave.⁶ In an interview later that afternoon with *USA Today*, Reagan reiterated his view that “there is a parallel between the Federal involvement in education and the decline in quality over recent years.”⁷ He then wrapped up his day by hosting a dinner for forty guests followed by a movie, but at some point that evening, he made note of the commission’s report in his personal diary. “It’s a great report & lays it on the line,” he wrote.⁸

The president and those around him could be forgiven for thinking that of all the matters he had tackled that day, from foreign policy to national security to job creation, the *least*

impactful might have been the slim report he received about K-12 education. The opposite proved true. As the chapters in this series detail, *A Nation at Risk (ANAR)* not only proved to be a sensation, it led to a wave of school reform efforts across the nation. Indeed, the report would give rise to what we know today as the modern school reform movement. With its calls for increased academic rigor, more productive use of instructional time, more effective teaching, and more impactful leadership, *ANAR* would set in motion policy and practice changes at every level of the education system.

But after forty years, what has been the result? And where do we go from here? The chapters in this series each tackle a distinct area of school reform that emerged in response to the needs described in *ANAR*. Each chapter provides background and context, describes the evidence of impact, draws conclusions, and makes recommendations for policymakers. The goal: to evaluate the evidence and determine what can be learned from four decades of effort to transform the nation's schools.

Forty years on, significant challenges remain. The COVID-19 pandemic has had devastating effects on student learning, and chronic absenteeism remains at alarming rates. Even prior to the pandemic, student achievement, as measured by standardized tests, seemed to have plateaued despite ever-increasing resources—in time, dollars, research, technology, and human capital—being devoted to school reform. How the nation and its schools confront the challenges ahead will be informed in no small part by the reform journey of the past forty years, a journey set in motion by the thunderous response to the commission's report.

1983: A NATION AT RISK ARRIVES

On the day it was released, the *ANAR* report hit, it was said, “with a bang that still echoes.”⁹ In Bell's telling, the response to *ANAR* was “overwhelming.” It made the front page of every major newspaper, was included in evening news broadcasts, and triggered calls and letters to the US Department of Education from across the nation.¹⁰ Sensing that he had a hit on his hands, Reagan took to the road and, with Bell in tow, participated in education events all over the country. More than six million copies of the report would be distributed by the US Government Publishing Office over the course of the following year.¹¹

That *ANAR* triggered a response unlike any federal report before or since was largely attributed to the provocative portrait it painted of an education system failing the nation. “For the first time in the history of our country,” it reported, “the educational skills of one generation will not surpass, will not equal, will not even approach, those of their parents.” The report famously warned of a “rising tide of mediocrity” that was “eroding” the “educational foundations” of the nation. “What was unimaginable a generation ago has begun to occur,” the authors wrote: “others are matching and surpassing our educational attainments.”¹²

What was at stake, the commission claimed, was not simply the future prospects of the nation's learners, but those of the nation itself. Education is a "major foundation for the future strength of our country," the authors wrote, "the foundation for a satisfying life, an enlightened and civil society, a strong economy, and a secure nation."¹³ This linkage, between the quality of the schools and the broader civic, economic, and spiritual health of the nation, was a key sentiment of the report and, as it turned out, an enduring one. K-12 education, by long tradition a local issue despite a small but growing federal role, was now catapulted into the national spotlight.

And then there was the report's tone. In the years following its release, commentators would variously describe *ANAR* as everything from "harsh" and "hard hitting"¹⁴ to "incendiary"¹⁵ and "apocalyptic."¹⁶ It accused the country of "squandering" achievement gains made in the shadow of Sputnik, a comparable wake-up call from a generation prior, claiming that the nation had been "committing an act of unilateral educational disarmament." The report summoned Cold War visions of an "unfriendly foreign power" imposing the system on us against our wills. This we would view, the authors indelibly asserted, as an "act of war."¹⁷

The commission was not content merely to sound the alarm, however. The report contained nearly forty recommendations for change, divided into five broad categories: content, standards and expectations, time, teaching, and leadership and fiscal support. In remarkably short order, policymakers would begin adopting these and other reforms and, perhaps more importantly, would largely embrace *ANAR*'s central argument that prosperity and school quality were closely linked.

Importantly, the authors of *ANAR* did not call for a wholesale reimagining or reinvention of public education. They credited the "American educational system" for rising to past challenges and celebrated the public's long-standing commitment to its schools, which they saw as the "most powerful" tool for reform. What the nation needed to overcome, they wrote, was "weakness of purpose, confusion of vision, underuse of talent and lack of leadership," and they were confident that thoughtful implementation of their proposals would achieve that goal.¹⁸

A NATION AT RISK IN CONTEXT

While *ANAR* proved to be the most high-profile report highlighting the need for school reform, it was far from the only one. In the fall of 1975, for example, Americans had been shocked by reports that scores on the Scholastic Aptitude Test had been dropping for a dozen years. Searching for an explanation, a 1977 report denounced a pervasive school culture where absenteeism is "condoned," where "promotion from one grade to another has become almost automatic," and where "homework has apparently been cut about in half."¹⁹ The National Science Foundation claimed in 1981 that academic rigor at the high school level had dropped to the point that "a third of the nation's school systems required

only one year of mathematics and one year of science.”²⁰ The only graduation requirement in California, it was reported, was two years of physical education.²¹

Even prior to the commission’s report, state leaders had begun to see education as an area of policymaking worthy of their focus. Fully two years before the release of *ANAR*, the Southern Regional Education Board (SREB), an education-focused association of fourteen southern states, had released the report of its own education task force. Like the commission, SREB called for raising standards, increasing salaries and training for teachers, addressing teacher shortages, and investing in school and district leadership.²² Shortly after the release of *ANAR*, the Education Commission of the States put forward its own report, which included a similar series of “action recommendations” developed by a mix of governors, state legislators, business leaders, and educators.²³

ANAR added to this call for reform, but the report’s authors also suggested that something bigger was going on in the country. They described “a growing impatience” with the “shoddiness” so common in life in the United States and a “national sense of frustration” characterized by a “dimming of personal expectations and the fear of losing a shared vision for America.”²⁴

The nation’s malaise, driven by rising crime and urban decay, a lost war in Vietnam, the resignation of a president, and widespread economic turmoil, led to a collapse in the public’s satisfaction with the state of the nation. As the 1970s came to an end, an overwhelming 78 percent of Americans told Gallup pollsters they were “dissatisfied” with the way things were going in the United States.²⁵ Confidence in the schools had fallen as well. From 1974 to 1983, the percentage of respondents giving their local school a grade of A or B dropped from 48 percent to 31 percent—a “negative change of opinion by 25 to 30 million people,” reported pollsters.²⁶

In short, the commission’s impassioned call to restore the “intellectual, moral and spiritual strengths of our people” by reforming the nation’s schools fell on fertile soil. The country was ready for a change. The only question that seemed to remain was who would lead the way.

1983–1989: THE STATES RESPOND

It was *ANAR* itself that helped put governors and other state leaders in the education reform driver’s seat. “What had been a sleepy backwater of state politics mostly delegated to state boards of education and state superintendents,” wrote an observer, “was all of a sudden a critical issue” for state leaders.²⁷ Governors of both parties jumped at the chance to advance education reform as a core priority, and many, including then governor of Arkansas Bill Clinton, became national leaders on the issue. As an indicator of the seriousness of their commitment to school reform, then governor of Kentucky Martha Layne Collins took the novel step of naming herself state secretary of education.²⁸

As a group, the governors wasted little time getting underway. In a paper marking the twentieth anniversary of *ANAR*, Robert Schwartz would write that it was “difficult to overstate the sheer volume of education reform activities unleashed in the states” following the release of *ANAR*:

In a document issued only fifteen months after the Commission’s report, [the Education Commission of the States] reported that over 250 state task forces had been created to study virtually every aspect of education. In that short period, forty-four states raised graduation requirements, forty-five strengthened teacher certification and evaluation requirements, and twenty-seven states adopted measures to increase instructional time.²⁹

In chapter 6 of this series, **Eric Bettinger** describes a series of reforms coming out of *ANAR* that were aimed at rethinking the basic architecture of schooling. Educators and policymakers sought, for instance, to restructure the school day and year, adjust class sizes, and create opportunities for school- and district-led innovations. There was a renewed focus on education leadership and increasing interest in new school models. A “small schools” movement emerged, for example, out of concern for the growing size of US schools and a sense that such schools were “too impersonal to reach every child.”³⁰

Teachers and teaching were a particular focus for the authors of *ANAR* and were seen by state leaders as areas where they could make rapid progress. Two chapters are dedicated to policy proposals in those areas. As **Thomas S. Dee** describes in chapter 4, a number of states raised teacher pay and reformed professional development programs, and several launched career-ladder programs, which had been a specific recommendation of *ANAR*.³¹ States also experimented with new teacher evaluation systems and pay-for-performance models.

Attracting skilled people to the profession was another focus of the report, and as **Michael Hansen** explores in chapter 3, states began reworking their teacher certification and training programs as well as adopting alternative routes to the classroom. The late 1980s, for example, saw the launch of Teach for America, an effort to strengthen the nation’s teacher corps by recruiting recent college graduates and preparing them for the classroom with training and support. The National Board for Professional Teaching Standards emerged in this era as well, as an alternative means of designating highly effective teachers.

As reforms of this type were put into place—it was estimated that three thousand reform measures were enacted in this era—governors quickly faced the formidable challenge of measuring their impact.³² In the early 1980s, few states had any reliable data about how well their students were doing and virtually no data on how their states compared to others. In response to this need, the ever-helpful Secretary Bell produced what came to be known as the “wall chart,” which used scores on the SAT and ACT exams to rank the states on student performance. Governors pushed back almost instantly—the two assessments were used almost exclusively by college-going students, and the percentage of test takers in a given student population varied dramatically from state to state—but the wall chart was an attention-getter, and in the dismay it caused the governors who were sensitive about their rankings,

it helped stimulate a robust discussion about how best to assess academic achievement both within and across the states.

Whether they could measure against them yet or not, the governors also quickly realized that having each state set its own education goals made little sense. To convince taxpayers that increased spending on K-12 education was worth it, states would have to show that progress was being made, both on each state's own terms and relative to the progress of neighboring states. Through the mid-1980s, the governors, working through organizations such as SREB and the National Governors Association, began pushing the idea of national education goals, shared "North Stars" toward which their own state-level efforts could be oriented.³³

As the end of the 1980s approached, though, it became clear that the flurry of state-level reforms launched in the years after *ANAR* had not resulted in the kind of seismic change in outcomes the report had called for. These initial reform efforts, some claimed, had suffered from a top-down, "one-size-fits-all" approach that was disconnected from the realities of the classroom as well as a lack of internal coherence that would have made them work together in thoughtful ways. To truly transform the nation's schools, it was thought, something more—and more systemic—was needed.

1989–2002: STANDARDS AND SYSTEMS

While *ANAR* is seen as having kick-started the modern school reform movement, a handful of events in the late 1980s and early 1990s gave the movement its enduring shape. First, in September 1989, then president George H. W. Bush hosted an education summit for the nation's governors in Charlottesville, Virginia. Bush had campaigned on becoming the "education president," and while he shared his predecessor's sense that K-12 education should be a state-led issue, he was intrigued by the growing interest in national education goals. Establishing these goals seemed like a smart way to signal the administration's commitment to K-12 education without creating a more substantial federal footprint, and their development was the key focus of the convening.³⁴ It would not be until the passage of the Goals 2000 legislation in 1994 that the National Education Goals would be formally established in law, but for the first time, the nation's leaders had put into writing what they hoped public schools would accomplish.

Interestingly enough, the very first of those goals spoke to ensuring that young children were actually prepared to go to school in the first place. Throughout the 1970s and 1980s, interest in preschool programming had grown, driven by a small number of studies showing promising results from high-quality programs.³⁵ As **Deborah Stipek** describes in chapter 1, though *ANAR* made scant mention of early childhood, policymakers were quick to embrace it as an improvement strategy. In the years that followed, states and districts would dramatically expand public support for early childhood education. Enrollment in state pre-K programs, for example, would "jump from about 290,000 to nearly 725,000" by the end of the 1990s.³⁶

With both *ANAR* and the new national goals calling for increased academic rigor for all students, policymakers next faced the challenge of translating these high-level aspirations into something more concrete and measurable. To meet the new national goal of having students demonstrate “competency in challenging subject matter,” for example, clear definitions of both “competency” and “challenging” would need to be developed for each content area.³⁷ An attempt was made to create truly national learning standards—federal grants were provided to fund their development—but this proved to be a disaster. The first federally funded standards out of the gate, in the subject of history, of all things, were released in 1994 and prompted such widespread pushback that the US Senate voted ninety-nine to one to condemn them. The one holdout, it was noted, “wanted an even stronger condemnation.”³⁸ It would ultimately fall to the states to individually develop their own content standards, an approach, as would later become apparent, with pitfalls of its own.

The question of how these standards would actually drive real reform at the level of the individual classroom was answered in part by a landmark paper by Marshall Smith and Jennifer O’Day, released in 1990. In “Systemic School Reform,” they observed that “even when standards are raised and more or better resources are allocated, little lasting change occurs.”³⁹ The reason for this, they posited, was that “the fragmented, complex, multi-layered educational policy system” that governed the nation’s schools undermined the development of coherent strategies to improve teaching and learning.⁴⁰ Smith and O’Day suggested that this coherence could be had by having the disparate elements of the system—everything from curricula, assessments, and instructional materials to teacher training and in-service professional development—aligned around a common set of learning standards. Smith and O’Day’s paper helped form the intellectual basis for the standards-based strategies that would dominate education reform efforts in the years to come.

Others thought differently. Many in the K-12 space saw the layering on of new regulations and top-down mandates, which had accumulated dramatically in the years following *ANAR*, as only making a complicated and bureaucratic system even worse. Rather than try to bring alignment to a vast, disconnected system by somehow getting all of the various players to row in the same direction, why not place key decision-making about teaching and learning where it belonged—inside the schools themselves? Accordingly, policymakers in a number of states enacted laws enabling “bottom-up” reforms, creating opportunities for greater innovation at the local level. As an example of this thinking, a number of states, beginning with Minnesota in 1990, passed legislation creating charter schools. These schools, publicly funded but independently run, would continue to grow in popularity and would prove to be among the most enduring reforms from this era. By the 2020–21 school year, 3.7 million students would attend nearly eight thousand charter schools.⁴¹

Reformers with a more free-market sensibility embraced private-school choice as an alternative approach, believing that school leaders would respond to competitive pressure more readily than additional mandates from above. A number of states and districts, beginning with Milwaukee, launched private-school voucher programs to test this approach. In chapter 10, **John D. Singleton** explores the ways that states adopted and expanded these and other choice options, both public and private. Despite the controversies this approach would inspire in the

years ahead, it proved popular with parents and families. By 2016, just more than 30 percent of the nation’s schoolchildren would be homeschooled or be attending public or private schools of choice.⁴²

The use of education technology was expanding as well. Throughout the 1980s and 1990s, personal computers became a common feature of the nation’s classrooms, and with the advent of the internet and the passage of various state and federal funding mechanisms, billions of dollars were spent on school connectivity. The deployment of all of this technology was typically accompanied by the near-ubiquitous promise that it would utterly and swiftly transform teaching and learning. In chapter 7, **Tom Vander Ark** reviews the rapid evolution of education technology since the era of *ANAR* and judges its impact, drawing lessons for how we might think about coming innovations such as artificial intelligence.

Throughout the 1990s, these various strands of reform were accompanied by growing levels of federal involvement. The 1994 reauthorization of the federal Elementary and Secondary Education Act, titled the Improving America’s Schools Act (IASA), created a federal grant program to support the creation of charter schools and put the federal imprimatur on the standards-based model of systemic reform. For the first time, federal law not only directed states to develop standards and regularly assess students against them, but it also required states to disaggregate this assessment data and report student outcomes for various student subgroups in order to identify achievement gaps. Perhaps more importantly, given where federal law was to go in the years to come, IASA also required states to set performance targets for Title I schools and determine whether they were making “adequate yearly progress” on their journey of improvement. If not, “corrective action” was to be taken to turn them around.⁴³

There was scant enforcement of these provisions from Washington, which was still cautious about overstepping its traditionally arm’s-length relationship with the schools, but a new federal role was established in this era—a shift, as an observer put it, “from the historical focus on ensuring equity for disadvantaged students and impoverished schools to a new commitment to improve academic performance of all students and schools.”⁴⁴

2002–2015: THE NCLB ERA

By the turn of the century, it was clear that the bold ambitions of the post-*ANAR* reform era, including the National Education Goals, were not going to be realized. Then senator Daniel Patrick Moynihan was among those who had predicted as much in 1994, having compared the nascent education goals—US students would be “first in the world in mathematics and science achievement” was among them—to Soviet-era grain production quotas.⁴⁵

The states, to their credit, continued to push forward with key reforms. For example, the number of states establishing academic standards continued to grow. By 2000, forty-nine states had standards in place for English, math, social studies, or science, with forty-four states having adopted standards in all four content areas. Forty-one states were also

assessing student achievement in at least one content area by that point, with twenty testing in all four content areas.⁴⁶ States were expanding choice options as well, with more than fifteen hundred charter schools in place by 2000, serving nearly three hundred forty thousand students.⁴⁷

Even so, there was concern that these strides were not enough. States had adopted standards, but the rigor of those standards was in question. A 2000 analysis by the Thomas B. Fordham Institute found only eight states worthy of an “honors grade” for their state standards, with the remainder “vague, uninspired, timid, full of dubious educational advice, and generally not up to the task.”⁴⁸ States were also struggling with what to do about schools that were continuing to fail their kids. By 2001, forty-one states required the reporting of student outcomes through school “report cards,” but only seventeen states rated or ranked their schools in any way, and only ten identified their most underperforming schools. Few had any power to impose sanctions on failing schools or mandate that they take corrective actions.⁴⁹

Heading into a presidential election year, policymakers questioned whether to double down on the standards-based reform model, with some school choice and other “bottom-up” approaches included as well, or to go in a different direction.

That question was answered with the election of George W. Bush and the enactment, in early 2002, of the No Child Left Behind Act (NCLB). Much of what the new law required—the adoption of rigorous standards, regular assessments of student achievement, a focus on underperforming schools—was directionally consistent with reforms since *ANAR* and with *IASA*. As **Michael J. Petrilli** describes in chapter 11, it was in the area of school accountability that the twelve-hundred-page law broke new ground. Seeming to have learned little from the overambition of the National Education Goals, NCLB established a new goal that all students would demonstrate proficiency in reading and math by 2014. Building on the adequate yearly progress (AYP) concept enshrined in *IASA*, states were to track and report school-level and student group progress. The law also required that states take a series of steps to turn around any schools that were falling behind.

States and districts struggled mightily with these new requirements. At the state level, the identification of schools in need of improvement, combined with the mandated implementation of a series of cascading intervention strategies, quickly strained state capacity. Before long, a majority of states reported that they lacked the staff capacity to implement elements of the law.⁵⁰ As Petrilli also notes, it quickly 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.”

The lack of leadership capacity at all levels quickly emerged as an issue. Leaders at the school and district levels, who were typically “trained to be managers, rather than instructional leaders,” struggled with the large-scale organizational changes school turnaround required.⁵¹ State education agencies, whose primary role had long been to ensure box-checking compliance with state and federal regulations, likewise lacked the expertise to lead the kind of systems

change NCLB envisioned. At the federal level, “shifting expectations, lengthy delays in offering states feedback, and sometimes conflicting advice depending on which federal administration a state official happened to encounter” complicated matters still further.⁵²

Schools and districts were feeling the pressure to improve, however, and reform efforts continued. As **Michael T. Hartney** details in chapter 9, a number of states looked to governance reform as a way to facilitate systemic change. While policymakers of the 1980s and 1990s had explored site-based management and other bottom-up governance models, policymakers in the NCLB era looked to mayoral control or state takeover as a means of overcoming the intransigence of local political actors such as entrenched school boards. In cities like Washington, DC, and post-Hurricane Katrina New Orleans, significant reforms were undertaken under these new governance approaches.

As frustrations with NCLB mounted, the Bush administration began providing flexibility to states by waiving certain provisions of the law. By the time Barack Obama was elected president in 2008, however, it was clear that NCLB was starting to fray at the edges. Thirty-five percent of all schools failed to make AYP that year, a number that would grow to 48 percent by 2011.⁵³ A polarized Washington was in no mood to compromise on revisions to the law, so the Obama administration attempted a workaround by using federal economic recovery funding appropriated in response to the 2008 economic downturn to underwrite the Race to the Top (RttT) initiative. This \$4 billion competitive grant program was designed to entice states into adopting the administration’s preferred reform strategies. Consistent with reform efforts to date, these included improving teacher effectiveness, expanding charter schools, strengthening state standards, and turning around underperforming schools. While it did prove effective in nudging states to adopt these reforms, RttT’s impact on student achievement was questionable, to say the least. In 2016, the US Department of Education, in its own 267-page analysis of the program, came to the definitive conclusion that “student outcomes could be interpreted as providing evidence of a positive effect of [RttT], a negative effect of [RttT], or no effect of [RttT].”⁵⁴

Among the initiatives backed by the Obama administration was the Common Core State Standards, a voluntary set of national standards for English language arts and math, developed by the states. Between 2010 and 2012, all but four states adopted the Common Core standards in the hope that not only would they bring the academic rigor that had been called for since the days of ANAR, but also, in their national reach, they would better facilitate the creation of high-quality instructional materials, student assessments, and professional development resources.⁵⁵ Over time, though, the Common Core standards were sucked into the same vortex of partisanship and ill feeling that came to engulf NCLB. By 2017, eight Common Core states had dropped the new standards entirely, and another twenty-one “had either changed or were in the process of changing” them.⁵⁶

The Obama administration had hoped to have better luck with school turnaround, launching the federal School Improvement Grants (SIG) program, which aimed to “turn around 1,000 schools every year for five years.” Utilizing the existing federal Title I funding structure, the administration flooded almost \$7 billion in funding to SIG from FY 2007 to FY 2014, with \$3.5 billion

allocated in the 2009–10 school year alone. In yet another indication of the immense challenges accompanying turnaround efforts of this kind, a 2017 study by the Institute of Education Sciences found that this staggering investment of resources had “failed to produce meaningful results.”⁵⁷ School district leaders who received SIG funding told researchers that they had struggled with having enough time to implement needed change, given the sheer complexity of putting transformative strategies into place.⁵⁸

In Newark, New Jersey, **Cami Anderson** led one of the most high-profile turnaround efforts in the nation, inheriting a district with plunging enrollment, where “only 40 percent of third-graders could read and write at grade level” and where school buildings dated back to the presidency of Abraham Lincoln. In chapter 12, she depicts the numberless challenges of districtwide turnaround efforts and details the research and best practices that drove the development and implementation of the “One Newark” plan, which anchored subsequent reform efforts there.

As members of Congress came to the table in 2015 to update NCLB, criticism of the law seemed to be coming from every direction. On the political right, there was anger that, in the face of congressional inaction in reauthorizing federal education law, the Obama administration had used waivers and initiatives such as RttT to advance its own reform goals. Obama’s support for charter schools and for the use of student achievement data in teacher evaluation systems won him no fans among the teachers’ unions and those on the political left.⁵⁹ From left, right, and center, an anti-testing backlash erupted. In 2015, for example, 20 percent of New York students in grades three through five opted out of state testing altogether, a protest reflecting deep concerns that under NCLB, school curricula were narrowing, pressures on kids and teachers were mounting, and “the joy of learning” was suffering.⁶⁰

To many, NCLB had been an important step forward despite the pushback, especially with regard to its requirement that schools shine a bright light on achievement gaps by disaggregating student assessment data and reporting outcomes by student group. Also lost in the debate over the law was the fact that student performance had indeed risen, particularly among poor and minority students.⁶¹ The challenge for policymakers was how to address concerns about the law while preserving, and hopefully building on, the modest achievement gains and increased transparency that had been realized.

2015–2023: ESSA AND COVID-19

When later asked how it was that Congress managed to successfully develop a replacement for NCLB when few expected it to, then US representative John Kline said that “the secret sauce was the fact that everybody was fed up with No Child Left Behind.”⁶² The resulting legislation, titled the Every Student Succeeds Act (ESSA), was signed into law in December 2015 and would dial back on the widely acknowledged flaws with NCLB, though key elements of that law remained. The anti-testing movement notwithstanding, the annual assessment provisions stayed in place, as did the requirements to identify achievement gaps by disaggregating

performance data by subgroup. States would still be required to identify and work to improve underperforming schools but were given far more flexibility both in identifying those schools and in the improvement strategies they then implemented. The goal of 100 percent proficiency was dropped, but data and reporting requirements were stepped up in order to provide greater transparency.⁶³

Perhaps most importantly, the law scaled back the federal role generally, providing the states with greater autonomy and increased responsibility for improving student outcomes. The standards-based “triad” that had been in place since the 1990s—state-level learning standards, aligned student assessments and progress determinations, and mandated accountability provisions—remained in place under ESSA, but it was up to the states to make it work in their own local contexts. There was a risk in this, to be sure, and there were concerns that, “left to their own devices, districts and states might not serve all students equally well.”⁶⁴ To help allay these concerns, ESSA required states to deeply engage with their key stakeholders as they developed their new accountability plans.

Away from Washington, there were efforts to learn from what had worked—and had not worked—under NCLB. From the very earliest days of the accountability movement, for example, there had been calls to not lose sight of what Richard Elmore called the “instructional core,” which is the complex interplay, inside the classroom, between teacher, student, and academic content. The lesson of the standards-based reform movement, Elmore wrote, was that it was “impossible to improve student performance without eventually improving the quality of teaching and learning that occurs in classrooms and schools.”⁶⁵ As **Robert Pondiscio** describes in chapter 5, supporting teachers in their classroom practice by strengthening curricular materials and improving instructional pedagogy received renewed attention in these years. A number of states, for instance, have invested heavily in high-quality instructional materials, incentivizing districts to adopt them and providing professional development to support their use.

The broader welfare of students and families has been an area of increasing interest as well. The success of models such as Harlem Children’s Zone, which provides a host of education and support services to the low-income families it serves, led to the adoption of similar “whole-child” approaches elsewhere. In chapter 2, **Maria D. Fitzpatrick** describes how the thinking underlying these approaches—that students are better able to learn if their basic needs are met and that schools are “the most effective places” to address many of those needs—has driven the development of a variety of new approaches such as “community schools,” which streamline and integrate support services for students and families.

What progress was being made by these and other efforts was largely halted in March 2020 by the COVID-19 pandemic. Virtually overnight, schools across the nation were closed to in-person instruction, and school systems struggled to provide learning opportunities and other school-based supports, such as school meals, remotely. Federally required testing and related accountability provisions were suspended, and in some areas of the country, schools remained closed for in-person instruction well into the following year.

Even today, the sheer scale of the pandemic’s impact on students and families is difficult to fathom. Once they were resumed, standardized test scores revealed staggering levels of lost learning, erasing twenty years of achievement gains. For impacted students, this could result in a lifetime of lower earnings.⁶⁶ Though schools and districts have invested heavily in remediation strategies to boost academic recovery, research suggests that absent more fundamental reforms to improve the pace of student learning, there is little chance these efforts will succeed to the degree required.⁶⁷ Worse still, recovery efforts rely on students attending school, and data suggests that levels of chronic absenteeism have skyrocketed. It is estimated that an additional 6.5 million students across the nation have become chronically absent since the pandemic began.⁶⁸

To support states and districts in their recovery efforts, Congress provided unprecedented levels of relief funding. More than \$190 billion was committed to K-12 education through the Elementary and Secondary School Emergency Relief fund (ESSER).⁶⁹ By way of comparison, the largest pot of federal money that flows into K-12 schools each year, Title I, was funded at \$18 billion for the most recent fiscal year.⁷⁰ Reports of ESSER spending by states and districts suggest that recovery dollars are going toward tutoring, extended learning opportunities, and other strategies to address learning loss, but whether they will have the needed effect remains to be seen.⁷¹

Even prior to *ANAR*, debates about school spending have always accompanied efforts to reform the nation’s schools. Education spending jumped in the years immediately following the report’s release, with total K-12 spending in the United States growing from \$128 billion to \$184 billion in the five years between 1982 and 1987—an inflation-adjusted increase of 25 percent.⁷² In constant terms, total K-12 revenues more than doubled in the twenty-five years between 1980 and 2005, as reforms were launched at both the state and federal levels.⁷³

As **Eric A. Hanushek** notes in chapter 8, with the single exception of a drop in revenue following the 2008 recession, “real per-pupil spending (i.e., adjusted for inflation) has risen continuously for more than one hundred years.” The impact of all that spending, Hanushek continues, is “highly variable,” making it nearly impossible to “describe when funds are particularly effective or ineffective.” Demands for increased funding, which have accompanied every effort at reform, will almost certainly emerge again in the fall of 2024, when ESSER relief funds are set to expire.

The nation’s schools will face additional burdens by that point as well, just as they do today. Despite the efforts described above, researchers report there has been “little, if any, progress” in closing pandemic-related learning gaps.⁷⁴ An increasing number of students are seeking mental and behavioral health supports, and schools continue to confront staffing shortages.⁷⁵ School closures drove parents to seek other learning options, and enrollment declines in many districts are leading to school closures and budget shortfalls.⁷⁶ Perhaps most concerning, the country’s confidence in its schools continues to deteriorate. Twenty-six percent of those surveyed by Gallup in 2023 reported “a great deal” or “quite a lot” of confidence in the public schools, down from the 29 percent who shared that view in 2019. In 1983, the year of *ANAR*’s release, 39 percent of respondents signaled that level of confidence in the nation’s schools.⁷⁷

CONCLUSION

Perhaps sensing that they needed to leaven their report's dire tone, ANAR's authors closed it on a hopeful note. "We are the inheritors of a past that gives us every reason to believe that we will succeed," they wrote. "Americans have succeeded before and so we shall again."⁷⁸

They were also clear-eyed about what lay ahead. Enacting their proposed reforms would "take time and unwavering commitment" as well as "widespread, energetic, and dedicated action." The nation's "willingness to take up the challenge, and our resolve to see it through," they wrote, would mean the difference between success and failure.⁷⁹

As the chapters in this series detail, widespread, energetic, and dedicated action is indeed what their remarkable little report inspired. Such action remains necessary today. The nation and its schools face challenges that are as great or greater than those they confronted forty years ago, and there is much to learn in the pages to follow about what was tried and why and what that tells us about the best path forward.

NOTES

1. White House, "White House Diaries, April 26, 1983," Ronald Reagan Presidential Foundation and Library, <https://www.reaganfoundation.org/ronald-reagan/white-house-diaries/diary-entry-04261983>.
2. White House, "Daily Diary of President Reagan, April 26, 1983," Ronald Reagan Presidential Foundation and Library, <https://www.reaganfoundation.org/media/102281/042683.pdf>.
3. Ronald Reagan, "Remarks on Receiving the Final Report of the National Commission on Excellence in Education," Ronald Reagan Presidential Library and Museum, <https://www.reaganlibrary.gov/archives/speech/remarks-receiving-final-report-national-commission-excellence-education>.
4. National Commission on Excellence in Education (NCEE), *A Nation at Risk: The Imperative for Educational Reform* (Washington, DC: US Department of Education, 1983), 1.
5. Terrence Bell, *The Thirteenth Man: A Reagan Cabinet Memoir* (New York: Free Press, 1988), 130.
6. Reagan, "Remarks."
7. "Interview with *USA Today*," Ronald Reagan Presidential Library and Museum, <https://www.reaganlibrary.gov/archives/speech/interview-usa-today>.
8. White House, "Daily Diary."
9. Chester E. Finn Jr., *Troublemaker: A Personal History of School Reform Since Sputnik* (Princeton NJ: Princeton University Press, 2008), 98.
10. Bell, *Thirteenth Man*, 131.
11. James W. Guthrie and Matthew G. Springer, "A Nation at Risk Revisited: Did 'Wrong' Reasoning Result in 'Right' Results? At What Costs?," *Peabody Journal of Education* 79, no. 1 (January 2004): 11.
12. NCEE, *A Nation at Risk*, 5, 11.
13. NCEE, *A Nation at Risk*, 17.
14. Thomas Toch, *In the Name of Excellence: The Struggle to Reform the Nation's Schools, Why It's Failing, and What Should Be Done* (New York: Oxford University Press, 1991), 15.

15. David Gordon, "Introduction," in *A Nation Reformed? American Education 20 Years after A Nation at Risk*, ed. David Gordon (Cambridge, MA: Harvard Education Press, 2003), 1.
16. Edward B. Fiske, "A Nation at a Loss," *New York Times*, April 25, 2008.
17. NCEE, *A Nation at Risk*, 5, 14.
18. NCEE, *A Nation at Risk*, 15, 16.
19. Bill Peterson, "TV, Variety of Social Factors Blamed in Fall of SAT Scores," *Washington Post*, August 24, 1977.
20. Toch, *In the Name of Excellence*, 8.
21. Diane Ravitch, *Left Back: A Century of Battle over School Reform* (New York: Simon & Shuster, 2000), 408.
22. Southern Regional Education Board, "The Need for Quality: A Report to the Southern Regional Education Board by Its Task Force on Higher Education and the Schools" (Atlanta: Southern Regional Education Board, June 1991), 7-9.
23. Task Force on Education for Economic Growth, *Action for Excellence: A Comprehensive Plan to Improve Our Nation's Schools*, Education Commission of the States, June 1983.
24. NCEE, *A Nation at Risk*, 11-12.
25. "Americans' Satisfaction with the Way Things Are Going in the US (Trends)," *Gallup News Service*, March 1-8, 2018, <https://news.gallup.com/poll/230249/americans-satisfaction-things-going-trends.aspx>.
26. Stanley M. Elam, ed., *Gallup Polls of Attitudes Toward Education, 1969-1984. A Topical Summary* (Bloomington, IN: Phi Delta Kappa, 1984).
27. Jal Mehta, *The Allure of Order: High Hopes, Dashed Expectations and the Troubled Quest to Remake American Schooling* (New York: Oxford University Press, 2015), 187.
28. Toch, *In the Name of Excellence*, 19.
29. Robert B. Schwartz, "The Emerging State Leadership Role in Education Reform: Notes of a Participant Observer," in *A Nation Reformed? American Education 20 Years after A Nation at Risk*, ed. David Gordon (Cambridge, MA: Harvard Education Press, 2003), 133.
30. Peter Meyer, "New York City's Small Schools Revolution," *Education Next*, May 12, 2015, <https://www.educationnext.org/new-york-citys-small-schools-revolution>.
31. Pam Grossman, "Teaching: From a Nation at Risk to a Profession at Risk?," in *A Nation Reformed? American Education 20 Years after A Nation at Risk*, ed. David Gordon (Cambridge, MA: Harvard Education Press, 2003), 71.
32. Toch, *In the Name of Excellence*, 36.
33. Maris Vinovkis, *From A Nation at Risk to No Child Left Behind: National Education Goals and the Creation of Federal Education Policy* (New York: Teachers College Press, 2009), 20.
34. Finn, *Troublemaker*, 152-53.
35. Elizabeth Rose, *The Promise of Preschool: From Head Start to Universal Pre-Kindergarten* (New York: Oxford University Press, 2010), 90-91.
36. Rose, *The Promise of Preschool*, 104.
37. Chester Finn Jr., "Agenda-Setters and Duds," in *Carrots, Sticks and the Bully Pulpit: Lessons from a Half-Century of Federal Efforts to Improve America's Schools*, ed. Frederick M. Hess and Andrew P. Kelly (Cambridge, MA: Harvard Education Press, 2011), 219.
38. Ravitch, *Left Back*, 435.
39. Marshall S. Smith and Jennifer O'Day, "Systemic School Reform," *Journal of Education Policy* 5, no. 5 (April 2008), 236.
40. Smith and O'Day, "Systemic School Reform."

41. Libby Stanford, "Charter Schools Find Quiet Support in a World Focused on Private School Choice," *Education Week*, August 21, 2023.
42. Ke Wang, Amy Rathbun, and Lauren Musu, "School Choice in the United States: 2019," NCES 2019-106 (Washington, DC: National Center for Education Statistics, 2019), <https://nces.ed.gov/pubs2019/2019106.pdf>.
43. Patrick McGuinn, "From ESEA to NCLB: The Growth of the Federal Role and the Shift to Accountability," in *The Every Student Succeeds Act: What It Means for Schools, Systems and States*, ed. Frederick Hess and Max Eden (Cambridge, MA: Harvard Education Press, 2017), Kindle.
44. Patrick McGuinn, "No Child Left Behind and the Transformation of Federal Education Policy, 1965-2005" (Lawrence: University Press of Kansas, 2006), 100.
45. George F. Will, "It's All Family," *Baltimore Sun*, February 17, 1994.
46. Craig D. Jerald, "The State of the States 2000," *Education Week*, January 12, 2000, <https://www.edweek.org/education/the-state-of-the-states-2000/2000/01>.
47. National Center for Education Statistics, Common Core of Data, "Public Elementary/Secondary School Universe Survey, 1999-2000 through 2012-13" (Washington, DC: National Center for Education Statistics, 2014), https://nces.ed.gov/programs/digest/d14/tables/dt14_216.90.asp.
48. Chester E. Finn Jr., "Foreword," in *The State of the State Standards 2000*, ed. Chester E. Finn Jr. and Michael Petrilli, The Thomas B. Fordham Foundation, January 2000, vii, <https://fordhaminstitute.org/national/research/state-state-standards-2000>.
49. Thomas S. Dee, "Learning from the Past: School Accountability before ESSA," Hoover Institution, Hoover Education Success Initiative, June 15, 2020.
50. Center on Education Policy, *From the Capital to the Classroom: Year 3 of the No Child Left Behind Act* (Washington, DC: Center for Education Policy, 2005).
51. Paul Manna, *Collision Course: Federal Education Policy Meets State and Local Realities* (Washington, DC: CQ Press, 2011), 87.
52. Manna, *Collision Course*, 158.
53. Jeffery R. Henig, David M. Houston, and Melissa Arnold Lyon, "From NCLB to ESSA: Lessons Learned or Politics Reaffirmed?" in *The Every Student Succeeds Act: What It Means for Schools, Systems and States*, eds. Frederick Hess and Max Eden (Cambridge, MA: Harvard Education Press, 2017), Kindle.
54. Lisa Dragoset, Jaime Thomas, Mariesa Herrmann, John Deke, Susanne James-Burdumy, Cheryl Graczewski, Andrea Boyle, Courtney Tanenbaum, Jessica Giffin, and Rachel Upton, *Race to the Top: Implementation and Relationship to Student Outcomes*, NCEE 2017-4001 (Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education, 2016), xv.
55. Catherine Gewertz, "The Common Core Explained," *Education Week*, September 30, 2015.
56. Amelia Harper, "New York Is Latest State to Tweak, Rebrand Common Core," *K-12 Dive*, October 10, 2017.
57. Emma Brown, "Obama Administration Spent Billions to Fix Failing Schools, and It Didn't Work," *Washington Post*, January 19, 2017.
58. Matt Barnum, "8 Lingering Questions to Confront after the Failure of Obama's \$7B School Turnaround Plan," *74 Million*, January 26, 2017.
59. Henig, Houston, and Lyon, "From NCLB to ESSA," Kindle.
60. Scott Levy, "Making Sense of the Opt-Out Movement," *Education Next* 16, no. 4, 54-64.
61. Chester E. Finn Jr., "School Accountability—Past, Present, and Future," Hoover Institution, Hoover Education Success Initiative, November 19, 2020.

62. Alyson Klein, "How ESSA Passed: The Inside Scoop," in *The Every Student Succeeds Act: What It Means for Schools, Systems and States*, eds. Frederick Hess and Max Eden (Cambridge, MA: Harvard Education Press, 2017), Kindle.
63. Charles Barone, "What ESSA Says: Continuities and Departures," in *The Every Student Succeeds Act: What It Means for Schools, Systems and States*, eds. Frederick Hess and Max Eden (Cambridge, MA: Harvard Education Press, 2017), Kindle.
64. Max Eden and Frederick Hess, "The Long Path to ESSA," *Education Next*, February 15, 2017.
65. Richard Elmore, "Change and Improvement in Educational Reform," in *A Nation Reformed? American Education 20 Years after A Nation at Risk*, ed. David Gordon (Cambridge, MA: Harvard Education Press, 2003), 30.
66. Eric Hanushek, "The Economic Cost of the Pandemic: State by State," Hoover Institution, Hoover Education Success Initiative, January 4, 2023.
67. Margaret Raymond, "The Pace of Recovery after COVID," Center for Research on Educational Outcomes, 2023.
68. Thomas S. Dee, "Higher Chronic Absenteeism Threatens Academic Recovery from the COVID-19 Pandemic," August 2023, OSF Preprints August 10, doi:10.31219/osf.io/bfg3p.
69. "Federal Stimulus Funds for Education," National Council of State Legislators, May 13, 2022, <https://www.ncsl.org/resources/details/federal-stimulus-funds-for-education>.
70. Anne Hyslop and Rebeca Shackelford, "Title I Funding Has Doubled—But Its Most Potent Formula Is Stuck in the Past," *74 Million*, February 21, 2023.
71. "States Leading: How State Education Agencies Are Leveraging the ESSER Set-Aside," Council of Chief State School Officers, <https://learning.ccsso.org/states-leading-how-state-education-agencies-are-leveraging-the-esser-set-aside>.
72. Toch, *In the Name of Excellence*, 37.
73. US Department of Education, National Center for Education Statistics, *Biennial Survey of Education in the United States, 1919–20 through 1949–50; Statistics of State School Systems, 1959–60 and 1969–70; Revenues and Expenditures for Public Elementary and Secondary Education, 1979–80; and Common Core of Data (CCD), "National Public Education Financial Survey," 1989–90 through 2018–19*.
74. Matt Barnum, "Recent School Year Saw Little Academic Recovery, New Study Finds," *Chalkbeat*, July 11, 2023, <https://www.chalkbeat.org/2023/7/11/23787212/nwea-learning-loss-academic-recovery-testing-data-covid>.
75. Kevin Mahnken, "COVID's 'Complicated Picture': Mental Health Worse, Staffing Tight, Enrollment Frozen at Nation's Schools," *74 Million*, May 24, 2023, <https://www.the74million.org/article/staffing-down-enrollment-frozen-federal-data-offer-complicated-picture-of-schools-during-the-pandemic>.
76. Linda Jacobson, "Exclusive Data: As Post-Pandemic Enrollment Lags, Schools Compete for Fewer Students," *74 Million*, May 10, 2023, <https://www.the74million.org/article/exclusive-data-as-post-pandemic-enrollment-lags-schools-compete-for-fewer-students>.
77. Gallup News Service, "Education," June 1–22, 2023, <https://news.gallup.com/poll/1612/education.aspx>.
78. NCEE, *A Nation at Risk*, 34, 36.
79. NCEE, *A Nation at Risk*, 36.



1. The Imperative for High-Quality Pre-K

Deborah Stipek

Executive Summary

This chapter summarizes the history of the development of public preschool in the United States and its current role in education reform. The chapter reviews research on preschool's effectiveness and research related to the many decisions that states need to make after deciding to invest in public preschool. Pros and cons of each decision are discussed, along with specific recommendations for state policymakers.

Since the publication of *A Nation at Risk*, extending school downward to include younger children has become the most expansive and deeply rooted strategy for improving achievement and reducing the achievement gap. The theory of change for public investment in preschool is that enrolling children in an educational program before they enter elementary school can help them develop the skills they need to succeed in kindergarten and can thus set children on a trajectory for school success.

Preschool programming can take many forms, which are explored in depth here, but with regard to student outcomes, the operative phrase is "high quality." Only programs that meet high quality standards have shown long-term effects. States have many policy levers to affect quality, including program licensing standards, teacher and administrator credentialing standards along with ongoing resources to support effective teaching, child learning standards and assessments, program quality monitoring systems, and curriculum guidance.

Policies related to preschool should not be considered in isolation. If teachers in the upper grades do not build on what children learned in preschool, the benefits will fade. States can help sustain benefits by making sure that learning standards, program standards, curriculum, assessments, and teacher credentialing are aligned across preschool and the early grades.

Preschool is not a panacea. It alone will not improve achievement scores and other important child outcomes or reduce that achievement gap. But of the many school improvement

initiatives this series discusses, preschool is among the most studied, and it may have the most potential to move the needle on improving student achievement.

- Preschool has become the norm in children’s educational experience, although controversies still exist about specific goals and how to provide it.
- Policymakers increasingly view preschool as a strategy for launching children on a positive trajectory for academic success and reducing the achievement gap related to economic circumstances that exists before children enter kindergarten.
- Research demonstrates long-term benefits of preschool only for high-quality programs, especially when followed by high-quality and coherent instruction in the early elementary grades.



PRESCHOOL CONTEXT

Since the publication of *A Nation at Risk*, efforts to improve student achievement have taken many forms.¹ Extending school downward to include younger children has become the most expansive and deeply rooted strategy. Nearly every state in the country has made significant investments in public preschool, and preschool has become the norm in children’s education experience.

Preschool is defined as an education program provided to three-to-five-year-olds during the two years before entering kindergarten. Traditionally, preschool has been an education-focused half-day program, which does not meet working parents’ childcare needs. To address this need, preschools are increasingly offering extended-day (“wraparound”) childcare.

Given that children are learning something all day, whether that is intentional or not, the distinction between childcare and preschool is blurry on a practical level. But the distinction has policy implications. Funding for childcare and preschool often comes from different sources, with different family eligibility requirements and standards. This chapter concerns programs (or those parts of programs) that intentionally focus on promoting children’s learning and development, not caretaking.

Unlike the K-12 system, in the United States preschool has no common governance structure, and administrative oversight rests with different levels of government. Children can attend a private, tuition-based preschool; a public preschool funded by the federal government (Head Start) or the state government (state preschool); a program funded by local resources (e.g., local taxes or philanthropy); or a preschool that involves funding from several sources.²

A commitment to educating children before school has its roots in the infant school, designed in the 1820s and 1830s by social reformers to offer free daycare and education emphasizing “moral habit” for the children of poor working parents. In many respects, this is the original “compensatory education” for young children, although it emphasized values and behavior, not preparing children for school success. Soon after infant schools were created for the poor, more affluent parents began sending their young children to private infant schools where the emphasis was on early enrichment and social skills rather than moral reform. But preschools as we currently know them were not widespread in the United States for either poor or more affluent children until the 1960s.³

At about the same time that infant schools were introduced in the United States, kindergarten was developed by German educator Friedrich Froebel. He believed that children should be in school from a young age, and the lessons he designed for children ages three through six emphasized music, nature, stories, and play. The first kindergartens in the United States were created by German immigrants in the mid-1800s. Over time, kindergarten for five-year-olds became a grade of schooling in the United States. In 1965, eighteen states funded public kindergarten. By 2000, all states funded some sort of kindergarten, most of them free to all children who met the age qualifications. Today nearly all (about 98 percent) children attend kindergarten prior to first grade, although kindergarten is not compulsory in most states.⁴ As kindergarten became viewed as “school,” preschool for three-to-four-year-olds (sometimes referred to as prekindergarten or pre-K) took its place as the education program for the year or two before school. Preschool has become the norm in the United States, and although private programs exist, preschool, like kindergarten, is increasingly supported with public funds. The theory of change for public investment in preschool is that an education program before children enter elementary school can help them develop the skills they need to succeed in kindergarten and can thus set them on a trajectory for school success. The next section offers a brief explanation and history of private and public preschools in the United States.

PRIVATE PRESCHOOL

In the decades prior to Head Start, children younger than five years of age were mostly in private and tuition-based preschool programs affiliated with universities or with religious or other community-based organizations. Now there are many community-based nonprofit programs as well as for-profit chains. Families that enroll their children in private preschool have, on average, higher incomes than families that enroll their children in public programs.

Although academic-focused private preschools exist, most focus primarily on social development and socialization—giving children an opportunity to learn how to get along in a setting with peers—as well as general cognitive development, including communication and problem-solving skills. Specific models, such as Montessori and Reggio Emilia, are prominent in the United States. As publicly funded preschools become more available, the proportion of children in community, tuition-based programs is declining.⁵

PUBLIC PRESCHOOL

Specific goals of public preschool vary across funding sources, states, and localities. Kindergarten readiness, however, is ubiquitous, signified by the widespread state adoption of kindergarten readiness tests for children. There is considerable research to support the belief that kindergarten readiness predicts success in school,⁶ which in turn predicts education attainment⁷ and accompanying benefits.⁸ Because some families cannot afford private programs, public funds are needed to make preschool accessible to all children.

Public preschool has had its critics. Some opposition comes from the more conservative view that the government should not be in the business of educating young children; that is parents' responsibility. A comment in 1971 by then president Richard Nixon sums up this argument: "For the Federal Government to plunge headlong financially into supporting child development would commit the vast moral authority of the National Government to the side of communal approaches to child rearing over against the family-centered approach."⁹ Some opposition has also come from skeptics about its effectiveness. A related concern is the isolation of low-income children in a program, especially given evidence that low-income children benefit from being in programs with more economically advantaged peers. Despite such objections, the federal government launched the first expansive public preschool, Head Start, in 1965; state and locally funded programs followed.

Head Start

Head Start was created in 1965 as part of then president Lyndon B. Johnson's war on poverty, and it is one of the only Office of Economic Opportunity programs to survive. It was launched with a clear and ambitious goal: to reverse the cycle of poverty. Head Start was considered compensatory education in the sense that the education program was compensating for inauspicious home and community environments associated with poverty.

The specific program reflects beliefs about the academic and life skills required to escape a life of poverty. Unlike most other publicly funded preschool programs, Head Start is a comprehensive program; in addition to promoting cognitive and social-emotional development, it is designed to promote children's physical and mental health with nutrition and social supports that affect their home environment. Health screenings and provisions for physical health (e.g., dental care) and parent involvement are key components. Head Start thus overlaps in approach with the whole-child reforms described by Maria D. Fitzpatrick in chapter 2 of this series.

Head Start has served an increasing number of children over its history, from fewer than two hundred thousand in the yearlong program in 1967 to over eight hundred fifty thousand in 2020. There has been ongoing tension between the desire to serve more children and at the same time increase quality, both of which add to the program's costs. Quality has been given considerable attention in the past couple of decades. The 2007 reauthorization of Head Start included a revision of the quality standards and a 40 percent set-aside for quality enhancement (used in part to increase teacher salaries). The reauthorization also required that by 2013 half of Head Start teachers have a BA degree and assistant teachers have at least a Child Development Associates credential.

Opposition specific to Head Start has been consistent but has come in different forms. Many conservatives opposed all of Johnson’s Great Society antipoverty programs. Critics concerned about parent choice have suggested that the money could be better spent by providing parents with vouchers to purchase their own childcare or education program for their children.¹⁰ There has also been opposition to the broad whole-child goals of Head Start by those who seek a narrower focus on academic skills. Among supporters of Head Start, the primary concern has been about effectiveness. Despite opposition, the program has survived and in fact grown with bipartisan support.

State and Local Preschool

During the 1980s, states began to develop preschool programs for students from low-income families. State-funded and locally funded preschool goals are typically more limited than Head Start, focusing primarily on kindergarten readiness. Few state preschool programs offer the health and social supports for children and families that Head Start offers.

A total of forty-four states, plus the District of Columbia, now provide at least some state funding for preschool programs, enrolling in 2021 about 29 percent of four-year-olds.¹¹ A number of cities have also used various sources of funding to increase access to preschool. For example, San Antonio, Texas, increased its sales taxes; San Francisco added a rental tax, and Seattle added a property tax. Funding levels vary substantially from state to state, from a low of \$420 (North Dakota) to a high of \$19,228 (Washington, DC) in 2021.¹²

There are several reasons for the growth in state and local preschool programs. First, preschool among middle-class four-year-olds was becoming the norm. Children in families that could not afford tuition-based preschool were therefore at a disadvantage when they entered kindergarten. Second, policymakers were increasingly aware of the large achievement gap related to economic circumstances that existed at kindergarten entry and persisted through school. It was clear that any effort to reduce the achievement gap needed to start early in children’s lives. Third, there were significant concerns about the inadequacy of public education and frustration from failed reform initiatives at the K-12 level. For example, disappointed by the results of \$40 million granted to reform the Philadelphia School System, the Board of the Pew Foundation turned to preschool education with the hope that it would be more amenable to change.¹³ Fourth, there was accumulating research evidence for the short- and long-term benefits of high-quality preschool, including programs implemented at a very large scale. And finally, neuroscientists were demonstrating the significant effects of children’s early experience on the architecture of the brain, making clear that early brain development served as the foundation for future learning and development. Although funding for state and local programs ebbs and flows, the general direction has been toward growth.

A Nation at Risk, published in 1983, played a significant role in the rise in states’ interest in early education. While the report initially stimulated reforms mostly at the high school level, findings at the time related to the benefits of preschool education amplified interest in early childhood education (ECE) as a strategy to improve education outcomes. Many national reports after the publication of *A Nation at Risk* make clear that promoting “school readiness”

became an important element in the school reform agenda. In 1987, the Council of Chief State School Officers included pre-K for low-income four-year-olds in its list of recommendations for improving high school graduation rates.¹⁴ In the six national goals for 2000, created at a national education summit in 1989 under then president George H. W. Bush, “All children will start school ready to learn” was listed first. The National Association of State Boards of Education recommended that elementary schools create early childhood (EC) units for children ages four to eight and develop relationships with community preschools to help coordinate the fragmented services.¹⁵

In the decade following, preschool for economically disadvantaged children was also frequently included in state school reform legislation. By 1989, thirty-one states either funded their own preschool program or contributed funds to increase access to Head Start. Nearly all of these state programs were instituted in the 1980s.¹⁶ The programs were typically administered by state departments of education and were located in school buildings. Some states funded preschool through separate grants, and some included preschool in their school funding formulas.¹⁷

Most of the publicly funded state preschool programs have targeted children from low-income families in a mixed-delivery system, which can include public schools, Head Start, private preschools, and family childcare homes. States vary in whether they offer a program for half or full school days, although the trend has been toward more full-day programs to meet working parents’ needs.

Some of the opposition to state preschool comes from within the ECE field. Programs for infants and toddlers that cannot avail themselves of preschool funding often take a financial hit when the older children move to a public preschool. The problem stems from reimbursement rates being too low for infant and toddler programs to survive with the low child-teacher ratios required. Programs manage financially by including preschool for older children in the mix, which allows higher child-teacher ratios overall. Also, in mixed-delivery programs, private or community-based programs are sometimes at a disadvantage because they lack the resources and the ability to pay teachers the salaries that public programs offer. As a consequence, some people working in community settings view public preschool as an unfair competitor.

Universal Preschool

Most publicly funded preschool programs target children in low-income families. But since the early 1990s, interest in universal programs available to all four-year-olds has increased. This movement was encouraged in part by research finding preschool benefits for nonpoor as well as economically disadvantaged children.¹⁸

Social programs that are universally available are also considered to be politically more robust than programs targeting the poor. A universal program extends preschool opportunities to the working poor—families that are economically fragile but with incomes just above the eligibility cutoff for programs targeting impoverished families—and it relieves middle-class

families of the cost of private programs. It thus develops a stronger constituency. Greenstein suggests furthermore that a program that serves not only the poor but also people significantly above the poverty line and the middle class may lessen the racial imagery of the program.¹⁹

One argument some advocates make for universal prekindergarten (UPK) concerns the value of mixed socioeconomic status (SES). Programs with low-income eligibility requirements isolate the poor. Research evidence shows, however, that children from low-income families make greater gains in programs with more affluent peers than in segregated programs.²⁰ De facto segregation will continue, but children from low-income families are not officially segregated where universal preschool is available.

Georgia (1995), New York (1998), New Jersey (in Abbott districts serving low-income students, 1998), and Oklahoma (1998) were pioneers in the universal preschool movement, committing to making publicly funded preschool open to all children whose parents chose to enroll. Currently, seventeen states have legislation committing the state to universal preschool. Despite claims of “universality,” however, most states that have legislation to support universal preschool do not serve all four-year-olds, partly because of inadequate funding and partly because of parent choice and staff and facilities shortages. In the 2019–20 school year (the last year not impacted by the COVID-19 pandemic), only four states and Washington, DC, had more than 70 percent of their four-year-olds enrolled in state preschool.²¹

The infrastructure for universal preschool programs varies considerably. For example, public preschool in Oklahoma operates through the public school system. In Georgia, the program has been set up with voucher-like subsidies; money from the state is given to parents who can choose from among private and public programs that have been certified by the government. In West Virginia, state funding for public preschool flows to both local education agencies (LEAs) and non-LEAs through County Offices of Education. Most states and cities that have passed legislation to expand preschool to all four-year-olds rely on a combination of funding sources, including federal Head Start and state preschool. The one exception is California, which has used state education funds to expand preschool to all four-year-olds by essentially adding a grade to elementary school (referred to as “transitional kindergarten”).

Objections specific to universal preschool focus mainly on public subsidies for middle-class and affluent families when, critics point out, they have shown they are willing to pay for preschool. Critics have also claimed that the state should not expand preschool education when so many of the country’s public schools are failing.²²

ISSUES AND CONTROVERSIES

Although not uncontested, preschool is deeply rooted in the education landscape; the debates are less about whether public resources should be used to support it than they are about which type of preschool the resources should support (targeted or universal, as

discussed above), where the preschool should reside (in public schools, community programs, or both), and how to ensure high quality. The latter two issues are discussed next.

COMMUNITY VERSUS SCHOOL-BASED PROGRAMS

Advocates of placing preschool in public schools argue that it is the most efficient approach to offering ECE at scale and that it contributes to alignment in curricula and teaching between preschool and the early elementary grades, which can help maintain the benefits of preschool over time. Alignment with kindergarten is difficult when children come from many different preschool contexts, and the public school they enter has no authority over the standards, expectations, curricula, and other policies at the preschool level. Another argument for public schools is that they have an infrastructure, including all the back-office supports (e.g., finance, facilities, human resources) needed. Finally, advocates point out that teachers in preschools that are part of public school systems are generally more educated and better paid and have easier access to special education supports.²³

But public school-based preschool has its critics. Private program providers complain that they are at a financial disadvantage, lacking the purchasing power of a school district. Another concern is that a direct link to K-12 schooling will result in a focus on academics that can divert attention from the more holistic set of goals that most experts in the field of ECE champion. Many critics expect school-based programs to “push down” from higher grades an emphasis on academics using structured, didactic instruction that is developmentally inappropriate for young children. They also worry that K-12 schools are less likely to communicate with and involve parents, a central principle of Head Start and many non-school-based programs. And there is evidence that public school-based programs often have higher child-to-teacher ratios than is recommended for young children.²⁴

For practical reasons, almost every state with a public preschool program uses a mixed-delivery approach, with some classrooms in public schools and others in community-based organizations (CBOs). A mixed-delivery system expands access by using a broad array of existing programs—including Head Start agencies, childcare centers, private schools, faith-based centers, charter schools, and family childcare homes—giving families choice and supporting small businesses. It also saves the state the cost of new facilities and allows it to take advantage of existing staff, enrollment, and organizational structures. Furthermore, including private and community providers increases political support for state programs. Another advantage of including CBOs is that teachers tend to be more diverse and likely to speak the native language of dual-language learners. A study in New York City found that CBOs also offered care for longer hours on average and were more likely to provide mental health services.²⁵

On the downside, depending on how it is organized, there are many challenges associated with coordinating preschool providers that operate in very different settings with different funding structures and often fewer resources. Also, in many states, childcare and preschool systems are overseen by the state’s human services agency, which can create a disconnect between the human services and the education aspects. It is a challenge to effectively braid

and distribute funding as well as maintain quality across a wide variety of settings with different administrative oversight.

Evidence also suggests differences in the quality of CBOs and public school-based preschools. Studies show that preschool teachers in public school settings are generally better educated and better paid.²⁶ A study of five large-scale, mixed-delivery preschool systems (Boston, New York City, Seattle, New Jersey, and West Virginia) showed differences in quality, despite explicit steps to improve equity in quality across settings.²⁷ All five systems implemented similar program standards in both public school and CBO programs. For example, all required lead teachers in both settings to have at least a BA degree, and New Jersey, Seattle, and Boston paid teachers the same in CBOs and public schools. (New York City did so in the years after the study was conducted.) Nevertheless, analyses by Weiland et al. (2022) reveal that children in CBOs, who were disproportionately children of color and children from low-income families, were taught by less educated teachers in all localities and showed lower student gains than their peers in public schools.²⁸ The authors point out, however, that differences in child gains by setting were smaller in New Jersey and Seattle, the two systems in which policies for CBOs and public schools were the most equitable. The Abbott program in New Jersey stands out for its significant efforts to ensure quality in community-based programs. The state requires frequent site visits from master teachers to coach staff, gives districts tools to conduct assessments, and employs university researchers to assess classroom quality and track children's learning.²⁹

To summarize, there are some practical reasons for mixed-delivery systems, but ensuring quality is more challenging than in a more centralized system such as public schools. If a mixed-delivery system is used, considerable attention needs to be given to supporting equity in quality, including pay equity for comparably educated teachers across settings.

DEFINING AND MEASURING QUALITY

There are three primary levers for ensuring quality: (1) state preschool program licensing standards, including teacher-to-child ratios and teacher credentialing requirements; (2) monitoring of programs that have been licensed; and (3) resources for improvement, such as teacher professional development. In many states, monitoring and improving are combined into what is referred to as a Quality Rating and Improvement System (QRIS). Programs are rated along particular dimensions and are typically offered opportunities to improve on those dimensions rated low. In some states, funding levels are based on a program's QRIS ratings. The National Association for the Education of Young Children (NAEYC) also offers accreditation to programs that meet its high standard, although few programs go through the arduous process to receive and retain NAEYC accreditation.

Two categories of quality are included in most state preschool program licensing standards and monitoring systems. The first involves "structural" indices that can be relatively easily regulated and measured, such as requiring particular teaching credentials, teacher-to-child ratios and group size, and aspects of the physical environment considered important for

safety. The second involves “process” variables having to do with the learning environment and the interactions between teachers and children and among children.

There is fair agreement on what constitutes quality regarding teachers’ general interactions with children. Research evidence indicates the value of (1) an overall classroom climate or tone that is emotionally warm, accepting, and supportive; (2) positive, proactive, and consistent classroom management practices that include more affirmation and warmth and fewer disapproving and behavioral controls; (3) educators’ positive, non-conflictual relationships with individual children; and (4) explicit modeling, teaching, and scaffolding of social-emotional skills.^{30,31,32} There is emerging agreement as well that effective teachers need to engage in bias-free and culturally responsive teaching.³³

Defining and measuring quality related to instruction is difficult and controversial for all education programs. But unlike K-12 education, where there is consensus on academic achievement as the primary goal, the goal of preschool is disputed. At the heart of the controversy is how much academic achievement should be a goal and, if it is, what teaching strategies are appropriate to attain it.

What Should Preschoolers Learn?

Most EC educators believe that preschool should attend to multiple domains of development—that is, the whole child. Proponents of focusing on the whole child argue that even learning basic academic skills requires self-confidence and self-regulation skills, such as paying attention, as well as the social skills required to avoid wasting learning time engaged in conflict with peers or the teacher.³⁴

On the other hand, the accountability movement, instantiated in No Child Left Behind legislation, created pressure for preschools to emphasize academic skills. School districts became focused on raising scores on standardized achievement tests, and many believed that starting early would help them achieve that goal. Current evidence that young children are able to develop foundational literacy and math skills supports advocates’ attention to academic skills,³⁵ as does evidence that early literacy and math skills when children enter kindergarten are highly predictive of reading and math achievement in school.³⁶ Although Head Start has from the beginning been committed to supporting the development of the whole child, there have been efforts to focus it more on academic preparation. The George W. Bush administration, for example, proposed shifting Head Start from a comprehensive intervention to a program that focused on language and literacy.³⁷

The debate rests to some degree on a false dichotomy. A focus on academic skills in math, literacy, science, or any other domain does not preclude attention to other dimensions of children’s development. Nonacademic dimensions of the whole child, such as self-regulation and social skills, can be developed even in the context of academic instruction. For example, math activities can involve cooperative learning opportunities that are designed to help children develop such social skills as listening, taking turns, and negotiating. Moreover, EC educators are less critical of the goal of academic achievement if other important dimensions of

development are also stressed and if literacy and math are taught in what EC experts consider developmentally appropriate ways, explained next.

How Should Preschoolers Be Taught?

The debate about *how* preschoolers should be taught is characterized variably as play based or child centered versus teacher directed. In the extreme of play-based/child-centered programs, children choose and initiate most activities, actively exploring and manipulating concrete materials. The teacher may build on children’s activities (“How many blocks do you have in your tower?”), but they are responding to child-initiated activity. In the extreme of teacher-directed instruction, teachers lead children through rote learning exercises, such as counting and identifying letters, and children work individually on tasks that often involve worksheets. People endorsing a more child-centered approach worry about the effects of structured, teacher-directed instruction on children’s motivation and enjoyment of learning. People endorsing a more teacher-directed approach argue that children do not learn foundational academic skills through self-initiated play.

Research suggests that child-initiated play is important for children to develop general problem-solving and social skills, but the development of subject-matter skills requires more intentional teacher guidance. Research suggests that whole-child, child-centered, play-based curricula fail to produce gains in either math or literacy.³⁸ In contrast, several meta-analyses of research have shown that math curricula involving teacher-led activities can have strong effects on math learning and that literacy curricula are modestly successful in boosting literacy achievement.^{39,40,41,42}

Academically focused instruction does not necessarily mean rote learning and worksheets. Literacy, math, and science can be taught to young children in a developmentally appropriate, engaging way. Research suggests that literacy and math are most effectively taught through intentional but playful activities in which the teacher leads children through a learning activity, but children have some discretion in how they achieve the goal of the activity.^{43,44,45} In addition to the benefits of teacher-guided, playful activities, research supports the value of a language-rich environment in which teachers listen to and engage children in conversation and small-group instruction, which is well designed to engage children’s interest in the context of a positive emotional climate.⁴⁶

MEASURING AND SUPPORTING QUALITY

The research is somewhat mixed on how well-structural dimensions of quality, such as teacher-to-child ratios⁴⁷ and teacher credentials,⁴⁸ predict child outcomes and does not pinpoint specific regulations, such as determining whether a 1:12 or 1:10 teacher-to-child ratio is ideal. The evidence is also mixed on the associations between teacher credentials and quality measured by either classroom observations or child outcomes.⁴⁹ There is some evidence that the amount of specific training in child development and ECE may matter more than the level of the credential.⁵⁰ And there is compelling evidence that professional development (PD) can have an impact on the quality of children’s experiences. But not all PD programs are created

equal. The evidence on PD suggests the value of coaches working directly with teachers, focusing on content knowledge and effective pedagogy within an identified domain. Brief workshops are ineffective in promoting lasting changes in instruction.^{51,52,53}

Classroom observations are needed to assess process quality. Those that are typically used provide valid information on general teacher-child interaction, but they do not assess the quality of literacy or math teaching. The field needs to develop assessment instruments that states can use to measure instructional quality in preschools.

CHALLENGES TO QUALITY

The most significant obstacle to quality is cost, and the critical variable is the teacher. Although research does not provide strong guidance on exactly what kind of preparation is ideal, there is good evidence that preparation is important, as is professional development. But both come with costs. Increasing the rigor of training can improve quality, but it will exacerbate the teacher supply problem without commensurate increases in salary. Professional development also requires time on the part of the teachers and funding for the people who do the PD (and ideally for the participants).

Teacher pay affects quality indirectly through its effect on the economic stress teachers experience, their classroom behavior, and ultimately, turnover.⁵⁴ Because preschool teaching is typically relatively low paid, turnover is high. Turnover undermines quality because programs that invest in professional development do not get much of a return on their investment, and high turnover causes children to experience instability, which undermines the quality of the relationships they can develop with their teachers.

The field is also challenged by inequity in quality for low-income and more affluent students. A study of 1,610 preschool sites in New York City found classroom quality to be lower in sites located in poor neighborhoods and in centers serving higher percentages of Black and Latino children.⁵⁵ Even in Georgia, a national leader in universal preschool, state preschool classrooms in low-income and high-minority communities were rated significantly lower in classroom quality.⁵⁶

IMPACT

Public preschool is costly, and policymakers understandably want evidence that it helps them achieve the goal of improving education outcomes. In brief, the findings indicate that preschool can but does not necessarily have both short- and long-term effects for children that also yield economic and social benefits.

The most notable long-term impact study, which has had a significant (perhaps outsized) impact on ECE policy, is the Perry Preschool Project. The program provided very high-quality preschool education to 123 three- and four-year-old African American children.

Findings based on participants and the control group long into adulthood show many long-term benefits of the program, including higher high school graduation rates, higher earnings, and lower arrest rates. Cost-benefit analyses suggest a substantial return, which economist James Heckman concluded is about 8 to 1.⁵⁷ A substantial portion of the return comes from reduced incarcerations. The findings from this one small study have been used to convince many policymakers of the value of investing in preschool. The public programs that have been developed since, however, are not near the level of quality of Perry Preschool, which employed primarily teachers with graduate degrees in ECE and which offered a staff-to-child ratio of 1:6, teacher weekly visits to children's homes, and parent education. It cost \$21,800 per child in 2017 dollars, compared to the \$5,867 spent per child on average in state preschool programs.⁵⁸

Results of the many studies on the impact of preschool cannot be reviewed here. Meta-analyses, however, indicate meaningful positive effects overall on school academic readiness, reductions in grade retention, and special education placement as well as small effects on social-emotional development, especially when this is a specific goal of the program.⁵⁹

Many studies, including a large-scale study of Head Start,⁶⁰ show an initial benefit of preschool, but the benefit fades over the first few years of elementary school. One study in Tennessee found that the control group had higher achievement scores in fourth grade than the children who had been randomly assigned to state preschool.⁶¹ These findings raised questions about the quality of the Tennessee state preschool program and what early educational experiences the control group had. But along with the many studies that have documented the fading of preschool benefits in elementary school, the Tennessee study makes clear that simply implementing a large-scale preschool program does not guarantee the desired effects on children.

There are several possible explanations for the fade-out effect. One explanation, supported by research, is that kindergarten teachers often repeat material that children had already learned in preschool,^{62,63} allowing children who did not have preschool to catch up with those who did. Another is that teachers in the early grades do not differentiate instruction for children with varying skill levels, which impedes children's opportunities to continue to exhibit growth.⁶⁴ Both explanations suggest that preschool is more likely to yield long-term academic benefits if it is followed with instruction that builds on initial gains.

Although there are only a few studies of long-term effects, there is evidence that preschool implemented at scale can have significant effects into adulthood. A random assignment study in Boston found increases in high school graduation, SAT scores, and college attendance and a decrease in juvenile incarceration, especially for boys, but no detectable impact on state achievement test scores.⁶⁵ A recent study of the Tulsa preschool likewise found effects on college enrollment.⁶⁶ Reviews of studies that assessed long-term effects, including some studies of Head Start participants, describe increases in college enrollment and decreases in incarceration rates and teen pregnancy.⁶⁷

A task force created by the Brookings Institution, which included individuals with very different perspectives on the value of state preschool, worked through all of the evidence on state-funded preschool to produce a consensus summary of impact.⁶⁸ Following is a summary of the task force’s conclusions:

- Studies of different groups of preschoolers often find greater improvement in learning at the end of the pre-K year for economically disadvantaged children and dual-language learners than for more advantaged and English-proficient children.
- Pre-K programs are not all equally effective. Several effectiveness factors may be at work in the most successful programs. One such factor supporting early learning is a well-implemented, evidence-based curriculum. Coaching for teachers as well as efforts to promote orderly but active classrooms may also be helpful.
- Children’s early learning trajectories depend on the quality of their learning experiences not only before and during their pre-K year but also following it. Classroom experiences early in elementary school can serve as charging stations for sustaining and amplifying pre-K learning gains. One good bet for powering up later learning is elementary school classrooms that provide individualization and differentiation in instructional content and strategies.
- Convincing evidence shows that children attending a diverse array of state and school district pre-K programs are more ready for school at the end of their pre-K year than children who do not attend pre-K. Improvements in academic areas such as literacy and numeracy are most common; the smaller number of studies of social-emotional and self-regulatory development generally show more modest improvements in those areas.
- Convincing evidence on the longer-term impacts of scaled-up pre-K programs on academic outcomes and school progress is sparse, precluding broad conclusions. The evidence that does exist often shows that improvements in learning induced prior to kindergarten are detectable during elementary school, but studies also reveal null or negative longer-term impacts for some programs.

ECONOMIC RETURN TO PRESCHOOL

Some impacts on individuals (improved school readiness, higher achievement, increased high school graduation, higher education attainment, higher earnings) bring social benefits, such as increased taxes, reduced use of welfare, and improved health. Other public-sector benefits come in the form of reduced child abuse and neglect and reduced incarceration. Public school savings are found in reduced grade retention and reduced special education use.

Because the large-scale programs are relatively recent, most cost-benefit analyses project benefits from shorter-term impacts. For example, a study of the Chicago Child-Parent Centers (CPC) used data from subjects at age twenty-six on special education use, grade retention, juvenile and adult crime, and adult earnings to project impact beyond age twenty-six, such

as earnings, taxes on earnings, incarceration, depression, smoking, and substance abuse. The net economic benefits to society were estimated to reach almost \$97,000 per child (in 2016 dollars), a return of nearly \$11 for every \$1 invested.⁶⁹ Looking across cost-benefit studies, including Chicago, Tulsa, and Head Start, estimates show a more realistic return of between 2:1 and 4:1; however, such analyses are based on many assumptions for which there is not total agreement. Another caveat is that the conclusions are based on relatively high-cost, high-quality preschool programs and are not likely to apply to programs of lower quality.⁷⁰

CONCLUSION AND RECOMMENDATIONS

Publicly supported preschool began with the goal of giving children living in poverty a “head start” in school, but over time it became a strategy for reducing the achievement gap and improving academic achievement. Will public preschool close the achievement gap? No. There are many other factors associated with poverty—including poor nutrition, crowded and unstable housing, poor medical care, and stress—that have significant effects on children’s learning. But access to high-quality preschool can give children from low-income families a fairer chance of succeeding in school, and the research on long-term impact suggests it can alter life trajectories. The operative term is “high-quality.” All of the research on long-term benefits with economic returns to individuals and society are based on programs that employed well-trained and well-supported staff.

This chapter has outlined a number of issues that need to be addressed after the decision to invest in public preschool has been made, such as whether to support a targeted or universal program, whether to implement a mixed-delivery or school-based delivery system, and how to define, measure, and ensure quality. Below are a few recommendations related to specific choices.

Clearly articulated goals are important. The goals guide program development as well as the strategies and measures used to assess program quality, support program improvement, and track progress in achieving the goals.

Achieving consensus on the goals is no easy task, because there are strong differences of opinion. But many of the differences rest on false dichotomies. Even if academic skills are the primary goal, all dimensions of children’s development play a role. Furthermore, developmentally appropriate, teacher-guided instruction can be playful, and time can be put aside for children to play freely and explore.

Quality is the primary consideration in any preschool system. Only preschool programs that meet high quality standards have shown long-term effects. States have many levers to affect policy, including program licensing standards, teacher and administrator credentialing standards, and ongoing resources to support effective teaching, child learning standards and assessments, quality monitoring systems, and curriculum guidance. These policies need to align with the articulated goals of preschool and with each other. For example, teacher

credentialing requirements and ongoing supports for teachers should prepare teachers to help children achieve the state learning standards.

Of the many dimensions affecting quality that these policy levers can be used to promote, the following are particularly important:

- Adult-to-child ratios that allow adults to form close, caring relationships with children
- Preparation and ongoing support for teachers in developmentally appropriate, differentiated instruction, including developing a safe, secure, and inclusive environment for all students
- Curricula that offer sufficient and developmentally appropriate attention to language, literacy, math, and social-emotional development as well as other important dimensions of child development such as creativity and motor development
- An assessment system that can be used to guide teacher decisions at the classroom level and track children’s progress related to the standards

Depending on the approach to be used, different policy considerations apply.

- If the decision is made to use public funds only for children in low-income families:
 - Because children from low-income families learn more on average in a mixed-SES setting, avoid isolating the poor by offering hybrid programs in which some children are subsidized by the state and some families pay tuition, or implement a sliding scale.
- If universal preschool is implemented:
 - Ensure that community-based providers for infants and toddlers are reimbursed enough to cover costs when they lose older children.
- If a school-based delivery system is employed:
 - Ensure that programs are developmentally appropriate for young children and that teachers and principals have training in ECE.
- If a mixed-delivery system is employed:
 - Establish strong program standards across school-based and CBO settings, and implement other policies (e.g., teacher pay equity) to support equity in quality. Findings suggest that even in states where efforts have been made to equalize quality, extra resources may be needed for CBO programs.

What happens after preschool affects the long-term benefits of preschool. Policies related to preschool should not be considered in isolation. If teachers in the later grades do not build on what children learned in preschool, the benefits will fade. Districts need to reflect on the nature and organization of instruction through the early grades, joining the movement expanding throughout the United States to develop greater P-3 alignment (i.e., alignment from preschool to third grade). States can help by making sure that learning standards, program standards, curricula, assessments, and teacher credentialing are aligned across preschool and the early grades. They also need to ensure strong connections among state agencies overseeing preschool and K-12 education.

Streamline connections between preschool and childcare. Working parents need childcare. The different funding sources and eligibility requirements cause difficulties for both programs and parents. States need to streamline funding and requirements for preschool and childcare as much as possible, and they need to facilitate wraparound childcare for preschool programs.

The politics of pre-K can be complicated to navigate. Many constituencies are affected by decisions related to state-funded preschool, including state agencies that oversee its implementation, community-based programs that are affected by whether and how they are included, higher education where teachers are prepared, school districts, Head Start, the people who staff the programs, and parents. All policy decisions should include their voices.

Preschool is not a panacea. It alone will not improve achievement scores and other important child outcomes. But of the many school improvement initiatives this series discusses, preschool is among the most studied and may have the most potential to move the needle.

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.

Deborah Stipek reflects well on the promises and challenges of preschool as an important policy decision and strategy for leaders who want to positively impact our citizens' outcomes. As demonstrated, there are wide-ranging differences in the quality and approaches to preschool. When policymakers meaningfully invest in high-quality early childcare, especially for our most vulnerable families, the benefits to our communities are clear.

The greatest challenge from my experience as the founder of a bipartisan early childhood caucus in North Carolina is the inability of families to pay the costs required to meet the measures of high-quality preschool that research has documented so well. This fiscal challenge,

along with an increased understanding of the relationship between early childhood and our workforce, offers an opportunity to engage business leaders and policymakers in innovative approaches to strengthen this important strategy in improving our collective future.

—Rep. Ashton Clemmons, North Carolina General Assembly

Preschool, whether universal or targeted (for lower-income families), or whether community-based or school-based, shows tremendous long-term impacts and returns. Certainly, as the piece highlights, there have been a few studies that show that the effects of preschool “fade” over time, but the vast majority of the larger, more longitudinal studies show otherwise. Preschool programs—especially those determined to be of quality—often have a substantial long-term impact on future high school graduation rates, college-going rates, and incarceration rates, to name a few.

One of the strongest arguments to be made—highlighting the return on investment of preschool—was found in two places in this piece. One study (Perry Preschool Project) found a return on investment of 8:1. Another study (Chicago Child-Parent Centers) resulted in a 6:1 return, although others later estimated the return could be somewhere between 2:1 and 4:1. Regardless, there is sufficient longitudinal evidence that there is a significant return on investment of preschool—both economically (a return of between 2:1 and 8:1) and in terms of personal performance and success (education and employment).

For these reasons, I have no concerns about moving forward with this as a major area of reform, but I would advise lawmakers to continue to look at best practices and the research related to quality, access (universal or targeted), and model (community or school based) to determine the most appropriate policy for each state, considering financial and political context.

—Chad E. Gestson, EdD, Arizona Institute for Education and the Economy, Northern Arizona University

NOTES

1. National Commission on Excellence in Education (NCEE), *A Nation at Risk: The Imperative for Educational Reform* (Washington, DC: US Department of Education, 1983).
2. There are other public funding sources used in a relatively small number of programs. For example, school districts are allowed to use federal Title I dollars for preschool, and some school districts use some of their public education funds for preschool.
3. Emily D. Cahan, “Past Caring: A History of US Preschool Care and Education for the Poor, 1820–1965,” National Center for Children in Poverty, Mailman School of Public Health, Columbia University, 1989.
4. Sheila B. Kamerman and Shirley Gatenio-Gabel, “Early Childhood Education and Care in the United States: An Overview of the Current Policy Picture,” *International Journal of Child Care and Education Policy* 1, no. 1 (May 2007): 23–34.
5. Kevin McElrath, Kurt Bauman, and Erik Schmidt, “Preschool Enrollment in the United States: 2005 to 2019,” Working Paper no. SEHSD-WP2021-25, US Census Bureau (November 2021), www.census.gov/content/dam/Census/newsroom/press-kits/2021/paa/paa-2021-presentation-preschool-enrollment-in-the-united-states.pdf.

6. Hanover Research, "Early Skill and Predictors of Academic Success," Hanover Research Report, Arlington, VA (November 2016).
7. Greg J. Duncan and Katherine Magnuson, "The Nature and Impact of Early Achievement Skills, Attention Skills, and Behavior Problems," 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), 47-69.
8. Dana Mitra, "Pennsylvania's Best Investment: The Social and Economic Benefits of Public Education," Education Law Center Report (June 2011), https://www.elc-pa.org/wp-content/uploads/2011/06/BestInvestment_Full_Report_6.27.11.pdf.
9. Edward F. Zigler and Susan Muenchow, *Head Start: The Inside Story of America's Most Successful Educational Experiment* (New York: Basic Books, 1992), 146.
10. Edward F. Zigler and Sally Styfco, *The Hidden History of Head Start* (Oxford, UK: Oxford University Press, 2010).
11. "Percentage of Pre-school Enrollment in the United States in 2021, by Age Group and Type of Program," Statista, 2022, <https://www.statista.com/statistics/315074/percentage-of-pre-school-enrollment-in-the-us-by-age-and-type>.
12. Allison H. Friedman-Krauss, W. Steven Barnett, Karin A. Garver, Katherine S. Hodges, G. G. Weisenfeld, Beth Ann Gardiner, and Tracy Merriman Jost, "The State of Preschool 2021: State Preschool Yearbook," New Brunswick, NJ: National Institute for Early Education Research (2022), https://nieer.org/wp-content/uploads/2022/09/YB2021_Full_Report.pdf.
13. Elizabeth Rose, *The Promise of Preschool: From Head Start to Universal Pre-Kindergarten* (Oxford, UK: Oxford University Press, 2010).
14. Council of Chief State School Officers, "EC and Family Education—Foundations for Success." Reprinted in the US House Committee on Education and Labor, Hearings on Child Care, 183-85, 1988.
15. National Association of State Boards of Education, "Right from the Start: The Report of the NASBE Task Force on ECE," Alexandria, VA, 1988.
16. Andrew J. Mashburn and Robert C. Pianta, "Opportunity in Early Education: Improving Teacher-Child Interactions and Child Outcomes," in *Childhood Programs and Practices in the First Decade of Life: A Human Capital Integration*, ed. Arthur J. Reynolds, Arthur J. Rolnick, Michelle M. Englund, and Judy A. Temple (Cambridge, UK: Cambridge University Press, 2010), 243-65.
17. Rose, *Promise of Preschool*.
18. William Gormley, Ted Gayer, Deborah Phillips, and Brittany Dawson, "The Effects of Universal Pre-K on Cognitive Development," *Developmental Psychology* 41 (2005): 872-84.
19. Robert Greenstein, "Targeting, Universalism, and Other Factors Affecting Social Programs' Political Strength," Washington, DC: Brookings Institution (2022), <https://www.brookings.edu/articles/targeting-universalism-and-other-factors-affecting-social-programs-political-strength>.
20. Portia Miller, Elizabeth Votruba-Drzal, Meghan McQuiggan, and Alyssa Shaw, "Pre-K Classroom-Economic Composition and Children's Early Academic Development," *Journal of Educational Psychology* 109, no. 2 (2017): 149-65.
21. Libby Stanford, "Which States Offer Universal Preschool? It's More Complicated Than You Might Think," *Education Week*, January 25, 2023. <https://www.edweek.org/teaching-learning/which-states-offer-universal-pre-k-its-more-complicated-than-you-might-think/2023/01#:~:text=At%20its%20simplest%2C%20universal%20pre,year%2Dolds%20who%20can%20participate>.
22. Darcy Olsen, "Universal Preschool Is No Golden Ticket: Why Government Should Not Enter the Preschool Business," Cato Institute Policy Analysis, 333 (February 1999), <https://www.cato.org/sites/cato.org/files/pubs/pdf/pa333.pdf>.

23. Walter Gilliam, "Preschool Programs Should be Coordinated in Public Schools with Supports from Head Start and Child Care," in *The Preschool Debates: Current Controversies & Issues*, ed. Edward Ed Zigler, Walter S. Gilliam, and W. Steven Barnett (Baltimore: Paul Brookes Publishing, 2011), 120-26.
24. Gilliam, "Preschool Programs."
25. Bruce Fuller, Talia Leibovitz, Declan Chin, Kaitlyn Du, Naomi Garcia, and Yoonjeon Kim, "Equity and Institutions—Distributing Preschool Quality in New York City." Report from the University of California, Berkeley, July 2020.
26. Dan Bellm, Alice Burton, Marcy Whitebook, Linda Broatch, and Marci P. Young, *Inside the Pre-K Classroom: A Study of Staffing and Stability in State-Funded Prekindergarten Programs* (Washington, DC: Center for the Child Care Workforce, 2002).
27. Karin Garver, G. G. Weisenfeld, Lori Connors-Tadros, Katherine Hodges, Hanna Melnick, and Sara Plasencia, "State Preschool in a Mixed Delivery System: Lessons from Five States," Learning Policy Institute (March 2023).
28. Christina Weiland, Meghan McCormick, Jennifer Duer, Allison H. Friedman-Krauss, Mirjana Pralica, Samantha Xia, Milagros Nores, and Shira Mattera, "Mixed-Delivery Public Prekindergarten: Differences in Demographics, Quality, and Children's Gains in Community-Based versus Public School Programs across Five Large-Scale Systems," EdWorkingPaper 22-651, Annenberg Institute at Brown University, 2022, <https://doi.org/10.26300/pncz-2233>.
29. Weiland et al., "Mixed-Delivery Public Prekindergarten."
30. Dale Farran, "Characteristics of Pre-Kindergarten Programs That Drive Positive Outcomes," in *Issues in Pre-Kindergarten Programs and Policy*, ed. Kenneth A. Dodge (Washington, DC: Brookings Institution, 2017), 45-49.
31. Patricia A. Jennings and Mark T. Greenberg, "The Prosocial Classroom: Teacher Social and Emotional Competence in Relation to Student and Classroom Outcomes," *Review of Educational Research* 79, no. 1 (2009): 491-92, <https://doi.org/10.3102/0034654308325693>.
32. Mashburn and Pianta, "Opportunity in Early Education."
33. Deborah Phillips, Anna Johnson, and Iheoma Iruka, "Early Care and Education Settings as Contexts for Socialization: New Directions for Quality Assessment," *Child Development Perspectives* 16 (2022): 127-33, <https://doi.org/10.1111/cdep.12460>.
34. Sheila Bishop-Josef and Edward Zigler, "The Cognitive/Academic Emphasis versus Whole Child Approach: The 50-Year Debate," in *The Preschool Debates: Current Controversies & Issues*, eds. Edward Zigler, Walter Gilliam, and Samuel Barnett (Baltimore: Paul Brookes Publishing, 2011), 83-88.
35. National Research Council, *Eager to Learn: Educating Our Preschoolers* (Washington, DC: National Academies Press, 2001).
36. Greg J. Duncan, "The Importance of Kindergarten-Entry Academic Skills," in *The Preschool Debates: Current Controversies & Issues*, ed. Edward Zigler, Walter Gilliam, and Samuel Barnett (Baltimore: Paul Brookes Publishing, 2011), 89-93.
37. Cybele C. Raver and Edward Zigler, "Public Policy Viewpoint: Another Step Back? Assessing Readiness in Head Start," *Young Children* 59 no. 1 (January 2004): 58-63.
38. Preschool Curriculum Evaluation Research Consortium, "Effects of Preschool Curriculum Programs on School Readiness (NCER 2008-2009)," Washington, DC: National Center for Education Research, Institute of Education Sciences, US Department of Education, 2008; Washington, DC: US Government Publishing Office.
39. Bette Chambers, Alain C.K. Cheung, and Robert E. Slavin, "Literacy and Language Outcomes of Comprehensive and Developmental-Constructivist Approaches to Early Childhood Education: A Systematic Review," *Educational Research Review* 18 (May 2016): 88-111, <https://doi.org/10.1016/j.edurev.2016.03.003>.

40. Greg J. Duncan, Jade M. Jenkins, Anamarie Auger Whitekar, Margaret Burchinal, Thurston Domina, and Marianne Bitler, "Boosting School Readiness with Preschool Curricula," Working Paper, Irvine Network for Interventions in Development, University of California, Irvine, March 2015.
41. Tutrang Nguyen, Jade Marcus Jenkins, and Anamarie Auger Whitaker, "Are Content-Specific Curricula Differentially Effective in Head Start or State Prekindergarten Classrooms?," *AERA Open* 4, no. 2 (June 2018): 1-17, <https://doi.org/10.1177/2332858418784283>.
42. Aubrey H. Wang, Janine M. Firmender, Joshua R. Power, and James P. Byrnes, "Understanding the Program Effectiveness of Early Mathematics Interventions for Prekindergarten and Kindergarten Environments: A Meta-Analytic Review," *Early Education and Development* 27, no. 5 (January 2016): 692-713.
43. Kayleigh Skene, Christine M. O'Farrelly, Elizabeth M. Byrne, Natalie Kirby, Eloise C. Stevens, and Paul G. Ramchandani, "Can Guidance during Play Enhance Children's Learning and Development in Educational Contexts? A Systematic Review and Meta-Analysis," *Child Development* 93, no. 4 (January 2022): 1162-80.
44. Deborah Stipek, "Playful Math Instruction in the Context of Standards and Accountability," *Young Children* 72, no. 3 (July 2017): 8-13. Reprinted in *Serious Fun: Guiding Play to Extend Children's Learning*. National Association for the Education of Young Children.
45. Deena Skolnick Weisberg, Kathy Hirsh-Pasek, and Roberta Michnick Golinkoff, "Guided Play: Where Curricular Goals Meet a Playful Pedagogy," *Mind, Brain, and Education* 7, no. 2 (May 2013): 104-12.
46. Farran, "Characteristics of Pre-Kindergarten Programs That Drive Positive Outcomes."
47. Nina T. Dalgaard, Anja Bondebjerg, Rasmus Klokke, Bjørn C.A. Viinholt, and Jens Dietrichson, "Adult/Child Ratio and Group Size in Early Childhood Education or Care to Promote the Development of Children Aged 0-5 Years: A Systematic Review," *Campbell Systematic Reviews* 18, no. 2 (May 2022): e1239.
48. Edward Ed Zigler, Walter S. Gilliam, and W. Barnett, eds., *The Preschool Debates: Current Controversies & Issues* (Baltimore: Paul Brookes Publishing, 2011).
49. Diane M. Early, Kelly L. Maxwell, Margaret Burchinal, Soumya Alva, Randall H. Bender, Donna Bryant, Karen Cai et al., "Teachers' Education, Classroom Quality, and Young Children's Academic Skills: Results from Seven Studies of Preschool Programs." *Child Development* 78, no. 2 (March-April 2007): 558-80, <https://doi.org/10.1111/j.1467-8624.2007.01014.x>.
50. Nadia S. Orfali Hall, Diane M. Early, and Deborah Seok, "Understanding Links between Work Climate and Early Care and Education Classroom Quality," *Child Trends* (June 2020).
51. Karen E. Diamond, Laura M. Justice, Robert S. Siegler, and Patricia A. Snyder, "Synthesis of IES Research on Early Intervention and Early Childhood Education. NCSER 2013-3001," US Department of Education (2013).
52. Gery Marshall Mohler, Kimo Ah Yun, Amy Carter, and Deb Kasak, "The Effect of Curriculum, Coaching, and Professional Development on Prekindergarten Children's Literacy Achievement," *Journal of EC Teacher Education* 30 (February 2009): 49-68.
53. US Department of Education, "Toward the Identification of Features of Effective Professional Development for ECE Educators, Literature Review," Washington, DC: Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service; *Child Trends*, 2010.
54. Margaret Caven, "Understanding Teacher Turnover in Early Childhood Education." Yale University School of Medicine, July 29, 2001, <https://medicine.yale.edu/news-article/understanding-ece-teacher-turnover>.
55. Fuller et al. "Equity and Institutions."
56. Daphna Bassok and Eva Galdo, "Inequality in Preschool Quality? Community-Level Disparities in Access to High-Quality Learning Environments," *Early Education and Development* 27, no. 1 (2016): 128-44, <https://doi.org/10.1080/10409289.2015.1057463>.

57. James J. Heckman, "Skill Formation and the Economics of Investing in Disadvantaged Children," *Science* 312, no. 5782 (2006): 1900–902, <https://doi.org/10.1126/science.1128898>.
58. Friedman-Krauss et al., "The State of Preschool 2021."
59. C. Weiland, "Launching Preschool 2.0: A Road Map to High Quality Public Programs at Scale," *Behavioral Science & Policy* 2, no. 1 (April 2016): 37–46, <https://doi.org/10.1177/237946151600200106>.
60. Mike Puma, Stephen Bell, Ronna Cook, Camilla Heid, Pam Broene, Frank Jenkins, Andrew Mashburn, and Jason Downer, "Third Grade Follow-up to the Head Start Impact Study Final Report." *OPRE Report # 2012-45*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, US Department of Health and Human Services (October 2012).
61. Grover J. "Russ" Whitehurst, "Does State Pre-K Improve Children's Achievement?," *Brookings: Evidence Speaks Reports*, 2, no. 59 (July 2018).
62. Mimi Engel, Amy Claessens, and Maida A. Finch, "Teaching Students What They Already Know? The (Mis)alignment between Mathematics Instructional Content and Student Knowledge in Kindergarten," *Educational Evaluation and Policy Analysis* 35, no. 2 (2013): 157–78, <https://doi.org/10.3102/0162373712461850>.
63. Ann Gervasoni and Bob Perry, "Children's Mathematical Knowledge Prior to Starting School and Implications for Transition," in *Mathematics and Transition to School: International Perspectives*, ed. B. Perry, A. MacDonald, and A. Gervasoni (Singapore: Springer, 2015).
64. Engel et al., "Teaching Students What They Already Know?," 157–78.
65. Guthrie Gray-Lobe, Parag A. Pathak, and Christopher R. Walters, "The Long-Term Effects of Universal Preschool in Boston," *Quarterly Journal of Economics* 138, no. 1 (February 2023): 363–411, <https://doi.org/10.1093/qje/qjac036>.
66. William T. Gormley Jr., Sara Amadon, Katherine Magnuson, Amy Claessens, and Douglas Hummel-Price, "Universal Pre-K and College Enrollment: Is There a Link?," *AERA Open* 9, no. 1 (January 2023): 1–17, <https://doi.org/10.1177/23328584221147893>.
67. Martha J. Bailey, Shuqiao Sun, and Brenden Timpe, "Prep School for Poor Kids: The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency," *American Economic Review* 111, no. 12 (2021): 3963–4001, <https://doi.org/10.1257/aer.20181801>.
68. Deborah A. Phillips, Mark W. Lipsey, Kenneth A. Dodge, Ron Haskins, Daphna Bassok, Margaret R. Burchinal, Greg J. Duncan, Mark Dynarski, Katherine A. Magnuson, and Christina Weiland, "Puzzling It Out: The Current State of Scientific Knowledge on Pre-Kindergarten Effects. A Consensus Statement," in *Issues in Pre-Kindergarten Programs and Policy*, ed. Kenneth A. Dodge (Washington, DC: Brookings Institution, 2017), 19–30.
69. Lynn Karoly, "The Costs and Benefits of Scaled-Up Pre-Kindergarten Programs," in *The Current State of Scientific Knowledge on Pre-Kindergarten Effects*, ed. Deborah A. Phillips, Mark W. Lipsey, Kenneth A. Dodge, Ron Haskins, Daphna Bassok, Margaret R. Burchinal, Greg J. Duncan, Mark Dynarski, Katherine A. Magnuson, and Christina Weiland (Baltimore: Brookings Institution, 2017), 57–66.
70. Karoly, "Costs and Benefits," 57–66.



2. A Survey of Whole-Child School Reforms

Maria D. Fitzpatrick

Executive Summary

Whole-child education models are those that expand the ambit of schools beyond a traditional academic focus. While a range of whole-child models have been explored since at least the Progressive Era, use of these models has expanded greatly over the past twenty years.

Because nearly all children in the United States attend public schools, it can be a tempting place to provide near-universal access to programs and resources. However, for various reasons, some families and educators are wary of a more expansive role for schools in children's lives beyond academic training. I review several examples of whole-child reforms that have become popular over the past few decades: community schools, school-based health centers, wraparound service models, and social emotional learning curricula. After describing the general framework of each, I explore research into each model's effectiveness using standards of high-quality causal inference evidence defined by the US Department of Education.

While some models have proven effective at shifting child outcomes in certain settings, none have yet been proven—at large scale, using high-quality causal research methods—to be a silver bullet that can overcome the challenges many children face today in terms of improving academic outcomes. Though they may have other positive impacts on their own, without related investment in academic reforms, they are unlikely to be the panacea for the low academic performance that plagues children in the United States. Thus, at the end of this brief, I close with recommendations for policymakers to think carefully about implementation of these models in their own contexts.

- Whole-child education models are becoming better known in the United States.

- Their adoption in some public schools provides an opportunity to see which models contribute to academic success.
- However, they are a part of the topic of child welfare, not the entire picture.



In the past couple of decades, there has been a renewed interest in the idea that schools should expand their ambit to address a wider range of student needs around health and well-being. Often this is described as a focus on development of the “whole child” rather than just the academic aspects of child development.

Of course, promotion of a wider ambit for schools beyond the academic sphere is at least a century old, as is the debate about whether it is optimal. The intellectual leaders of the Progressive Era, in the nineteenth century, sought to bring a broader focus to education systems than the traditional academic one. This included various ways of engaging the whole child, some of which are similar to the models covered here, particularly the social and emotional learning curricula and community school models that have skyrocketed in popularity in the past several years.

Similarly, the roots of whole-child reforms that are focused on improving children’s physical health are deeply embedded in US education history. As early as 1850, states began requiring immunizations and sometimes hosted immunization clinics in schools, where there was easy direct access to children. Also, the beginning of what we now know as the standard school nurse model began in 1902 as a pilot program aiming to insert healthcare into schools in order to improve chronic absenteeism by managing easily treatable illnesses and focusing on prevention. Each of these foreshadowed the more recent creation and rapid expansion of school-based health centers, which insert healthcare providers directly into schools with the goal of improving academic and overall well-being.

Recent decades have seen a renewal in the popularity of whole-child models. To some extent, this renewed interest is partly a backlash to what many perceived as the laser focus of the No Child Left Behind era on student test score performance. The difficult periods of the Great Recession and the COVID-19 pandemic also contributed to this shifted focus. The recent version of this movement has also been helped by increased emphasis on the complex relationships between education, health, housing, and other social dimensions across a range of academic disciplines and policy spheres.

This whole-child movement in schools has taken many forms, some of which I describe in more detail below. Across all its forms, the theory of change driving whole-child reform has two main parts. First, many students struggle academically because their basic needs are not met. Second, supporting these basic needs directly by bringing healthcare and/or social service resources into the school itself will overcome the access barriers that some children face, particularly poor children, thereby increasing their ability to thrive academically and socially.

To some extent, this theory of change pervades the entire US education system. Almost all districts in the country provide some form of nonacademic care to students through the school nurse, school counselors, or expanded offerings like universal vision screening programs. And many provide extracurricular activities or partner with community organizations in a variety of ways. What differentiates the whole-child models of reform here from the standard public school environment is the broader range of services provided and the depth of engagement between the school and community partners.

Intuitively, the first part of this theory of change makes some sense. How can a child learn if they suffer from an ongoing undiagnosed disease or disorder that prevents them from attending school regularly, concentrating in class, or participating fully in the community around them? How can a child learn if they feel isolated in a community, are surrounded by violence, and lack strong support inside and outside of school?

There is little direct causal evidence to support this theory of change, and there are plenty of anecdotes about children thriving despite incredibly challenging experiences during childhood. Yet a majority of parents would agree that children thrive most when their basic needs are met. However, as with all aspects of childrearing, there is debate about which “needs” require fulfillment for children to thrive. Furthermore, there is debate about whether schools are the best provider of health and social services to support children.

For decades, people have debated whether schools are the most effective places to solve the deep-rooted societal problems, like poverty, that leave many children with their basic needs unmet. Some people see schools as the great equalizer, holding them uniquely responsible for the achievement and well-being of all students, regardless of their backgrounds or the social forces determining those backgrounds. Others argue that systemic poverty, isolation, violence, poor health, and other ills have such a strong role that schools cannot be responsible for overcoming them.

Because nearly all children in the United States attend public schools, it can be a useful place to provide nearly universal access to programs and resources. However, for various reasons, some families are wary of a more expansive role for schools in children’s lives beyond academic training. Some have concerns about the differences between their own values and beliefs and those promoted in the school environment, as is the case with the recent backlash among social emotional learning programs. Others have concerns about whether school employees have the bandwidth and expertise to provide an expansive range of high-quality care; instead, they suggest that a focus on academic knowledge would allow school employees, like teachers, to be more impactful. Still others distrust the push for schools to focus on issues beyond academics because of concerns about greater intrusion into the private lives of families.

Below I review several examples of whole-child reforms that have become popular over the past few decades. After describing the general framework of each, I explore research into each model’s effectiveness. Most have been described as effective by the literature, but this

assertion is generally based on research that is largely theoretical, comprises mixed methods, or is conducted either at a small scale or without the types of carefully constructed comparison groups that are essential for determining causal impacts. I focus on summarizing the subset of this literature that meets the Tier 1 or Tier 2 standard of the US Department of Education for strong or moderate evidence of effectiveness from either an experimental or a quasiexperimental design study (What Works Clearinghouse 2020).

Further, since many areas of research have shown patterns of effective programs in small studies that have limited effectiveness when taken to scale, I place particular emphasis on the relatively few studies that have analyzed the effectiveness of programs with large numbers of students across multiple school settings. While some have proven effective at shifting child outcomes in certain settings, none have yet been proven at scale, using high-quality causal research methods, to be a silver bullet that can overcome the challenges many children face today. Importantly, when looked at in total and given the scale of the existing research, the lack of conclusive evidence of a clear positive causal effect of these reforms on children's academic achievement casts doubt on the theory underlying these reforms. Though they may have other positive impacts, on their own and without attention to academic reforms they are unlikely to be the panacea for low academic performance that plagues children in the United States. Thus, at the end of this chapter, I close with recommendations for policymakers to think carefully about implementation of these models in their own contexts.

COMMUNITY SCHOOLS

The defining feature of a community school is the creation of a more comprehensive, deeper relationship between the school, its families, and the surrounding community than exists in the common public school framework. The theory motivating this design is that the strengthened relationships and related supports provided to children and families will allow children to thrive academically. Some also argue for the model's value in promoting the creation of a more engaged citizenry of the students, families, and neighbors of a community school.

There is no single definition of what a community school is—it is more a strategy than a particular model. In part, this is in acknowledgment that the needs and goals of individual schools and communities may vary, and so the design of the school and its partnerships should vary. In recent years, though, practitioners and researchers have theorized that the following four common core pillars of community schools may be important for their success (Maier et al. 2017):

1. Integrated student supports provided in partnership with social service agencies and health providers
2. Expanded learning time and opportunities—often operationalized as extended-day or year-round schooling—expanded access to extracurricular enrichment activities, or individualized academic support

3. Family and community engagement with shared decision-making responsibilities and expanded roles in onsite opportunities
4. Collaborative leadership and practice across participating groups, including district and school leadership, teachers, and community organizations, often with goals of data-driven assessment of operations

For more than a century, educators and politicians have touted the potential benefits of stronger engagement between schools and their surrounding communities. Like many education reforms, the use of a community school-type model is rooted in the progressive education movement. (John Dewey, in 1902, wrote an essay on the subject, titled “Schools as Social Center” [Dewey 1902].) Also, like other reforms, the community school movement has waxed and waned, often gaining in popularity in times of great economic and social change, such as the industrial revolution, the Great Depression, and the civil rights era. The most recent wave of expanded interest in the community school model likely began in the 1990s with the creation of popular national models and attention to the effects of living in poverty on children’s educational progress. More recently, the proliferation of community schools has accelerated in response to the strong focus on academic outcomes of the No Child Left Behind Act of 2001 and with increased national attention to particular models.

Although there is no centralized tracking of community schools across the country, estimates suggest that, as of 2023, between 8,000 and 10,000 schools identified as community schools in the United States (Quinn and Blank 2020). As with the other types of whole-child reform discussed here, there has been a marked expansion in recent years. This includes both conversion of individual schools to a community school model and systemwide initiatives across multiple schools. For example, the Community School Initiative in New York City was introduced in 2014 and began with forty-five schools. Seven years later this number had increased by nearly 1,000 percent. As of 2023, there are 421 community schools in New York City. This is more than the number of charter schools.

The availability of COVID-19 relief funds and the specific focus of those funds on supporting children’s mental health has further increased interest in community schools. For example, in 2022, California announced a \$3 billion investment in its Community Schools Partnership Program to provide up to seven years of funding to support the conversion of as many as four thousand Title I schools to a community school framework grounded in the pillars mentioned above (Fensterwald 2022).

Although not traditionally classified as community schools, the Promise Academy charter schools of the Harlem Children’s Zone in New York City have many of the features of community schools. Started in 2004 and expanded in 2007, the Academy has an extended school day and school year and offers free medical, dental, and mental health services in the school to all children. The Academy provides additional support to families as needed and connects to more than twenty community programs in the Harlem Children’s Zone that provide a variety of services and resources to children and families, such as extracurricular opportunities for children, health programs, and tax guidance programs.

Dobbie and Fryer (2011, 2015) examined the effectiveness of the Academy and related community services using multiple strategies, including the creation of comparison groups of students using those who were randomized into the Promise Academy and those who were randomized out. (This research therefore qualifies as moderate evidence in the classification schema laid out by the US Department of Education.) Dobbie and Fryer found that attending the Promise Academy led to reduced absences and increased test scores during elementary and middle school, as well as improvements in a variety of academic outcome measures such as high school completion. They also identified decreases in teen pregnancy for girls and incarceration for boys.

Crucially, the improvements for children in the Harlem Children’s Zone accrued to children who enrolled in the Promise Academy but were not different based on whether the children had access to the community resources provided by the program. To investigate this, the authors compared the outcomes of lottery winners and lottery losers who lived within the zone and had access to the additional resources, to the outcomes of lottery winners and losers who lived outside the zone and did not have access to the additional resources. This suggests that it is the Promise Academy itself that was impactful in improving child outcomes, rather than the broader array of community services in the Harlem Children’s Zone.

Similar studies have used methods of creating comparison groups of comparable students and schools to examine the effectiveness of community schools in other settings, such as Tulsa, Oklahoma, as well as Iowa, Maryland, Pennsylvania, and Washington State (Adams 2010; LaFrance Associates 2005a, 2005b, 2005c, 2005d). The results of these studies are mixed. There are positive effects of community school settings on academic and behavioral outcomes for children in some settings but not others. The range of effects could be driven by different characteristics of the community school architecture since there is some evidence that the most comprehensive interventions—those that include strong versions of each of the four core components listed above—show a stronger positive effect than those with only one of these dimensions or with only mild commitment to a particular dimension (Adams 2010). However, the differences could also be attributable to differences in design of the comparison groups and other dimensions of the studies, making it hard to generalize.

A recent study of the first set of schools in the New York City Community Schools Initiative is notable for its careful choice of schools for the comparison group and for the large number of community schools included (Covelli et al. 2022). The researchers made use of the decision rules about which schools would be among the first to become community schools in New York City to create a comparison group of schools that is arguably quite similar to the group of schools that became NYC Community Schools. They found that, in this initial group of forty-five community schools, chronic absenteeism fell immediately and test score performance increased within a few years.

This limited evidence is somewhat promising. We need additional studies of community school programs at scale to understand whether models that have worked in isolation in relatively small settings work more broadly. This is particularly true since the diffuse definition

of exactly what a community school is makes it hard to replicate successful models and bring them to scale across a broader range of schools. Evaluations thus far have focused on settings where school leaders and communities were initiating a new community-engaged schooling model on their own. Given that this model likely takes large investments by a range of leaders and stakeholders, results might be very different when schools are induced to adopt a vaguely defined community school model because of a funding incentive offered, say, by the state or federal government, particularly if available funding is not sustained and not as comprehensive as existing small-scale programs.

To date, research has identified some components of the model that are linked with more successful implementation of community schools. The four pillars outlined above are influential. Others have argued that a thorough initial strengths and needs assessment is essential, as are (1) connecting and coordinating across programs and services and (2) authentic community engagement (Quinn and Blank 2020). Some of the best-known models, which are considered by many to be successful, include specific staff, either at the school or at the state or district level, who coordinate across several schools and are dedicated to supporting the assessment and comprehensive coordination considered essential for community school success. Policymakers looking to explore a community school model may be able to learn a lot by connecting to or coordinating with these staff and other leaders who have successfully implemented this model.

Even supporters of the community school model emphasize that additional community resources are not, on their own, a panacea for improving student achievement, and they worry about the tension between using limited resources to provide strong school supports or the more “outside-of-school” supports that community schools are known for (Shapiro 2016). Policymakers interested in adopting a community school model would be best served by ensuring that the emphasis on community partnership does not come at the cost of investing in academic supports that have proven effective.

WRAPAROUND SERVICES

A cousin of the community school model is the wraparound services model (also called the comprehensive or integrated student support model). This model is focused on the first pillar of the community school model described above—the provision of integrated student support in partnership with social service agencies and health providers—without necessarily having the other components. In the most intensive form, wraparound services are school-wide. Each student, regardless of their needs, is paired with a staff-support person to coordinate with families, teachers, health professionals, and community agencies in creating an individualized support plan with services tailored to the student. In this most intensive model, data on both the student and community partners is monitored and used to inform activities.

Historically, wraparound services were used to coordinate care for children with special needs, such as those with individualized education plans or those in the foster care system.

Recently, enthusiasm is spreading for the more intensive school-wide model described above. The rationale for expanding to the school-wide model is that a wider range of students could benefit from assistance “navigating the system” and that it will allow for children’s needs to be identified early enough to provide preventive support rather than waiting until problems have reached a more difficult level.

There have been positive findings in evaluations of some wraparound school services programs. For example, City Connects is a program originally started in 2001 in the Boston Public Schools system in partnership with researchers at Boston College. The program embeds a coordinator into the school to evaluate the needs of all children. The coordinator then creates a support plan for each student that involves connecting them with a range of specific service providers depending on their needs. Using multiple different quasi-experimental evaluation methods, researchers found the City Connects program improves child outcomes such as absences and performance on achievement tests (City Connects 2016, 2020, 2022).

City Connects started in the 1990s as a collaboration between Boston Public Schools, researchers at Boston College, and community agencies to investigate ways to help children in Boston’s public schools deal with factors outside the schools that were negatively impacting their success. After years of convening, engaging with the school and community partners, and planning, the first City Connects program debuted in six public schools in 2001.

City Connects involves six key factors considered important for its success (City Connects 2016):

- 1. School site coordinator (SSC):** Each school has a coordinator who is trained as a counselor or social worker and is responsible for working with students, teachers, school staff, families, and community agencies to evaluate strengths and needs of students and to connect each student to a tailored set of in-school and out-of-school supports and programs.
- 2. Whole-class review:** Each classroom teacher and SSC conduct a whole-class review to assess each student on four domains: academic, social/emotional/behavioral, health, and family. The review then involves identifying the appropriate supports and enrichment services and connecting each child and their family with providers of those services. It also involves tracking the use of the service and following up to ensure the appropriate fit.
- 3. Individual student review:** Students determined to have intensive needs are assessed by a team of professionals including teachers, school psychologists, principals, nurses, and community agency staff led by the SSC. The aim is to develop specific measurable goals for the student and family.
- 4. Community agency partnership:** Part of the SSC’s role is to create a set of strong partnerships with community agencies, including an advisory board of citywide agency leaders and an advisory council of representatives working at the local level.

5. **Connecting students to services, tracking, and following up:** Using program-specific software, SSCs connect students to services and programs, track individual plans and engagement, and continue their interaction with students, families, and school staff to determine appropriate fit.
6. **Service provision within the school:** SSCs also provide some services, such as crisis intervention, in the school as needed.

The range of services it is possible to connect students with is extensive. It includes before-school and after-school programs, enrichment opportunities, health and wellness programs, social skills interventions, academic support or tutoring, family assistance, and family counseling. Each student's support plan is individually tailored to their strengths and needs. The SSC is integrated into the school and community and receives training and professional development through the City Connects program.

Over the years, the program has expanded to other school districts in Massachusetts, as well as to districts in Indiana, Ohio, and other parts of the country. Evaluations have demonstrated that the model has continued to succeed as it has scaled across communities, making this a promising model for policymakers and district leaders to consider.

Another wraparound services program with moderate evidence of effectiveness is Massachusetts Wraparound Zones (Gandhi et al. 2016). The researchers compared students in schools receiving wraparound programs with those in a carefully constructed comparison group of schools and looked for whether there were breaks in the pattern of differences in various outcomes over time with the onset of access to the wraparound services. While they found no differences in attendance, retention, or suspension, they did find significant improvements in standardized test score performance for children in elementary school and middle school.

There is something novel in this particular type of reform. Many tout the latest education reform—be it a new curriculum, new organizational structure, or something else—as *the* silver bullet that will cure all problems of the system for all children. The wraparound services model intervenes by providing to each student specialized services tailored to their needs and goals. Using data to provide a tailored program to each child and family helps remove barriers that exist, even in other whole-child models. Also, the model does not involve a change in curricular focus away from academic work and instead supplements the resources available to families to support the academic work.

As research on these promising potential models expands, it will be important to determine the key aspects of the model and whether the model is replicable. For example, are particular support services useful for students, and are particular service providers responsible for improved child outcomes? Can these programs be expanded successfully beyond Boston or Massachusetts? Are the differences in outcomes across studies related to differences in populations or to differences in the models themselves?

SCHOOL-BASED HEALTH CENTERS

Another model that has gained traction over the past few decades is one that more closely links healthcare providers with schools. In theory, directly linking schools with community primary care providers—or even embedding the providers within school buildings—will remove the geographic, administrative, and financial barriers many children face in receiving care. Their resulting improved health will lead to improved learning outcomes.

Although the first school-based health centers of the 1960s were focused primarily on family planning services and support for teenage parents, the role of school-based health centers has expanded greatly. Nearly all offer primary care through a nurse practitioner or physician that includes immunizations, diagnosis and treatment of acute illnesses and chronic conditions, referrals, and follow-ups. Some also offer dental, optical, and mental health care. Another common aspect of school-based health centers is population-level health education and primary prevention programming. All children enrolled in school-based health centers can receive these services upon parental consent.

According to a national survey of providers, the number of school-based health centers more than doubled from 1999 to 2017, increasing from 1,135 to 2,584 (Love et al. 2019). School-based health centers now provide services to more than ten thousand elementary, middle, and high schools across the country. Typically, these organizations are not funded by and are organizationally separate from school districts. Funding comes from a range of sources, including local, state, and federal grants; foundation support; and reimbursement for services provided. The proportion covering costs through reimbursement for services from public and private insurers has grown over time, such that now 90 percent of school-based health centers seek reimbursement from insurers (Price 2016).

There is a large body of literature promoting the efficacy of school-based health centers for improving children’s health and academic outcomes. This includes prominent review articles, including those from Geierstanger et al. (2004), Knopf et al. (2016), Arenson et al. (2019), and Thomas et al. (2020). However, most of this literature does not meet the criteria for inclusion here because it lacks high-quality causal inference research design or is not conducted on school-based health centers at scale. The limits of this evidence are outlined in some of these review pieces and are clear even to the federal government. In a recent call for research proposals on school-based health centers, the National Institutes of Health summarized: “Although research evidence supports the *feasibility* of SBHCs to provide preventive and primary healthcare services to youth, rigorous research is lacking regarding their *effectiveness*” (National Institutes of Health 2022, emphasis added).

One exception is the research conducted by Westbrook et al. (2020), which explored the effects of school-based health centers across Colorado high schools on graduation rates. To isolate the causal relationship between school-based health center access and high school graduation rates, the authors compared graduation rates across schools that opened school-based health centers to those that did not, before and after the centers opened. They found

that school-based health center access increased graduation rates in these high schools, but the effects were only weakly statistically significant. In small samples, like the resulting data sample of 132 school-level data points, we should take caution generalizing weakly significant results. Thus, even this study exemplifies the need for a stronger research base for the claims of school-based health center effectiveness.

One of the challenges of school-based health centers is that the dependence on reimbursement from insurers for services makes them sustainable only in schools where there are many children who are eligible for Medicaid or the Children’s Health Insurance Program. Thus, the model will not be able to support the health of the many poor children in other school settings. This dependence on charging for reimbursable services also shifts the emphasis away from preventive care or other programs that can have a great impact on children in favor of services that can be charged to insurers.

Even when school-based health centers are in operation, they face challenges reaching all students. For example, in New York City, only about 60 percent of students in schools with school-based health centers actually enrolled in the center. Fewer still utilized the care provided by the center. In many school systems, the school-based health center replaces all other forms of school-provided physical and mental health services, such as those of the school nurse or other community health programs. When this happens, school-based health centers can only provide urgent first aid care to students who are not enrolled. The low enrollment rates of school-based health centers may mean some children actually lose access to healthcare services under the school-based healthcare model. Whether this is the case is an important open question, as few researchers have focused on whether students and families are more likely to participate in care in a traditional school nurse program than in a school-based health center. Policymakers interested in implementing school-based health center models should pay particular attention to ensuring that all children are enrolled and have access to the needed care in order to avoid any potential unintended consequences of adopting this new model.

SOCIAL AND EMOTIONAL LEARNING PROGRAMS

The Collaborative for Academic, Social, and Emotional Learning (CASEL) defines social and emotional learning as “the process through which all young people and adults acquire and apply the knowledge, skills, and attitudes to develop healthy identities, manage emotions and achieve personal and collective goals, feel and show empathy for others, establish and maintain supportive relationships, and make responsible and caring decisions.” The Collaborative defines competencies in five areas: self-awareness, self-management, social awareness, relationship skills, and responsible decision-making.

Like the other models in this chapter, social and emotional learning programs encompass a wide range of activities, sometimes in partnership with families and communities. Most commonly, a social and emotional learning program is a set of curricula adopted by a school to be

implemented by teachers school-wide or in particular grades (e.g., elementary school). This means the model is different from others considered here in two important ways. First, rather than healthcare and social services provided by professionals, as with school-based health centers, for example, most social emotional learning programs are in the hands of teachers who often have little training in mental, social, and behavioral health. This leaves open the quality of programming, particularly when programs are implemented at scale where we know fidelity often fades. Second, social emotional learning programs typically are not coupled with shifts in the length of the school day or other additional community resources. Therefore, time devoted to social and emotional learning curricula likely crowds out time spent on other learning activities.

Several common programs exist, such as Second Step, a classroom-based social skills program built on cognitive behavioral therapy; and Tools of the Mind, an early childhood curriculum focused on developing self-regulation and executive functioning. However, there is no single arbiter of what defines a program as social and emotional learning, let alone which programs are effective for improving children's academic outcomes or other aspects of their well-being.

The vast majority of schools have adopted some form of social emotional learning program or curricula, and the use of these programs has grown over time. In 2018, 34 percent of elementary school teachers and 60 percent of elementary school principals reported implementing a social and emotional learning program. By 2022, those numbers had climbed to 60 percent and 81 percent of elementary school teachers and principals, respectively. (Because these numbers focus on formal implementation of programs, they likely miss the many school environments where teachers and staff are adopting some components of social and emotional learning programs without utilizing formal programming.)

Much of the growth over this period was driven by the COVID-19 pandemic. Concerns about social isolation due to school closure and about the emotional impact of living through a crisis environment increased the emphasis on child health, and not just for children or in schools. Recent research using claims data for insured individuals has demonstrated that utilization of mental health care services is now 39 percent higher than before the pandemic (Cantor et al. 2023). Given this heightened focus on mental health in society, it makes sense that there would be renewed interest in the use of social emotional learning programs in schools.

Recent changes to the Every Student Succeeds Act opened the door for districts to use federal funds to support implementation of social and emotional learning programs. The reauthorization of the bill in 2015 allowed states and districts to use federal funding to support any program that meets the criteria for Tier 1 (strong), Tier II (moderate), or Tier III (promising) evidence. This opens the door for states and districts to use a wide range of evidence, of varying quality, to motivate spending on social and emotional learning programs. Even further, states and districts can also adopt programs that are only at Tier IV (demonstrating a rationale), as long as they build in evaluation systems to measure effectiveness as they go.

Publications by research and policy organizations have provided various “how-to” guides that make it easy for policymakers to choose programs that fit these criteria and include them in proposals for various funding streams from the federal government (Title I, Title II, and Title IV) (Grant et al. 2017). This has undoubtedly had a large impact on the use of these funds and, therefore, the proliferation of social and emotional learning programs.

As mentioned, to help school leaders determine the efficacy of social emotional learning programs, multiple organizations have produced reviews of existing evidence of their effectiveness (CASEL 2013, 2015; Durlak et al. 2015; Grant et al. 2017; Jones et al. 2017). However, much of this evidence does not meet the US Department of Education standards for strong or moderate evidence. For example, none of the studies summarized in Grant et al. (2017) provide strong or moderate evidence that any of the dozens of specific programs studied improve children’s academic outcomes. There is some strong and moderate evidence of the effectiveness of these programs at improving children’s inter- and intrapersonal skills, particularly in the short term immediately after program participation. However, it is still an open question as to whether these improvements in social skills are maintained for very long after an intervention and whether they fulfill the promise of improving children’s ability to learn inside and outside the classroom.

A big concern among parents, professionals, and policymakers about social and emotional learning programs is that the term (and the way it is defined, for example, in the quote from CASEL above) is vague enough to be meaningless and to encompass a wide range of programs. Without clear definitions of program focus, a clear research base about the most important components, and a stronger research base about effective programs, almost any program can be sold—both literally and figuratively—to schools as important for investment. Recently, parents have begun objecting to programs billed as social and emotional learning programs, in part because these parents believe the programs have extended beyond traditional areas, like self-awareness and grit, into more controversial areas, like sex education and critical race theory. In this way, social and emotional learning programs are becoming a new front in the culture wars (Field 2022).

There is also the very real question of how to incorporate focus on social and emotional learning into what many consider an already full school day. Since few rigorous studies have shown positive effects of social and emotional learning programs on academic outcomes and because most of these programs are curricular interventions executed by teachers in the classroom, it is fair to assume that attention to these programs will crowd out attention to academic learning. That is a trade-off many parents and taxpayers are uncomfortable with, particularly given the large declines in children’s achievement during the COVID-19 pandemic.

Moreover, many people who are in favor of social and emotional learning programs would also argue that teachers are already overburdened. Without proper training and support, it is difficult to imagine teachers effectively implementing curricula on a new dimension of cognitive development. Even with proper training and resources, these programs will necessarily shift focus away from areas of academic learning that are in great need of

attention—particularly now, post-pandemic. Asking teachers to master and incorporate a set of curricula on social and emotional learning skills on top of existing curricula seems to risk them being unable to do well in either area.

COMMON THEMES AND CONTROVERSIES IN WHOLE-CHILD MODELS

In reviewing the evidence on recent whole-child reforms, a few themes emerge. First, the theories motivating these reforms are at least a century old, as is the underlying structure of the models themselves. Under different names, some of these models, like the community school model, have been around for many decades. Second, many are focused on deep engagement with community partners to provide a more extensive and more comprehensive set of health and social services support in schools than the traditional model focused on academics. The exception is the social and emotional learning model, which more commonly operates with teachers as the provider of curricula focused on particular nonacademic skills. Third, all of these models either explicitly or implicitly assume that the school is the best vehicle or central conduction point for intervening to improve nonacademic outcomes.

A few controversies surround whole-child models, either as a group or for specific models. The most common controversy is the debate over whether schools are the appropriate places to focus on nonacademic outcomes and on the broader societal ills that so negatively impact child well-being. As I discussed at the beginning of this chapter, there are many potential concerns about locating this work in schools. One of the largest concerns is that, given scarce financial and bandwidth resources in public schools, these efforts will crowd out attention to academic outcomes and the academic reforms that research has proven are effective at improving child outcomes. Relatedly, others are concerned that schools do not have the expertise to venture beyond the traditional academic focus.

Social and emotional learning programs are notable for the growing controversy in the public and political spheres. Particularly since the pandemic, as the country has increased its focus on mental health, there has been incredible growth in the adoption of a wide range of curricula labeled as social and emotional learning. Because the curricula are adopted without expanding time in the classroom and are often taught by teachers without much, or any, background training, concerns exist that the social and emotional curricula are crowding out time spent on building academic skills without even the potential benefit that might accrue if the social and emotional skill building were in the hands of professionals trained in these areas.

TAKEAWAYS FOR POLICY

The emphasis on whole-child reforms in recent decades is understandable from a rhetorical perspective. Who among us would argue that they only care about one part of our children's

development? Or that they do not want to support children’s development as productive citizens of society across a range of dimensions? And what parent who is trying to teach their child a new skill does not understand the value of concepts like grit and problem-solving skills?

However, the research evidence motivating the models of whole-child reform in schools is limited. To date, there is some promising evidence from a few programs in a few settings, but there is not enough to support any of these models’ ability to improve children’s educational outcomes at scale. There is also limited evidence on other outcomes of interest at scale. When examined in total, the lack of a more prominent relationship between improvements in social and behavioral outcomes and academic outcomes casts doubt on the underlying theory that school investments in a wider range of areas of child development are what is needed to enhance students’ academic success.

One takeaway from this is that policymakers interested in improving their students’ learning, particularly in the wake of the learning loss of the COVID-19 pandemic, should redouble their focus on academics by investing in tools that we know work, such as attracting and training a high-quality teaching workforce, extending the time children spend in high-quality learning environments, and math and reading curricula that demonstrate improvements.

For example, there is a growing body of high-quality causal evidence showing that teacher effectiveness can be improved. For example, teacher evaluation programs have improved teacher effectiveness (Taylor and Tyler 2012). Even low-cost peer-observation programs, which pair teachers to observe and provide feedback on each other’s teaching without the incentives and extra costs of evaluation systems, have had positive effects on teacher performance (Burgess et al. 2021; Papay et al. 2020).

Of course, it is unlikely that any single reform will be a silver bullet that can improve education outcomes for all children. It may be the case that, in some districts, a whole-child model seems like the best available intervention. In these settings, care should be taken to determine the needs and goals of the school or district; clearly articulate these for the community; choose a whole-child model focused on the relevant outcomes with research evidence validating its effectiveness; provide adequate resources to support the implementation; and commit to a process of continual evaluation and a willingness to change directions or to abandon the model if it is not effective.

In this case, the most promising whole-child reform reviewed here is the wraparound services model, such as the City Connects program. In part, this is because the model is not a one-size-fits-all model that is trying to improve the outcomes of all children with an intervention in a particular area. Instead, it is a model of providing direct, data-driven, personally tailored support to each student using existing academic and broader resources. It has the benefits of increased access to resources for individual families and children that is a hallmark of proponents of whole-child reforms—but without the potential drawbacks of entirely restructuring the school organization or environment or shifting curricula wholesale for all children in ways that detract from time spent in academic environments.

Policymakers looking to invest in any of these reforms, either whole child or academic focused, should take a careful look at existing research. The US Department of Education's taxonomy of research quality is helpful, as is the What Works Clearinghouse it operates. However, the current level of quality required to motivate federal funding through the Every Student Succeeds Act is well below the level that a discerning policymaker or school leader should set before adopting a reform.

Finally, policymakers looking to initiate whole-child reforms should take care to complement those investments with investments focused on academics. This will avoid the possible pitfall of not maintaining emphasis on strong academic programming that even advocates of whole-child reforms raise concerns about and that is an important component of the overall effectiveness of whole-child reforms.

HESI PRACTITIONER COUNCIL RESPONSE

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.

Policymakers play a crucial role in shaping and implementing education reforms. When it comes to investing in education, policymakers should prioritize the establishment of a robust accountability mechanism. This mechanism should go beyond de minimis or adequate yearly progress and instead aim for a baseline standard for the quality of education that children receive within any model.

By incorporating explicit measurements and guidelines for accountability, policymakers can ensure that education reforms are grounded in evidence, have a significant impact, and can be sustained over time. This accountability system should encompass the establishment of clear benchmarks and targets for student outcomes, school performance, and overall educational quality. Through rigorous evaluation, policymakers can identify the reforms that yield the highest returns on investment and replicate them to drive positive change across educational institutions.

A comprehensive accountability mechanism should extend beyond academic performance to encompass other crucial aspects of education, such as well-rounded development and student well-being. By including factors like social emotional skills, creativity, and critical thinking in the accountability framework, policymakers can ensure that reforms go beyond a focus solely on test scores and promote the holistic growth of students.

Policymakers should involve all stakeholders, including educators, parents, and community members, in the development and implementation of reforms. This collaborative approach will ensure that reforms address the specific needs and challenges of the education system.

—Christina Laster, CEO of Bold Enterprises LLC

REFERENCES

- Adams, Curt M. 2010. "The Community School Effect: Evidence from an Evaluation of the Tulsa Area Community School Initiative." Oklahoma Center for Educational Policy: University of Oklahoma.
- Arenson, Michael, Philip J. Hudson, NaeHyung Lee, and Betty Lai. 2019. "The Evidence on School-Based Health Centers: A Review." *Global Pediatric Health* 6 (January): 2333794X1982874. <https://doi.org/10.1177/2333794X19828745>.
- Burgess, Simon, Shenila Rawal, and Eric S. Taylor. 2021. "Teacher Peer Observation and Student Test Scores: Evidence from a Field Experiment in English Secondary Schools." *Journal of Labor Economics* 39, no. 4 (October): 1155–86. <https://doi.org/10.1086/712997>.
- Cantor, Jonathan H., Ryan K. McBain, Pen-Che Ho, Dena M. Bravata, and Christopher Whaley. 2023. "Telehealth and In-Person Mental Health Service Utilization and Spending, 2019 to 2022." *JAMA Health Forum* 4, no. 8 (August): e232645. <https://doi.org/10.1001/jamahealthforum.2023.2645>.
- CASEL (Collaborative for Academic, Social, and Emotional Learning). 2013. "Effective Social and Emotional Learning Programs: Preschool and Elementary Edition." Chicago: Collaborative for Academic, Social, and Emotional Learning.
- . 2015. "Effective Social and Emotional Learning Programs: Middle and High School Edition." Chicago: Collaborative for Academic, Social, and Emotional Learning.
- City Connects. 2016. "The Impact of City Connects: Student Outcomes." Boston: Center for Optimized Student Support.
- . 2020. "City Connects: Intervention and Impact." Boston: Center for Optimized Student Support.
- . 2022. "The Impact of City Connects: Progress Report 2022." Boston: Center for Optimized Student Support. <https://www.bc.edu/content/dam/bc1/centers/city-connects/Publications/The%20Impact%20of%20City%20Connects-Progress%20Report%202022.pdf>.
- Covelli, Lauren, John Engberg, and Isaac M. Opper. 2022. "Leading Indicators of Long-Term Success in Community Schools: Evidence from New York City." EdWorkingPaper no. 22-669. Annenberg Institute at Brown University. <https://doi.org/10.26300/59Q2-EK65>.
- Dewey, John. 1902. "The School as Social Center." *Elementary School Teacher* 3, no. 2 (October): 73–86.
- Dobbie, Will, and Roland G. Fryer. 2011. "Are High-Quality Schools Enough to Increase Achievement among the Poor? Evidence from the Harlem Children's Zone." *American Economic Journal: Applied Economics* 3, no. 3 (July): 158–87. <https://doi.org/10.1257/app.3.3.158>.
- . 2015. "The Medium-Term Impacts of High-Achieving Charter Schools." *Journal of Political Economy* 123, no. 5 (October): 985–1037. <https://doi.org/10.1086/682718>.
- Durlak, Joseph A., Celene E. Domitrovich, Roger P. Weissberg, and Thomas P. Gullotta. *Handbook of Social and Emotional Learning: Research and Practice*. New York: Guilford Press, 2015.
- Fensterwald, John. 2022. "California Ready to Launch \$3 Billion, Multiyear Transition to Community Schools." *EdSource*, January 31. <https://edsources.org/2022/california-ready-to-launch-3-billion-multi-year-transition-to-community-schools/666571>.
- Field, Kelly. 2022. "Social and Emotional Learning Is the Latest Flashpoint in the Education Wars." *Hechinger Report*, February. <https://hechingerreport.org/social-and-emotional-learning-is-the-latest-flashpoint-in-the-education-wars>.
- Gandhi, Allison, Rachel Slama, So-Jung Park, Patrick Russo, Robin Bzura, and Sandra Williamson. 2016. "Focusing on the Whole Student: Final Report on the Massachusetts Wraparound Zones." Waltham, MA: American Institutes for Research.

- Geierstanger, Sara Peterson, Gorette Amaral, Mona Mansour, and Susan Russell Walters. 2004. "School-Based Health Centers and Academic Performance: Research, Challenges, and Recommendations." *Journal of School Health* 74, no. 9 (November): 347–52. <https://doi.org/10.1111/j.1746-1561.2004.tb06627.x>.
- Grant, Sean, Laura S. Hamilton, Stephani L. Wrabel, Celia Gomez, Anamarie Auger, Jennifer Tamargo, Fatih Unlu et al. 2017. "Social and Emotional Learning Interventions under the Every Student Succeeds Act: Evidence Review." Santa Monica, CA: RAND Corporation.
- Jones, Stephanie, Katharine Brush, Rebecca Bailey, Gretchen Brion-Meisels, Joseph McIntyre, Jennifer Kahn, Bryan Nelson, and Laura Stickle. 2017. "Navigating Social and Emotional Learning from the Inside Out." Boston: Harvard Graduate School of Education.
- Knopf, John A., Ramona K. C. Finnie, Yinan Peng, Robert A. Hahn, Benedict I. Truman, Mary Vernon-Smiley, Veda C. Johnson et al. 2016. "School-Based Health Centers to Advance Health Equity." *American Journal of Preventive Medicine* 51, no. 1 (July): 114–26. <https://doi.org/10.1016/j.amepre.2016.01.009>.
- LaFrance Associates, LLC. 2005a. "Comprehensive Evaluation of the Full-Service Community Schools Model in Iowa: Harding Middle School and Moulton Extended Learning Center." San Francisco: Milton S. Eisenhower Foundation.
- . 2005b. "Comprehensive Evaluation of the Full-Service Community Schools Model in Maryland: General Smallwood Middle School." San Francisco: Milton S. Eisenhower Foundation.
- . 2005c. "Comprehensive Evaluation of the Full-Service Community Schools Model in Pennsylvania: Lincoln and East Allegheny Middle Schools." San Francisco: Milton S. Eisenhower Foundation.
- . 2005d. "Comprehensive Evaluation of the Full-Service Community Schools Model in Washington: Showalter Middle School." San Francisco: Milton S. Eisenhower Foundation.
- Love, Hayley E., John Schlitt, Samira Soleimanpour, Nirmita Panchal, and Caroline Behr. 2019. "Twenty Years of School-Based Health Care Growth and Expansion." *Health Affairs* 38, no. 5 (May): 755–64. <https://doi.org/10.1377/hlthaff.2018.05472>.
- Maier, Anna, Julia Daniel, Jeannie Oakes, and Livia Lam. 2017. *Community Schools as an Effective School Improvement Strategy: A Review of the Evidence*. Palo Alto, CA, and Washington, DC: Learning Policy Institute.
- National Institutes of Health. 2022. "Call for Research Proposals on the Effectiveness of School-Based Health Centers to Advance Health Equity." Washington, DC: Department of Health and Human Services. <https://grants.nih.gov/grants/guide/pa-files/PAR-21-287.html>.
- Papay, John P., Eric S. Taylor, John H. Tyler, and Mary E. Laski. 2020. "Learning Job Skills from Colleagues at Work: Evidence from a Field Experiment Using Teacher Performance Data." *American Economic Journal: Economic Policy* 12, no. 1 (February): 359–88. <https://doi.org/10.1257/pol.20170709>.
- Price, Olga Acosta. 2016. "School-Centered Approaches to Improve Community Health: Lessons from School-Based Health Centers." A Series of Discussion Papers on Building Healthy Neighborhoods. Washington, DC: Brookings Institution, July 2016.
- Quinn, Jane, and Martin J. Blank. 2020. "Twenty Years, Ten Lessons: Community Schools as an Equitable School Improvement Strategy." *VUE (Voices in Urban Education)* 49, no. 2. <https://doi.org/10.33682/3csj-b8r7>.
- Shapiro, Eliza. 2016. "New York's Community School Expansion Rattles Allies." *Politico*, July 28. <https://www.politico.com/states/new-york/albany/story/2016/07/allies-fret-as-new-york-becomes-nations-community-school-capital-104279>.
- Taylor, Eric S., and John H. Tyler. 2012. "The Effect of Evaluation on Teacher Performance." *American Economic Review* 102, no. 7 (December): 3628–51. <https://doi.org/10.1257/aer.102.7.3628>.

- Thomas, Casey L., Olga Acosta Price, Stephen Phillippi, and Ashley Wennerstrom. 2020. "School-Based Health Centers, Academic Achievement, and School Discipline: A Systematic Review of the Literature." *Children and Youth Services Review* 118 (November): 269–78. <https://doi.org/10.1016/j.chilyouth.2020.105467>.
- Westbrook, Marisa, Lisette Martinez, Safa Mechergui, and Sara Yeatman. 2020. "The Influence of School-Based Health Center Access on High School Graduation: Evidence from Colorado." *Journal of Adolescent Health* 67, no. 3 (September): 447–49. <https://doi.org/10.1016/j.jadohealth.2020.04.012>.
- What Works Clearinghouse. 2020. "Using the WWC to Find ESSA Tiers of Evidence." US Department of Education, Institute of Education Sciences. <https://ies.ed.gov/ncee/wwc/essa>.



3. Strategically Prioritizing Teacher Quality and Quantity Can Reinvigorate the Teacher Pipeline

Michael Hansen

Executive Summary

A Nation at Risk (ANAR) concluded that both the quality of current teachers and the quantity of available talent to fill teaching roles in public schools were sorely deficient. The report offered numerous recommendations to improve both dimensions of the teacher pipeline but failed to provide a coherent strategy to resolve the inherent tension between quality and quantity. Consequently, policy efforts that developed in the years following coalesced into two different approaches to teacher reform that frequently came into conflict and worked at cross-purposes.

In this essay, I focus on how *A Nation at Risk* helped to shape policies and practices impacting the teacher pipeline. I document the report's conflicting messages and review some of the evidence on key policies that grew out of the report's recommendations, even though they often worked against one another. Importantly, the report omitted consideration of socio-economic differences across schools and how they shape the teacher pipeline and limit disadvantaged students' access to instructional resources. Also, the report failed to recognize how pathway policies have been used to discriminate against people of color, contributing to continued underrepresentation in the workforce.

Moving forward, I identify the unique challenges for the teacher pipeline today and draw some parallels with 1983. I urge policymakers to learn lessons from the ANAR experience and plot a path forward by strategically combining promising reform efforts from both sides of the teacher conflict and implementing them in context-dependent ways. I also propose a big idea that would use generous scholarship money to directly build a cadre of individuals ready for teaching in the classroom.

- *A Nation at Risk* recognized the importance of both quality and quantity in the teacher pipeline.

- However, *ANAR*'s recommendations were at times contradictory, and they failed to consider policies that discriminated against people of color.
- Generous scholarship funding could build a cadre of teachers who are ready and suited for today's classrooms.



In its alarming and unapologetic critique of the American school system, *A Nation at Risk* directed many pointed barbs at the teacher workforce and those tasked with preparing teachers.¹ The report summarily concluded that both the quality of current teachers and the quantity of available talent to fill teaching roles in schools were sorely deficient; both dimensions needed immediate intervention to help achieve education excellence. It offered a range of explicit recommendations regarding teachers to address these staffing challenges, while simultaneously making other recommendations that implicitly demanded more and better teachers. What the report failed to do, however, was reconcile the inherent tensions in simultaneously pursuing both higher quality and quantity or offer a strategy to systematically develop the teacher workforce that was desired.

Consequently, *A Nation at Risk* unleashed a wave of teacher-focused reform that was both expansive and incoherent. With the report used to justify efforts as disparate as establishing best practices through the National Board for Professional Teaching Standards and creating new alternative pathways into teaching, *A Nation at Risk* seeded reform efforts that grew in opposing directions over subsequent decades. Essentially, the report helped organize both sides of the growing conflict that would later come to be dubbed the "teacher wars."²

In this chapter, I focus on how *A Nation at Risk* (henceforth *ANAR*) helped shape policies and practices impacting how new individuals are attracted into public education, prepared to be teachers, and deployed once they are ready for the classroom. Hereafter, I use the label "teacher pipeline" to refer to this whole preservice and induction process. In the sections that follow, I document the report's conflicting messages and review some of the evidence on key policies that grew out of the report's recommendations. I also highlight critical omissions from the report and the resulting missed opportunities for reform. Moving forward, I identify the unique challenges for the teacher pipeline today and draw some parallels with 1983. I urge policymakers to learn lessons from the *ANAR* experience and plot a path forward by strategically combining promising reform efforts from both sides of the teacher conflict and implementing them in context-dependent ways. I also propose a big idea that would use generous scholarship money to directly build a cadre of individuals ready for teaching in the classroom.³

STAGNATING AND LISTLESS: PUBLIC SCHOOLS IN 1983

Before diving into *ANAR*, let's quickly set the stage. In 1983, US public schools were at the tail end of nearly a decade of enrollment declines as the baby boomer generation aged out

of K-12 schools and were replaced by the less-populous Generation X kids.⁴ Public school enrollment had peaked in the early 1970s and shrunk by seven million students by 1983, amounting to a loss of roughly 15 percent.

People worried that not only was enrollment in decline but also the quality of public education itself. The public's responses to Sputnik (1957) and the technological advances that followed during the 1960s were well on their way to fading into collective memory. Though children in elementary school and middle school showed small achievement gains in the 1970s, SAT scores among high school graduates were in an unexplained decline.⁵ Some of the public's worries were certainly shaped by growing pessimism (especially among White parents) about school desegregation efforts. Over the course of more than twenty years following the 1954 *Brown v. Board of Education* Supreme Court decision, schools integrated in stages: first, resistance to desegregation; second, limited integration through school choice programs; and third, more forceful government actions, such as busing students, to achieve racial balance.⁶ The consequent White flight accelerated through the 1970s, pushing private school enrollment up while public school enrollment declined.⁷

The teacher workforce was also viewed as a depreciating asset. The profession was increasingly unionizing in the late 1960s and into the 1970s. Unions were often perceived as stymying school performance, hindering school leaders' ability to serve students. To wit, collectively bargained contracts grew in their specificity in this era, increasingly covering many non-compensation items—for example, class size maximums and procedures for reductions in force.⁸ Meanwhile, the professional workforce was opening for educated young women, and fewer college students chose to pursue teaching as a career. Retrospective analyses of college graduates' occupational choices during this era find that teaching attracted significantly fewer women with strong educational backgrounds.⁹

Finally, policymakers in this era did not possess a clear road map for success since much of the conventional wisdom about schooling was recently upended. The highly influential Coleman Report (1966) concluded that family and community factors—rather than schools—were more influential in explaining differences in long-term educational outcomes across different racial groups.¹⁰ School desegregation efforts were contemporaneously viewed as unsuccessful in promoting achievement gains and only furthering racial tensions.¹¹ Analyses by education economists were raising serious doubts that differences in education spending had any causal relationship with student outcomes.¹² Thus, policymakers were well aware that schools were in decline but had few ideas to remediate the situation.

A small opening came from the Coleman Report, however. In exploring the different schooling factors that influenced student outcomes, the authors concluded that “variation in school averages of teachers' characteristics accounted for higher variation than did all other aspects of the school combined, excluding the student body characteristics” (p. 316). In other words, teachers appeared to matter more than anything else schools were doing. Later analyses in the report (see pp. 316–19) suggested that it was teachers' education backgrounds and verbal ability that were particularly influential in explaining teachers' impacts. Hence, a seed focusing on teachers and their preparation was planted.

Meanwhile, policymakers grew increasingly worried that teacher training programs of the day were part of the problem, not the solution. In Virginia, a series of state legislative subcommittee meetings began working on teacher preparation and licensure issues in 1978 due to concerns about declining enrollment and uneven candidate quality. Contemporaneous reports showed college graduates in the liberal arts were at least as strong academically as those pursuing education majors. State policymakers grew to become narrowly supportive of a “liberal arts approach” to certifying teachers without completion of an approved training program, despite the vocal opposition of education deans in universities across the state.¹³ When the legislation finally passed in 1982, it became what we recognize today as the first state policy allowing for alternative certification for teachers.¹⁴

QUALITY VERSUS QUANTITY: PARSING THE REPORT’S RECOMMENDATIONS FOR THE TEACHER PIPELINE

With that background, *ANAR*’s publication was an electric jolt, spurring the public and school leaders out of their acquiescence of mediocrity by plotting a new course. I now turn my focus to *ANAR* itself and how its messages helped identify problems and solutions for the teacher pipeline.

First, the report’s findings, or statement of problems, on teachers echoed and further detailed the worries of the period. These included the following statements (pp. 22–23):

- Teachers were disproportionately sourced from the low end of the educational achievement distribution.
- Teacher candidates spent too little preparation time on learning content but instead focused on pedagogy.
- Teachers received uncompetitive salaries and had little input into critical education decisions.
- Ongoing subject-specific teacher shortages hampered quality instruction, especially in math and science.

Thus, the authors believed that the teacher workforce was underprepared for providing instruction aligned with their specialization fields and was undersupplied from high achievers and in specific subject areas. In other words, both the *quality* and *quantity* of the teacher workforce was deemed lacking. The report strongly implicates teacher compensation—both monetary payments and professional privileges—as being insufficient to attract the talent necessary to ensure excellence in education.

Note that as I discuss the contents of *ANAR*, I analyze its statements through the lens of their impact on *teacher quality* versus *teacher quantity*. Both the quality and quantity of teachers

available for the task are useful dimensions in producing education: higher-quality teachers can cover instructional material in less time, while more teachers can enable smaller class sizes and more personalized instruction. But keep in mind that there's an inherent trade-off between quality and quantity (all else equal); for example, increasing the quantity of the workforce is easiest if we apply few quality standards to new recruits. The quality-quantity trade-off is a useful framework for applying to the teacher workforce not only because it is a widely understood, intuitive concept but also because these are evergreen issues that frequently manifest in policy discussions about managing teachers (e.g., teacher evaluation systems, teacher shortages). Take note because this tradeoff will be a recurring theme throughout this essay.

The ANAR report explicitly recommended a set of seven teacher reforms (pp. 30-31) in response to the challenges detailed above:

- Increasing education standards for those preparing to teach and holding preparation programs accountable for candidates' qualifications
- Increasing salaries to be professionally competitive, market sensitive, and performance based, with employment decisions tied to teacher evaluations
- Offering an eleven-month teacher contract to lengthen instruction and preparation time and provide higher salaries to full-time teachers
- Developing teacher career ladders to provide opportunities for advancement without leaving the classroom
- Recruiting knowledgeable individuals from outside the classroom to help solve subject-specific teacher shortages
- Providing grants or loans to attract students into teaching, particularly in shortage fields
- Involving master teachers with the design of teacher preparation programs and supervising early-career teachers

Furthermore, the report included several other recommendations that are not directly focused on teachers but nonetheless had clear implications on teachers' work and preparing new teacher candidates in the pipeline. For example, recommendation #5 in the content section (p. 26) proposed content standards for teaching computer science to all high schoolers as part of the report's "New Basics" of the modern high school curriculum. Because computer science was not widely available to US high schoolers in 1983,¹⁵ implementing this recommendation represented a major shift in the quantity and type of new teachers demanded. In total, I count five policy recommendations that implicitly created pressure to recruit more and different teachers into the workforce and two policy recommendations that implicitly demanded higher-quality instruction.¹⁶

As evidenced by the expansive reach of these recommendations, *ANAR*'s authors believed comprehensive—not incremental—reform was necessary to improve the teacher workforce. Mirroring the problems identified above, these recommendations sought to improve instructional quality (e.g., increasing standards to teach) and draw a greater quantity of candidates into the teacher pipeline (e.g., grants for shortage fields). The report also included recommendations for how schools should compensate teachers to ensure that the teaching occupation could attract individuals in sufficient quantity (e.g., by being competitive) and quality (e.g., by being performance based).

Despite the expansive reach of the report's recommendations, the report did not prescribe a strategy for creating this new workforce. For example, beyond naming subjects that face ongoing staffing challenges, it provided little nuance about circumstances in which teacher quantity may be more important than teacher quality or vice versa. More importantly, it failed to recognize that these two dimensions are inherently in conflict—for example, a push for quality will tend to limit the quantity of people who can meet higher expectations. Improving the teacher workforce on any one dimension alone would have been difficult to achieve in isolation, but the combination of all recommendations simultaneously rendered the challenge more demanding, even impossible.

Consequently, since this key tension remained unaddressed in *ANAR*, it was not critically challenged in the policy discourse and ensuing reforms. Since both teacher quality and quantity were judged as key priorities, any reform efforts that aimed to further either goal implicitly earned the imprimatur of furthering the report's mandates. But because the mandates themselves sat in unresolved conflict, the stream of reforms that evolved were likewise caught in that same conflict, as I explore further in the next section.

DECISIVE AND UNCOORDINATED: POLICY DEVELOPMENTS AFTER A NATION AT RISK

Almost immediately following the report's publication, waves of interest in teacher pipeline reform emanated in different directions from various stakeholders, including policymakers, school leaders, and educators. Importantly, these waves moved with little direct coordination and eventually coalesced into two competing approaches to teacher pipeline reform. Before describing those differing sides, though, I quickly review the chain of events in rough chronological order.

Many state policy responses to *ANAR* were quick, especially those dealing with attracting more individuals into the profession. Though the report did not explicitly name what we today call "alternative certification," it did encourage bringing qualified people from other occupations into teaching to support hard-to-staff fields, and these policies gained momentum. As described previously, Virginia had already provided a route for noneducation majors to enter teaching, and several states quickly followed suit after *ANAR*'s publication. California developed a district intern program as a pathway to becoming a teacher in 1983, and New Jersey

and Texas enacted their own pathways in 1984.¹⁷ By 1986, twenty-three states had adopted alternative certification strategies.¹⁸

Other state policy developments took aim at reducing costs for teachers to go to college. For example, in 1985, California's legislature enacted the Assumption Program of Loans for Education, an early loan forgiveness program for students in teaching certificate programs, conditioned on working in California high-need schools for at least four years after graduation.¹⁹ The North Carolina Teaching Fellows program, created in 1986, was a service scholarship program that provided four-year scholarships for students who taught in the state's public schools for at least four years.²⁰ By 2016, more than forty states and the federal government offered some type of service scholarship or loan forgiveness program for teachers.²¹

Teacher pay was another focal point of quick policy experimentation. Average salaries nationwide had been slowly declining through much of the 1970s (after accounting for inflation), but they experienced a reversal in the 1980s, with the largest gains between 1984 and 1988, after the report's publication.²² Merit pay and career-ladder programs, which attempted to either differentiate pay or differentiate roles within teaching, gained a national spotlight thanks to then president Ronald Reagan's bully pulpit after *ANAR*'s publication.²³ Importantly, these models were believed to work in two ways: bring more people into teaching with larger expected payouts *and* attract new entrants who were top performers. Yet these theoretical predictions were given little time to surface, as all but a few of these experiments dried up quickly thereafter due to funding instability and teacher resistance.²⁴ Thus, single-salary schedules persisted in their dominance in determining teacher pay, despite a spate of post-*ANAR* experiments.

Some of *ANAR*'s ideas took a bit more time to percolate. Two different teacher educator groups immediately went to work metabolizing the recommendations from the report and developing their own detailed reform agendas. In 1986, two highly influential reports were released responding to the challenges of teacher education described in *ANAR*.²⁵ A consortium of deans representing colleges of education across the country writing under the name of the Holmes Group published *Tomorrow's Teachers* in April 1986.²⁶ The following month, the Carnegie Corporation of New York (also using teacher educators as the author team) published *A Nation Prepared: Teachers for the 21st Century*.²⁷ Both reports were similar in that they envisioned teachers—once unburdened of bureaucratic tasks—as taking on larger roles within schools, becoming the agents of knowledge curation about best practices in the field and leading professional development among colleagues. Notably, the reports offered slightly different routes to operationalize these changes. The Holmes Group saw state licensure and colleges of education as needing reform to better align with this expanded vision of teaching, whereas the Carnegie report sought to locate this authority over standards and the profession externally from state legislators and university systems. A common theme in both reports, however, was that teachers were to take on an elevated status in their schools and in the professional world broadly, which would reestablish the prominence of the teaching profession by rigorously curating teacher quality.

Several important developments originating in these reports came to be persistent features of the teacher workforce. First, the National Board for Professional Teaching Standards (NBPTS), initially proposed in *A Nation Prepared*, was established in 1987. The national standards board was founded to establish best practices and disseminate these practices in schools across the country. The organization continues today, primarily providing national board certification to experienced teachers who demonstrate excellence in the classroom.

Also, ideas connecting teacher candidate testing, graduate education, tiered state licensure, and the conferral of tenure were presented as necessary steps in improving the teaching profession in *Tomorrow's Teachers*. The state-based model of professional oversight put forward in that report (contrary to the external NBPTS model put forward in *A Nation Prepared*) served to strengthen the role of the state in establishing professional privileges, such as licensure and tenure. To be sure, many of these policies were already in practice pre-ANAR, though their implementation took on new roles in these waves of reform after ANAR. For example, teacher testing was not uncommon historically, but it was mostly adopted at the district level and often as a tool for determining which teachers to let go when reductions in force were necessary (there was an ugly racial motive to much of this testing, described in a following section). Post-ANAR, however, many states began adopting testing policies as a tool to ensure that teacher training programs were producing quality teachers for the classroom.²⁸

Teachers' unions also played a role in shaping post-ANAR reform efforts, even though they were skeptical of many of the reform aims. The report contained recommendations that both the American Federation of Teachers (AFT) and the National Education Association (NEA) had previously opposed, yet the unions did not directly voice opposition to the report or its public champions. Later, representatives from both national unions endorsed the recommendations of the 1986 Carnegie report (the NEA noted some reservations). Shortly thereafter, Mary Futrell, then president of the NEA, participated on the Carnegie Board that helped establish NBPTS. Both national unions counseled local units to take stances of moderated opposition to post-ANAR reforms to enable some experimentation (e.g., accommodating merit-pay programs), while exerting selective opposition to elements that were seen as unfair to teachers (e.g., using student test scores in evaluation systems).²⁹ It is noteworthy that unions also generally resisted differentiation within the profession (e.g., through career-ladder programs), even though these ideas were endorsed by university educators (among their closest allies), and the Holmes Group had included them in its report.

Looking back over the policy developments in the years following ANAR's publication, I see a clear delineation in the responses. The earliest experiments—in alternative certification, loan forgiveness or teacher scholarship programs, compensation reform—were responding primarily to the report's calls for bringing people into the teaching profession and keeping them there. In other words, they were attempting reform through focusing on quantity. To be sure, these efforts also sought to promote quality, but it was by attracting high-achieving people (which the teaching workforce was lacking) through either lowering barriers into the field or making the payoff of teaching more attractive. The later wave of policy responses, following the publication of the Holmes and Carnegie reports, sought to redefine the responsibilities of teachers in schools and the training that led to that new aspirational role. This group prioritized the quality

provision of instruction, the expertise that teachers provided in schools, and the qualification process necessary to develop those skills; the quantity of teachers was a secondary concern.

In other words, both camps were attempting to improve the state of the teacher workforce, but they were doing so through two different objectives. The quantity-based approach might be called reform from the “outside in,” since reform efforts from this side primarily focused on bringing outsiders into the classroom. The quality-based approach, conversely, could be called reform from the “inside out,” since it prioritized the continual development of skills starting in the preservice training period, proving those skills through licensure, and then developing those teachers into becoming master teachers.

Though these two different methods for reform started as complementary approaches to the same call for reform, they evolved with time. *ANAR* galvanized bipartisan support temporarily, but the different approaches to workforce reform expressed in the document slowly developed their own disparate sets of stakeholder groups. For example, the creation of easier pathways into teaching and bubbling opportunities for high performers enabled Teach for America (TFA) to take root in the late 1980s and flourish in the years after. TFA’s outside-in approach attracted a set of corporate philanthropists and funders who were interested in disrupting the status quo in public education. On the other side, the alignment of teacher training programs and teachers’ unions was highly critical of TFA, as their model was antithetical to their reforms from the inside out.

Over time, these different approaches and their respective camps developed into ongoing conflict that, at times, felt like a marriage that soured. The initial days of harmony and different approaches to tackle the same problem gave way to bickering, then patronizing dismissals, and then developed into feelings of disdain. In the early days of the No Child Left Behind era (during the 2000s), there may have been some pretense of tolerance on both sides, but the working relationship came to a stalemate during the Race to the Top era (2009-15) with its focus on testing, school turnaround, and teacher evaluation. And since the enactment of the Every Student Succeeds Act, it seems to me that both sides are forging their own divergent paths toward education reform.

But failing to work together would be a loss in my view, because we have gained valuable insights from the reform ideas expressed on both sides. The next two sections highlight the evidence on the policy innovations produced from these two sides.

THE EVIDENCE ON OUTSIDE-IN REFORMS

This section summarizes the research on innovations that came from this outside-in approach to reforming the teacher pipeline into the classroom: alternative certification, defraying the cost of college, and compensation reforms to attract and retain teachers. As I consider these policies, the motivating questions here are: Do these policies succeed in bringing in more

teachers? If so, what types? Do these teachers help stabilize the workforce, especially in places with weak teacher pipelines?

ALTERNATIVE CERTIFICATION

Alternative certification policies provide routes into the classroom that do not require an individual to complete a formal degree program in education prior to entry. Critically, the “alternative certification” label does not represent a single model but is an umbrella term covering many different types of programs. The design and rigor of these programs differ based on state licensure requirements (which necessitate differing amounts of student teaching, for example). Programs can also vary by providers, which can range from state colleges of education (which often feature programs adjacent to traditional training programs) to online companies (which tend to offer very different training experiences). In the paragraphs below, I offer an overview of the evidence on alternative certification, though given the variety just described, the evidence may not be representative of any single program.

Although in its infancy in 1983, alternative certification is now a major entry point into teaching, accounting for 18 percent of public school teachers’ points of entry in the 2015–16 school year, the most recent data available.³⁰ Clearly, this is a productive route based on the number of new entrants, but it’s less clear whether these additions are a net benefit to the workforce. Though prior evidence is mixed, it does suggest traditionally trained teachers have a slight edge on producing learning gains in their students in some (but not all) studies.³¹ However, the differences in productivity within routes are significantly greater than differences across routes. In other words, both alternative certification and traditional training routes can produce effective and ineffective teachers; neither route has a clear advantage based on classroom performance.³²

To be sure, there are some alternative certification programs that stand out, with TFA being a prime example. Several randomized controlled trials and dozens of other studies using rigorous empirical methods have consistently found that TFA teachers perform at levels comparable to their colleagues in reading and often perform significantly better in math. They also appear to improve students’ attendance and GPAs.³³ The literature on TFA suggests that it’s primarily the organization’s intense screening practice, which typically selects 10 percent to 20 percent of applicants, that produces these positive outcomes.³⁴

The biggest liability with hiring from alternative certification routes, however, is their high turnover rate. TFA has received much criticism on this point, as an estimated 85 percent of corps members leave the classroom within five years.³⁵ Other alternative entry routes fare better than TFA, though turnover rates among alternatively certified novice teachers are still roughly 10 percentage points higher than otherwise similar traditionally trained teachers. Survey evidence suggests these differences could be mitigated through more robust organizational support from their providers, though most providers’ ongoing support is inadequate.³⁶ Notably, alternatively certified teachers tend to fill teaching vacancies in schools with higher needs, which experience high levels of staff turnover already. Since high turnover in high-need

settings is costly to schools financially and costly to students academically, staffing with alternatively certified teachers is often considered a less-preferred option—a Band-Aid.³⁷

Not all alternative certification programs produce high turnover, though. Teacher residency programs—alternative routes sponsored by school districts that provide intensive apprentice and educational experiences for candidates—are becoming increasingly popular. The extensive training in schools and ongoing support that are built into residency programs result in significantly lower turnover rates in the first years after placement—on the order of about 20 percentage points lower—though they are costly to implement and sustain.³⁸

A 2004 report from the US Department of Education identified common elements of alternative certification programs that showed the most promise. Among these were careful candidate selection and extensive support in the training period and first year of teaching or beyond.³⁹ As I see it, TFA leans heavily on selection to produce standout performances, while teacher residency programs lean heavily on support to produce standout retention. Both are exemplars under the alternative certification umbrella. I would caution, however, against alternative certification programs that provide neither of these. A worrying development in recent years is the sharp rise in teacher candidates entering through for-profit providers. These programs have little incentive to screen candidates and often use online programs with little in-person training or support. Though there is little evidence to date on the efficacy of these programs, it seems that the model itself is a recipe for mediocre performance and high turnover.⁴⁰

SUBSIDIZING THE COST OF COLLEGE

Next, let's look at the track record on service scholarships or loan forgiveness as a vehicle to attract new entrants into the teaching profession. It is well known that college costs have risen dramatically in recent decades, along with student loan balances. Empirical work supports the notion that higher college costs discourage students from pursuing degrees that lead to low-paying fields such as teaching.⁴¹ Thus, it seems plausible that at least some new teachers are attracted to the classroom because of the availability of these programs, though exact numbers are hard to pin down. Because these programs are made widely available, the majority of those who utilize these incentives may have already been intending to pursue teaching as a career, and typically only a minority of participants report that it is influencing their decision to pursue teaching.⁴²

It is also not clear whether the availability of these programs induces high-achieving individuals into teaching, as most enter through traditional training programs and are not typically distinguished in education data. An evaluation of the long-running North Carolina Teaching Fellows program, however, is a notable exception here: those entering teaching through the selective, merit-based scholarship program outperformed their traditionally trained colleagues.⁴³

However, these scholarship or loan forgiveness programs appear to be more influential in shaping where teachers take jobs and how long they stay. Many of these programs require participants to teach in high-need schools or shortage fields. Multiple studies have found

that award recipients have a greater likelihood of teaching in high-need settings and spend a longer duration in these settings than otherwise expected.⁴⁴ Thus, it's clear these programs shore up teacher pipelines in specific areas, even if the effect on the overall teacher supply is less certain.

In their review of the research on service scholarship and loan forgiveness programs, Podolsky and Kini conclude that these programs can be effective ways to strengthen teacher pipelines, offering recommendations for increasing their impact.⁴⁵ These include covering all or large portions of college costs, conditioning benefits on teaching in high-need settings or fields, selecting strong and committed candidates, and limiting the financial consequences to those who cannot fulfill their teaching commitments. I would add to this by recommending that more-generous benefits through scholarships (rather than loan forgiveness) may optimize the impact on genuinely new entrants to the teacher pipeline, rather than subsidizing those already bound for the occupation.

REFORMING TEACHER PAY

The final area of outside-in reform that I address here is whether compensation reform—changing what and how teachers are paid—impacts the teacher pipeline. Without question, simply paying teachers more will induce more individuals into the field and could go a long way toward ending teacher shortages. There would be no need for alternative certification routes or subsidizing college costs if we were paying teachers a high enough wage. The problem with relying exclusively on teacher compensation, however, is that district employers are budget constrained in ways that make large, general wage increases prohibitively expensive.⁴⁶

Though, what if some teachers didn't need more money? After all, most difficult-to-staff vacancies are concentrated in specific high-need fields or settings.⁴⁷ If districts could be flexible in offering compensation to these weak spots in their teacher pipelines, they may be able to overcome staffing shortages with only modest increases in staffing expenses. Empirical evidence suggests such a strategy would help sustain the supply of teachers, even if the targeted pay differentials would need to be quite large.⁴⁸ This strategy is indeed very promising in reducing teacher pipeline challenges. The problem here, again, is that most districts have rigid contracts or other salary structures in place that often hinder this differentiated approach, even if it may be more direct and cost effective.

Teacher pay reform can come in many varieties beyond differential pay for select subjects or settings, and all face similar resistance due to rigid salary structures. Other criticisms of the dominant single-salary schedule are that it rewards teachers excessively for graduate degrees, defers too much compensation until teachers gain experience (including deferring too much into pension systems that too few teachers meaningfully benefit from), does not pay teachers according to their impact on students (either positive or negative), fails to adequately compensate for extra responsibilities teachers take on, and pays male teachers more for otherwise similar work. Making changes to address any one of these shortcomings could

feasibly make a beneficial impact on the teacher pipeline, though the expected impacts would differ depending on exactly what type of pay reform was pursued.

Further, even within these pay reforms, the outcomes often differ from what is predicted. For example, pay-for-performance schemes have been a popular method of experimenting with teacher compensation over the past decade. High-achieving teachers have been shown to earn higher outside wages in comparison to what they earn in schools, and thus pay-for-performance is expected to attract and retain high-performers.⁴⁹ While studies show they are associated with attracting high-achieving individuals, other evidence suggests men tend to disproportionately select into such settings (regardless of their ability to perform), which could undermine fair-compensation concerns.⁵⁰

Tensions between pursuing desired workforce objectives and pay fairness are often at the center of debates about pay reform. This does not mean that pay reform should be avoided. On the contrary, pay reform should be an important tool in shaping the teacher workforce and attracting more to the profession. But with the overlapping tensions and constraints involved, it should not be seen as a silver bullet. The key issue with pay reform, then, is for policymakers and school leaders to clearly prioritize specific workforce objectives, choose the most appropriate pay reform strategy, and then closely monitor responses on the ground to ensure that those objectives are being met.

Finally, I should also note that even though pay appears to be a strong attractant into teaching, it fades in importance once teachers are in schools. Instead, working conditions, leadership quality, and organizational support are often cited as primary reasons for keeping teachers in the classroom.⁵¹ Thus, efforts that pair teacher support with compensation reform are more likely to sustain the pipeline than a focus on salaries alone.

THE EVIDENCE ON INSIDE-OUT REFORMS

This section summarizes the research findings on policy innovations that came from the inside-out approach to reforming the teacher pipeline into the classroom. This includes efforts related to National Board Certification, aligning the teacher preparation experience with the classroom, and mentoring and induction support for new teachers. As I consider these policies, the motivating questions here are: Do these policies show evidence of developing teacher effectiveness among new teachers? And do these policies stabilize the pipeline of incoming teachers, especially in hard-to-staff settings?

NATIONAL BOARD CERTIFICATION

As described above, the NBPTS was established in response to the 1986 Carnegie report. Initially, the organization was envisioned as a gatekeeping entity for the profession, where NBPTS would assess rigor and adherence to best teaching practices for early-career teachers and supersede state-specific teacher licensure requirements.

Eventually, the scope of NBPTS's certification process shifted to focus on providing National Board Certification to experienced teachers who wanted to demonstrate their excellence in the classroom. This change was not trivial. By shifting from a focus on early-career applicants to experienced teachers, the organization had a less direct impact on the teacher pipeline. Also, by focusing on voluntary applicants (not all incoming teachers), the certification has become an important endorsement that is individually valuable to teachers, though its value to the profession and schools is less clear.

The process to become National Board certified (NBC) is intensive. Applicants must complete four different components, which include both computer- and portfolio-based assessments and are often completed over multiple years. The process currently costs a minimum of nearly \$2,000, with additional costs for retaking individual components as necessary.⁵² Clearly, the cost and process alone can be a deterrent to many teachers. Many states offer support to cover certification costs, provide bonuses for NBC teachers once certified, or both.⁵³

Research on NBC teachers has generally found that they are, indeed, relatively more effective than both unsuccessful applicants and non-applicants. The process of getting certified, however, shows no empirical evidence of enhancing applicants' classroom productivity.⁵⁴ In other words, the primary value of NBC is in its signal of individual quality, not its potential to develop skills for the workforce.

This point about signaling is important in understanding another established research finding: NBC teachers tend to be more mobile across the workforce than non-NBC colleagues. Part of NBC teachers' mobility is by design, as many states recognize the credential in license reciprocity policies, leading it to become widely viewed as a *de facto* national teaching license. Teachers already prone to mobility across state lines (e.g., military spouses) find the certification very valuable, though it's not just military spouses who are mobile. Rather, the signal of quality endows all NBC teachers with power in the job market that other teachers do not enjoy, moving them to opportunities that may personally benefit them. After earning certification, NBC teachers become more mobile and are much more likely to sort out of high-need schools toward those in low-need settings.⁵⁵

Thus, the evidence on NBPTS and NBC teachers is generally underwhelming, as far as the incoming teacher pipeline is concerned. Though the certificate does reliably identify effective teachers, the associated mobility that NBC teachers exhibit appears to erode the teacher workforce in settings that already have weak pipelines. Some states and districts offer bonuses to NBC teachers conditional on teaching in high-need settings. Evaluations of these policies suggest that teacher retention (especially among established teachers) is stronger in these high-need settings, which is a notable improvement.⁵⁶ These or similar criteria are necessary for policymakers to leverage NBC teachers to improve the supply and distribution of teachers for students at least as much as the credential personally benefits teachers.

ALIGNING TEACHER PREPARATION WITH THE CLASSROOM

Both *ANAR* and the subsequent Carnegie and Holmes reports promoted better connections between schools of education and public schools. Though the recommendation articulated in *ANAR* (master teachers in training programs) has evidently become a common practice, I do not know of any empirical studies at scale that can confirm whether and how this practice impacts new teachers. Yet an emerging body of empirical evidence has developed adjacent to this *ANAR* idea that explores the importance of the alignment between teacher preparation experiences and teachers' initial placements. This section briefly reviews the promising findings coming from this subfield.

The first key finding is that student teaching experiences can be key developmental experiences for new teacher candidates, though they appear to be underutilized in teacher training programs. For example, placing candidates into schools with high levels of collaboration among teacher colleagues helps prepare them for effective practice and enhances participant retention as they transition into their own classrooms. Yet these types of schools are underrepresented in student teacher placements.⁵⁷ An even stronger predictor of teacher candidate performance is found when student teachers are assigned to highly effective teachers (based on test score value-added estimates) as supervising mentors. Yet again, selecting mentors based on proven classroom performance is not a common practice; rather, matching student teachers with alumni teachers from the same training program appears to be far more frequent.⁵⁸ Thus, better curation of the student teaching experience represents a high-yield, low-cost policy shift in the teacher pipeline.

Second, proximity to teacher training programs is an underappreciated boost to the teacher pipeline. This finding has some easy intuition behind it: student teaching helps familiarize teacher candidates with area employers, and they become go-to candidates when vacancies open. About 15 percent of student teachers get their initial placement in the same school, and about 40 percent are in the same district.⁵⁹ The hitch is that teacher training programs are typically not located in communities that face ongoing staffing challenges; consequently, schools with weak pipelines tend not to be the beneficiaries of the supply lines. Education programs should expand their reach and open teacher supply lines to schools that face ongoing staffing challenges. The pandemic provided a unique opening, showing that with technology, apprenticeships and other formative experiences in the teacher training years can happen virtually over long distances. This is a fruitful strategy that states can encourage to connect supply lines with otherwise disconnected schools across dispersed regions.

And finally, technological advances can also help optimize the teacher training experience. A research team at the University of Virginia has spent several years developing classroom simulation models intended to provide training experiences for teacher candidates. The simulation approach offers many advantages over traditional student teaching models: it creates an opportunity for playback and repeated practice that live student teaching does not. Teacher candidates can be exposed to challenging situations that arise infrequently in real life while limiting children's exposure to poor teaching.⁶⁰ Currently, the researchers recommend the simulation exercises paired with coaching to provide feedback as a supplement to their

student teaching. Though these simulations are still in development, the early evidence suggests a lot of promise in developing quality in novice teachers for the classroom in a delivery model that can be made broadly accessible.

Overall, the evidence here suggests that teacher training experiences can be a meaningful tool in developing teacher quality while teacher candidates are in their preservice period, before they are overwhelmed by the daily demands of teaching. It's useful to contrast these promising results during the preservice period with the underwhelming evidence on professional development efforts for experienced, in-service teachers, as discussed by Thomas S. Dee in chapter 4. Investing in training at the front end clearly has the upper hand here.

MENTORING AND INDUCTION SUPPORT FOR NEW TEACHERS

The final area of inside-out reform is mentoring and related induction practices for new teachers to the workforce. This practice addresses overlapping challenges in the teacher labor market: new teachers are most at risk of leaving within their first five years on the job (this window is especially critical in high-need schools with weak teacher pipelines), and teacher survey responses consistently indicate that the lack of organizational support is a critical factor in choosing to leave a school or the profession.⁶¹

Overall, the empirical evidence on new-teacher mentoring shows generally positive results. A 2011 review concluded that induction supports, including mentoring, benefit new teachers in three ways: strengthening commitment to the profession and increasing retention, enhancing new teachers' instructional practices, and improving student achievement.⁶² A recent meta-analysis focused on mentoring preservice teachers comes to similar conclusions and argues that cognitive modeling (i.e., explicitly providing reasoning for a given teaching practice and demonstrating it) is a key driver of the value behind mentoring.⁶³

Echoing the findings of the student teaching experience above, studies on new-teacher mentoring also emphasize the importance of choosing good mentors. Those with a record of effective teaching, who make their implicit processes of teaching explicit, and who are adequately prepared and given time for the role are the most successful in offering support for new teachers.⁶⁴ Similar findings about the importance of effective coaches are also seen in studies of instructional coaching, which is essentially mentoring for all teachers (regardless of whether they are new).⁶⁵

With all this promising evidence, one might ask, why isn't mentoring and induction done more consistently across schools? The biggest hindrance is the lack of available resources, primarily time—both for the new teacher and the mentor—to devote to the task. Ironically, new teachers in high-need schools that could benefit the most from effective mentoring are less likely to have the time and other resources to support a quality coaching experience. In the conclusion, I offer some ideas that I believe can help overcome these constraints.

RACE AND INCOME CONSIDERATIONS AMONG STUDENTS AND TEACHERS: MISSED OPPORTUNITIES IN 1983

While acknowledging all the policy activity that ANAR prompted on the teacher pipeline, I also want to briefly reflect on a key element that the report omitted. Namely, the report failed to recognize the role of socioeconomic inequalities among students and teachers. These omissions relegated discussions of class and race to a lower status, limiting the integration of these considerations in policy solutions in the wake of the report.

First, the report failed to recognize how important race and income were in dictating students' access to public school resources. Perhaps it was assumed at the time of writing that existing gaps in access were waning with ongoing school integration efforts and thus fading from public consideration. However, with hindsight, it's clear that race and income continue to matter, even forty years later. Importantly, even as financial resources have narrowed over recent decades (a welcome development), gaps in access to instructional resources like teachers continue to lag.⁶⁶

The consequence of this omission is that the report failed to explicitly say that the largest deficiencies in teacher quantity and quality are in high-need, under-resourced school settings. This finding has been documented in multiple districts and states.⁶⁷ Outside of high-need contexts, teacher vacancies tend to be manageable and rarely compromise academic offerings in schools. But these settings, which disproportionately serve students of color, are frequently beset by high teacher turnover. Often, schools must cope with limited teacher supply by hiring from weak applicant pools, employing long-term substitutes, or reducing class offerings in a subject.⁶⁸

A proper remedy, in my view, would be to explicitly identify these settings as particularly understaffed both in quantity and quality and to offer a strategic approach to developing both dimensions in the workforce over time. I offer such a strategic approach in the following section.

Second, the report missed the opportunity to recognize the importance of racial factors in the teacher pipeline. Black teachers were systematically removed from the workforce during the decades following the 1954 *Brown v. Board* decision. They were removed through racially administered reductions in force, forced rehiring for Black teachers, and were pressured out through offering lower wages and fewer professional privileges. Qualified Black teachers were purged from public schools, making room for less qualified White teachers to replace them.⁶⁹

A particularly insidious method for removing Black teachers from the classroom, which continues to have relevance for the teacher pipeline today, is the use of teacher competency tests. The National Teacher Examinations (NTE) program was the Educational Testing Service's predecessor to the Praxis exams commonly used as licensure tests. The widespread adoption of the NTE across (primarily southern) states occurred once policymakers learned to exploit racial differences on the test as a pretext to provide lower wages to Black

teachers or force them out of their positions. In other words, the tests became a preferred method to indirectly discriminate against Black teachers.⁷⁰

The racialized history of gatekeeping in teaching needs to be acknowledged and confronted, even though the NTE has been replaced with the Praxis and other licensure exams. Also, the injustices in this history are not limited to Black teachers only but have impacted the representation of Native American, Hispanic, and Asian groups too. Even if modest progress has been made in recent decades, the racial diversity of the teacher workforce does not align with the racial makeup of the students they serve nor with other professional occupations.⁷¹ The pipeline into teaching continues to advantage White teachers over teachers of all other racial and ethnic backgrounds, and teacher licensure testing is a major chokepoint.⁷² In other words, teacher licensure testing is a policy that deserves closer scrutiny.

Research over the past several decades has now come to conclude that the lack of racial diversity among teachers is hindering public schools and students. A more racially diverse workforce has been shown to improve a variety of outcomes, ranging from test scores to school funding allocations. Students of color stand to benefit most from a more representative teacher workforce, and White students can benefit too. It is for these and many other reasons that my coauthors and I argued in our recent book that “teacher diversity is teacher quality.”⁷³ Of course, hindsight is twenty-twenty, but *ANAR* failed to recognize that teacher diversity could be a key asset and that teacher testing was a liability that prevented the teacher workforce from reaching full productivity. In making this omission, *ANAR* squandered potential progress on narrowing long-standing achievement gaps among students and building a more inclusive American society.

AMERICAN EDUCATION IN 2023 AND LESSONS FROM THE PAST

The state of public schools in 2023 is, on the surface, very different from public schools in 1983. We have just recently witnessed schools shutting down for months on end to limit the spread of COVID-19. The consequences of these shutdowns have been a major setback in achievement for students, significantly widening racial and income-based achievement gaps for the first time in decades.⁷⁴ Schools have also become new battlegrounds in culture wars. Clashes initially started by debates around pandemic reopening decisions and masking policies evolved into explosive school board meetings clashing around systemic racism in US history and LGBT accommodations in schools.⁷⁵

These developments appear to have impacted the teacher pipeline in multiple ways. First, teachers are frontline workers helping to aid children’s learning recovery, and signs point to increased burnout and elevated turnover in the wake of the pandemic. Second, the culture wars appear to have had a chilling effect on teachers.⁷⁶ These challenges are layered on top of a teacher pipeline that was already weakened before the pandemic hit, prompting some analysts to warn of a “perfect storm” forming in the teacher labor market.⁷⁷

The pipeline into teaching is objectively worse than ever, with 161,000 completers of teacher training programs in 2020–21, a decline of more than 40 percent from the 1970 peak in the number of completers at 284,000.⁷⁸ These teacher graduates are now serving a student body that is more than 10 percent larger today than it was in 1970.⁷⁹ The growing cost of college is also a widely perceived barrier for entry into the profession, especially for people of color who are more burdened by debts to get through school. Yet enrollment in colleges has also slowed, offering little reprieve for the state of the teacher pipeline.

Under the surface differences, though, I also observe several parallels between the state of American education in 1983 and the state in which we find ourselves in 2023. In 1983, like today, there was also growing dissatisfaction with the current state of public schools and student enrollments were dropping, ostensibly signaling lower confidence in public schools. In 1983, the halcyon days of the late 1950s and 1960s, when schools were responding to the challenge of Sputnik and organizing around the space race to show economic superiority against communism, were not long ago. Today our golden days might be the initial years following the enactment of No Child Left Behind—when accountability began to focus attention on performance in schools, achievement gaps were falling, and graduation rates were trending upward. It was clear then, as it is now, that our public schools are far from returning to that level of performance in our recent past.

Another fascinating parallel comes from a recent study by Kraft and Lyon on the state of the teaching profession since the 1970s.⁸⁰ Looking at historical trends in teacher prestige, interest in the profession, teacher preparation, and teacher satisfaction, the authors conclude that the teacher workforce is now near or at historic low points. The last time we were in this position was in the early 1980s, right around the release of *ANAR*. In this period, there was a quick upsurge in public support for and interest in the profession, though the authors could not pinpoint exactly what the catalyst was back then. Perhaps part of this turnaround was new messaging about the nobility of the teaching profession (championed by those reforming from the inside out). Or perhaps part of the new interest in teaching was due to easier access into the profession and visible pay reform efforts (thanks to the outside-in reformers). Regardless of the source, though, what it suggests is that a much-improved prognosis for the teacher pipeline may be around the corner.

Looking back and learning lessons from the past can help chart a productive way forward. As with the *ANAR* period, schools today have similar pressures to do many things at once, such as promoting learning recovery, dealing with teacher shortages, expanding teacher diversity, and professionalizing the workforce. All these things are valuable, but these multiple demands may divert focus and attention from completing any one of these objectives. If we respond as we did previously, trying our preferred model of reform everywhere and dismissing all alternative efforts, then both sides of the teacher quality and quantity debate will continue to work at cross purposes. Yet these different approaches can be complementary if we can find a way to productively work together.

Building on the lessons from the past, I offer recommendations in two forms as I conclude. The first is a proposed strategy that attempts to build both quality and quantity through context-specific prioritization. The second is a more aspirational idea that attempts to build capacity into the ranks of college graduates broadly.

BUILDING QUALITY AND QUANTITY THROUGH CONTEXT-SPECIFIC PRIORITIZATION

First, if we want to build quality and quantity simultaneously, we need a strategy to get there. If we can let go of the need to treat all teachers the same, regardless of context or specialization, then perhaps we can create individual pockets of progress that can be protected, then vetted, then expanded and emulated on a broad scale.

In practice, we must develop a systematic plan for workforce management that is sensitive to workforce needs on the ground. The most important context here—the one conspicuously omitted from *ANAR*—is school settings that serve high-need student populations. A secondary context for consideration could be difficult-to-staff subjects such as math, science, or special education; however, I view subject specialization as a second-priority category. For the argument that follows, I focus on high-need settings for simplicity, but the same logic extends to high-need specializations.

The primary concession here is that we recognize that these schools, by serving high-need student populations, have difficulty attracting and retaining teachers. Consequently, these schools will spend disproportionate amounts of money and time recruiting, interviewing, and onboarding new teachers. Even if these schools had the excess capacity to invest in building teacher quality, the high levels of turnover lower the expected return on that investment. In other words, these schools have a problem with teacher quantity first.

I propose that policymakers and school leaders prioritize efforts that build and sustain teacher quantity in these settings. The outside-in options will be most readily applicable. For example, monetary bonuses for teachers or generous service scholarships conditioned on working in high-need settings would be an excellent way to shore up the workforce. Alternative certification pathways will bring more candidates to these settings, though because these teachers tend to exhibit greater turnover and can lead to instability, I caution against relying too much on them. Also, recognizing the variety within alternative certification providers, if alternatively certified candidates are considered, I encourage school leaders to prioritize candidates from programs that invest heavily in selection (like TFA) or ongoing support (like teacher residency programs); avoid candidates coming from programs that do neither.

Some inside-out reforms can also provide valuable benefits to workforce stability and should be considered in these settings too. For example, connecting teacher training programs with otherwise disconnected schools can increase the teacher supply. Bonuses for NBC teachers in high-need settings may be useful, as evidence suggests the retention of quality colleagues could have positive benefits on the turnover of younger, non-NBC peers.

Prioritizing quantity does not mean that we ignore quality in these schools entirely, but we should look to building quality only once policies supporting quantity are firmly in place. In fact, some of the inside-out ideas could help move the needle on both quality and quantity. For example, investing in a strong mentoring program with effective teachers in these schools can lower turnover (and hence the quantity of new teachers needed) while also developing teachers' instructional skills.

The big surprise is that the resources to pursue these policies are within reach for most high-need schools already, but they simply need to be repurposed. Currently, class sizes in high-need schools are smaller than they are in low-need schools in the same districts. Though exact numbers vary across states and school levels, the average high-need high school employs nearly one additional teacher per one hundred students than non-high-need schools in the same district, nationwide.⁸¹ In other words, high-need schools are *consistently overdemanding teacher quantity* in the places that most struggle with anemic teacher pipelines.⁸²

Simply allowing class sizes to rise to meet those in low-need settings would save money on teacher salaries that could be repurposed to fund bonuses for teaching in those settings, generously reward mentors for new teachers, or pursue other policies focused on stabilizing the workforce. And let's be honest, the last teachers hired in high-need schools are most likely to be underprepared and least likely to stay in the profession. So not hiring them is itself a small step toward workforce stability. Even greater cost savings could be realized by allowing class sizes to rise even further.

But what about the benefits of small classes? Yes, it is the case that smaller classes promote greater learning and personalization, and these benefits are observed in both the year of exposure and over the long term. But the same thing is also true of excellent teaching, and then some. Studies have consistently shown that the benefits to maintaining teacher quality outweigh the benefits of lowering class size (increasing quantity) when the marginal teacher is of lower quality.⁸³ This trade-off almost certainly holds in high-need schools, whose marginal hire is more likely to be a long-term substitute than an experienced, well-qualified hire.

I acknowledge that the status quo of small class sizes may be useful to teachers. Perhaps smaller classes make classroom management easier or otherwise reduce teacher burden. However, remember that low class sizes come at the cost of workforce stability in schools already overburdened by this issue. There are often easier, lower-cost ways to reduce teacher burden (e.g., hiring more instructional aides, an often-overlooked resource) than hiring more teachers.⁸⁴

Outside of these high-need settings, where the quantity of the teacher workforce is not a pressing demand, we should prioritize quality-focused enhancements to the pipeline. This is where the promising inside-out ideas become the policies of choice. These are the contexts where new-teacher mentoring with effective mentors should be the default. We should be focusing on providing enriching experiences to enhance the quality of the new teachers in these schools, rather than tweaking things like pay or loan forgiveness to attract better teachers from the outside.

The overarching idea here is that both the inside-out and outside-in approaches offer useful ways to reform the teacher workforce. It's not clear that either side has a monopoly on the right approach here, and if we insist on doing all of one but not the other, we will just continue to snuff out opportunities for success. But if we prioritize the outside-in policies in high-need settings where quantity is most constrained and the inside-out policies where we want to build capacity in the existing workforce, we are more likely to succeed in simultaneously improving both quantity and quality.

BUILDING A CADRE OF TALENT READY FOR TEACHING

This final recommendation to bolster the teacher pipeline focuses on expanding the reach of service scholarships. As discussed in the review of outside-in ideas, the intuition here is that, since we have good reason to believe the cost of college is deterring young people from teacher training, defraying the cost of college could be a powerful attractant. Though service scholarship and loan forgiveness programs have become common across states, most offer relatively modest benefits (typically a few thousand dollars per year of teaching). I recommend leveraging this tool much more aggressively to build out the teacher pipeline and perhaps even build excess capacity.

I recommend state policymakers set up a menu of scholarship options for college students that ties scholarship aid with taking a recommended core battery of education classes. I'll consider each of these elements in turn.

First, what is meant by "a menu of scholarship options"? I recommend trading off more generous scholarship support with differing levels of commitment to teaching and teaching in high-need schools. Recall that one of the potential issues with service scholarship or loan forgiveness programs is how punitive consequences are for not fulfilling the program's teaching commitments. Providing a transparent menu of options for students while in their undergraduate years allows students to select into the level of aid they want with the level of commitment they are comfortable with. For example, those ready to take the education classes and commit to four years of teaching in a high-need setting would qualify for maximum aid (e.g., free tuition), while those who do not commit to any teaching but simply take the core battery of education classes receive the minimum aid (say, a 20 percent reduction in tuition costs). A tier or two in between these extremes offers a middle level of aid while requiring some teaching but offers flexibility on the length or the requirement to teach in a high-need setting. Perhaps the program could offer an option to convert teaching commitments after entering the workforce where one year in a high-need setting receives equal credit to two years teaching in a low-need setting. Also, selectivity could be built into the menu of options: for example, all students with at least a 3.0 high school GPA qualify for the program, and those with a 3.5 or higher get an extra 5 percent discount. Aid amounts or selectivity criteria could also be used to attract more candidates in specific subjects (e.g., offering especially generous aid and lowering selection criteria for math and science majors and doing the opposite for oversupplied specializations like elementary education).

The specifics could be tweaked, but the main idea is that generous aid plus commitment levels that students can choose from will attract many more people to education and a career

in teaching. Note that, by design, some people will benefit from minimal aid, taking education classes without committing to teaching. This is an acceptable trade in my view, since it gets people engaged with education in a meaningful way. Note that taking education classes alone is a commitment, and in the process, students can explore their own interests in teaching. Some students may find themselves willing to increase their commitment to teaching, while others may realize teaching is not a fit for them; both outcomes can be seen as a positive for the teacher pipeline. And taking the long view, many people pursue other careers for a time before choosing to enter teaching at a later point. This training would be a beneficial head start that can facilitate entering through alternative certification routes at a later point in time.

Now let's consider the "recommended core battery of education classes." Courses in this core should be foundational education courses aligned with the student's intended major, be scientifically based, and provide opportunities for the student to get some early live practice experiences in front of students (even simulated ones). The National Council on Teacher Quality's ongoing Teacher Prep Review project evaluates programs across the country on their course requirements and the syllabi of these courses to judge their alignment with best practices.⁸⁵ Also, UTeach is a program for attracting math and science majors into teaching; it has a model that emphasizes content mastery, foundational pedagogical learning, with early practice experiences.⁸⁶ Both of these resources and others can help steer policymakers' judgments about what belongs in the battery of courses. Also, to avoid undesired institutional variation, the composition of the core battery should be determined by a state committee and be consistently offered across all training programs in the state.

Importantly, an expansive scholarship program like this is intended to build capacity for teaching in a broad range of students, directly building pressure into the teacher pipeline. Given the uneven college debt burdens faced by people of color, I anticipate this program could be especially useful in attracting teachers of color into the profession. Importantly, since teachers of color disproportionately staff high-need schools, this would be a major boon to teacher supply in these settings.⁸⁷

I acknowledge this scholarship idea would require significantly more resources than are currently allocated to the teacher pipeline. But with the deficient quality and quantity in the teacher pipeline currently, a surge of investment into it is probably the only way out of the predicament.

A STRONGER AND MORE ROBUST TEACHER PIPELINE COULD BE JUST AROUND THE CORNER

In summary, the teacher pipeline is inadequate now and will continue to be a drag on schools—especially those serving high-need students—without significant policy intervention. But I am optimistic that a much healthier pipeline is within reach. We don't need to think of new solutions; we just need to deploy the ones we have with more strategy and purpose. These lessons from the past offer a road map to get there.

I thank Ayanna Platt, Elizabeth Gellman, Nicolas Zerbino, and Sana Sinha for excellent research assistance in preparing this chapter. Macke Raymond and Steve Bowen also provided key editorial feedback that strengthened the argument presented here.

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.

I appreciated the approach of reflecting on “strategically combining promising reform efforts” and implementing the reforms in “context-dependent ways.” While the historical examination was interesting, I also think some of the contextual reforms could be further explored. For example, the section on aligning teacher preparation with the classroom has great promise if current evidence-based programs are outlined. One such model is the professional development school, which has been modernized to have candidates work directly in schools at all phases of their program with dedicated personnel from the district mentoring, advising, and coordinating.

How mentors are trained, supported, and matched is a key factor in the success of beginning teachers. I also appreciated the focus on missed opportunities for equitable access to effective teachers for all students. Equity labs are one way to help districts and schools understand the “who” of the student-teacher relationship and how getting qualified teachers who represent the community can aid in outcomes.

Finding ways to keep teachers in the classroom part time while giving them other opportunities for leadership has aided Utah in keeping some of our most effective teachers in the classroom. Endorsement for instructional coaching with state support is another strategy that is paying dividends in teacher capacity and retention. I feel that we are beyond the surface level and have discovered related contextual strategies that are working. I think the three pillars of quality, quantity, and context are spot-on and we should highlight programs that are really working.

—Dr. Sydnee Dickson, state superintendent of public instruction for Utah

This paper provides both rich historical context and actionable steps for the future to address the complexities of teacher recruitment and retention. The author acknowledges the need to address both the quantity and the quality of our teacher workforce—doing one without consideration for the other would be a “loss,” and I agree. While progress has certainly been made, it is clear there is more work to be done. Understanding that the teacher is the single greatest school-level factor that impacts student performance, future investments should

ensure that all children—especially those who struggle the most—have access to highly skilled and trained educators.

After spending several months with business leaders who are helping solve teacher recruitment and retention issues in our state, I have learned that it is challenging to get buy-in from stakeholders who may question the “step and lane” model. And, while teachers certainly desire and deserve an increase in salary, most tell me that more than monetary fulfillment is needed to keep them in the classroom. The stressors they cite have escalated significantly since *ANAR* was published.

I couldn’t agree more with the paper’s conclusion that significant policy intervention is needed; the policies presented remain relevant and must be strategically prioritized, and a much healthier pipeline is within reach. I would also suggest that for true comprehensive reform to occur, we must examine or at least acknowledge where policy levers fall short in addressing some of the top concerns of today’s classroom teachers.

—Dr. Margie Vandeven, commissioner of education, Missouri

NOTES

1. National Commission on Excellence in Education (NCEE), *A Nation at Risk: The Imperative for Educational Reform* (Washington, DC: US Government Printing Office, 1983). <https://files.eric.ed.gov/fulltext/ED226006.pdf>; references to specific items or page numbers reflect the original pagination from this version.
2. Dana Goldstein, *The Teacher Wars: A History of America’s Most Embattled Profession* (New York: Doubleday, 2014).
3. I want to be clear here that I am not an education historian. While I endeavor to reference the historical record as best I can, I make several speculations and inferences about connections between historical developments rather than laying out a clean causal connection. My argument here is mostly an attempt to make sense of the past and offer some perspectives about teacher policy moving forward, rather than to offer a detailed history.
4. National Center for Education Statistics, “Table 208.20. Public and Private Elementary and Secondary Teachers, Enrollment, Pupil/Teacher Ratios, and New Teacher Hires: Selected Years, Fall 1955 through Fall 2031,” *Digest of Education Statistics*, March 2023, https://nces.ed.gov/programs/digest/d22/tables/dt22_208.20.asp.
5. T. C. Venable, “Declining SAT Scores: Some Unpopular Hypotheses,” *Phi Delta Kappan* 62, no. 6 (February 1981): 443–45, <http://www.jstor.org/stable/20385940>.
6. David J. Armor, *Forced Justice: School Desegregation and the Law* (New York: Oxford University Press, 1995).
7. National Center for Education Statistics, *Digest of Education Statistics*.
8. Loraine M. McDonnell and Anthony Pascal, “Teacher Unions and Educational Reform,” Center for Policy Research in Education; RAND Corporation, April 1988, <https://files.eric.ed.gov/fulltext/ED293837.pdf>.
9. Whether this relative decline was due to pull factors (like attractive wages in other industries) or push factors (compressed teacher pay scales from unionization) has been the subject of academic debate; Sean P. Corcoran, William N. Evans, and Robert M. Schwab, “Changing Labor-Market Opportunities for Women and the Quality of Teachers, 1957–2000,” *American Economic Review* 94, no. 2 (May 2004): 230–35, <http://www.jstor.org/stable/3592888>; Caroline M. Hoxby and Andrew Leigh, “Pulled Away or Pushed Out? Explaining the Decline of Teacher Aptitude in

- the United States," *American Economic Review* 94, no. 2 (May 2004): 236–40, <http://www.jstor.org/stable/3592889>.
10. James S. Coleman, Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York, *Equality of Educational Opportunity*, National Center for Educational Statistics (Washington, DC: US Government Printing Office, 1966), <https://files.eric.ed.gov/fulltext/ED012275.pdf>.
 11. David J. Armor, "The Evidence on Busing. Research Report," *Public Interest* 28 (Summer 1972): 90–126, https://www.nationalaffairs.com/public_interest/detail/the-evidence-on-busing.
 12. Eric A. Hanushek, "Throwing Money at Schools," *Journal of Policy Analysis and Management* 1, no. 1 (Fall 1981): 19–41, <https://doi.org/10.2307/3324107>.
 13. Wayne Dixon Lett, "An Analysis of the Influences Affecting the Standards for Certifying Public School Teachers in Virginia," PhD diss., College of William & Mary, School of Education, 1984, available at W&M ScholarWorks, <https://scholarworks.wm.edu/cgi/viewcontent.cgi?article=1623&context=etd>.
 14. Kenneth M. Zeichner and Ann K. Schulte, "What We Know and Don't Know from Peer-Reviewed Research about Alternative Teacher Certification Programs," *Journal of Teacher Education* 52, no. 4 (September 2001): 266–82, <https://doi.org/10.1177/0022487101052004002>.
 15. Though much progress has been made since the latter half of the 2010s, the most recent data shows that computer science still has a way to go to be universally accessible to high schoolers—see Michael Hansen and Nicolas Zerbino, "Exploring the State of Computer Science Education amid Rapid Policy Expansion," Brookings Institution, Washington, DC, April 11, 2022, <https://www.brookings.edu/articles/exploring-the-state-of-computer-science-education-amid-rapid-policy-expansion>.
 16. See NCEE, *A Nation at Risk*, 26–30, for a detailed breakdown of recommendations. Those demanding more teacher quantity are Content recommendations #5 and #6 (p. 26), Standards and Expectations recommendation #2 (p. 27), and Time recommendations #1 and #3 (p. 29). Those demanding more teacher quality are Standards and Expectations recommendation #3 (p. 28) and Time recommendation #4 (p. 29). Only one recommendation, Time recommendation #7 (p. 30), which recommends reducing administrative burdens for teachers, implicitly demanded less of the teacher workforce; all other recommendations implicitly demanded more quality, more quantity, or both.
 17. Zeichner and Schulte, "What We Know and Don't Know."
 18. Warren S. Corbin, "Alternative Certification Programs: Problems and Prospects," *Clearing House: A Journal of Educational Strategies, Issues and Ideas* 65, no. 4 (1992): 241–44.
 19. Robert Shireman, Jen Mishory, and Sandy Baum, "Appendix 6: Summary and History of California Financial Aid Programs," in *Expanding Opportunity, Reducing Debt: Reforming California Student Aid* (New York: The Century Foundation, 2018), https://production-tcf.imgix.net/app/uploads/2018/03/03154827/CSAC_Appendix-6.pdf.
 20. *A Legacy of Inspired Educators: A Report on the North Carolina Teaching Fellows Program, 1986–2015*, Public School Forum of North Carolina, February 2015, https://www.ncforum.org/wp-content/uploads/2016/09/PSF_TeachingFellowsReport_HRsingle.pdf.
 21. Anne Podolsky and Tara Kini, *How Effective Are Loan Forgiveness and Service Scholarships for Recruiting Teachers?* (Policy Brief), Learning Policy Institute, Palo Alto, CA, 2016, <https://learningpolicyinstitute.org/product/how-effective-are-loan-forgiveness-and-service-scholarships-recruiting-teachers>.
 22. "Findings from the Condition of Education 1994, No. 3: America's Teachers Ten Years After 'A Nation at Risk,'" National Center for Education Statistics (NCES) no. 95-766, August 1994, <https://nces.ed.gov/pubs95/95766.pdf>.
 23. Jean Protsik, "History of Teacher Pay and Incentive Reforms," Consortium for Policy Research in Education, February 1995, <https://files.eric.ed.gov/fulltext/ED380894.pdf>.
 24. Harry P. Hatry, John M. Greiner, and Brenda G. Ashford, *Issues and Case Studies in Teacher Incentive Plans*, 2nd ed. (Washington, DC: Urban Institute, 1994), 131–36.

25. A special issue of *Teacher Education Quarterly*, released in early 1987, focused entirely on “teacher education in the aftermath of Holmes and Carnegie,” demonstrating the urgency and importance of these new reports for the field; see Alan H. Jones, “Introduction,” *Teacher Education Quarterly* 14, no. 1 (1987): 2–3, <https://www.jstor.org/stable/i23474602>.
26. *Tomorrow’s Teachers: A Report of The Holmes Group*, Holmes Group Inc. (April 1986), <https://files.eric.ed.gov/fulltext/ED270454.pdf>.
27. *A Nation Prepared: Teachers for the 21st Century, Task Force on Teaching as a Profession*, Carnegie Corp. of New York, 1986, <https://eric.ed.gov/?id=ED268120>.
28. Ann Jarvela Wilson, *Knowledge for Teachers: The Origin of the National Teacher Examinations Program*, paper presented at the Annual Meeting of the American Educational Research Association (Chicago, March 31–April 4, 1985), <https://files.eric.ed.gov/fulltext/ED262049.pdf>.
29. McDonnell and Pascal, “Teacher Unions.”
30. *Characteristics of Public School Teachers Who Completed Alternative Route to Certification Programs* Condition of Education, National Center for Education Statistics, US Department of Education, Institute of Education Sciences, 2022, <https://nces.ed.gov/programs/coe/indicator/tlc>.
31. Christopher Redding and Thomas M. Smith, “Easy In, Easy Out: Are Alternatively Certified Teachers Turning Over at Increased Rates?,” *American Educational Research Journal* 53, no. 4 (August 2016): 1086–125.
32. Thomas J. Kane, Jonah E. Rockoff, and Douglas O. Staiger, “What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City,” *Economics of Education Review* 27, no. 6 (2008): 615–31, <https://doi.org/10.1016/j.econedurev.2007.05.005>.
33. Ben Backes and Michael Hansen, “The Impact of Teach for America on Non-Test Academic Outcomes,” *Education Finance and Policy* 13, no. 2 (Spring 2018): 168–93, https://doi.org/10.1162/edfp_a_00231.
34. Will Dobbie, “Teacher Characteristics and Student Achievement: Evidence from Teach for America,” July 2011, published at Harvard Kennedy School, <https://www.hks.harvard.edu/publications/teacher-characteristics-and-student-achievement-evidence-teach-america>.
35. Morgaen L. Donaldson and Susan Moore Johnson, “TFA Teachers: How Long Do They Teach? Why Do They Leave?,” *Education Week*, October 4, 2011, <https://www.edweek.org/teaching-learning/opinion-tfa-teachers-how-long-do-they-teach-why-do-they-leave/2011/10>.
36. Redding and Smith, “Easy In, Easy Out.”
37. Nicole Simon and Susan Moore Johnson, “Teacher Turnover in High-Poverty Schools: What We Know and Can Do,” *Teachers College Record* 117, no. 3 (March 2015): 1–36; Astrid Galván, “Hiring Uncertified Teachers Is a Band-Aid Response to Shortages, Experts Say,” *Axios*, August 22, 2023, <https://www.axios.com/politics-policy/shortage-teachers-hiring-uncertified-2d3d7ce0-094c-4a37-93e0-6e53b38cc6ca.html>.
38. Lauren Worley and Nicolas Zerbino, “Teacher Residencies Offer Compelling Solution to Staffing Shortages, Although at a Large Investment,” Brookings Institution, May 22, 2023, <https://www.brookings.edu/articles/teacher-residencies-offer-compelling-solution-to-staffing-shortages-although-at-a-large-cost>.
39. *Alternative Routes to Teacher Certification*, US Department of Education, Office of Innovation and Improvement, November 2004, <https://www2.ed.gov/admins/tchrqual/recruit/altroutes/report.pdf>.
40. Ben Backes and Michael Hansen, “Teach for America Is Shrinking—Is This Cause for Celebration?,” Brookings Institution, January 31, 2023, <https://www.brookings.edu/articles/teach-for-america-is-shrinking-is-this-cause-for-celebration>.
41. Jesse Rothstein and Cecilia Elena Rouse, “Constrained after College: Student Loans and Early-Career Occupational Choices,” *Journal of Public Economics* 95, no. 1–2 (2011): 149–63, <https://doi.org/10.1016/j.jpubeco.2010.09.015>.

42. Podolsky and Kini, *Loan Forgiveness*.
43. Gary T. Henry, Kevin C. Bastian, and Adrienne A. Smith, "Scholarships to Recruit the 'Best and Brightest' into Teaching: Who Is Recruited, Where Do They Teach, How Effective Are They, and How Long Do They Stay?," *Educational Researcher* 41, no. 3 (April 2012): 83-92, <https://doi.org/10.3102/0013189X12437202>.
44. Li Feng and Tim R. Sass, "The Impact of Incentives to Recruit and Retain Teachers in 'Hard-to-Staff' Subjects," *Journal of Policy Analysis and Management* 37, no. 1 (Winter 2018): 112-35, <https://doi.org/10.1002/pam.22037>; *Answering the Call for Equitable Access to Effective Teachers: Lessons Learned from State-Based Teacher Preparation Efforts in Georgia, Indiana, Michigan, New Jersey, and Ohio* (Princeton, NJ: The Woodrow Wilson National Fellowship Foundation, 2015).
45. Podolsky and Kini, *Loan Forgiveness*.
46. Michael Hansen, "Ways to Reform Teacher Compensation, Beyond a \$60,000 Minimum Salary," Brookings Institution, March 8, 2023, <https://www.brookings.edu/articles/ways-to-reform-teacher-compensation-beyond-a-60000-minimum-salary>.
47. Thomas S. Dee and Dan Goldhaber, "Understanding and Addressing Teacher Shortages in the United States," Brookings Institution, April 2017, https://www.brookings.edu/wp-content/uploads/2017/04/es_20170426_understanding_and_addressing_teacher_shortages_in_us_pp_dee_goldhaber.pdf.
48. Charles Clotfelter, Elizabeth Glennie, Helen Ladd, and Jacob Vigdor, "Would Higher Salaries Keep Teachers in High-Poverty Schools? Evidence from a Policy Intervention in North Carolina," *Journal of Public Economics* 92, no. 5-6 (June 2008): 1352-70, <https://doi.org/10.1016/j.jpubeco.2007.07.003>.
49. Matthew M. Chingos and Martin R. West, "Do More Effective Teachers Earn More Outside the Classroom?," *Education Finance and Policy* 7, no. 1 (Winter 2012): 8-43, https://doi.org/10.1162/EDFP_a_00052.
50. Michael Jones and Michael T. Hartney, "Show *Who* the Money? Teacher Sorting Patterns and Performance Pay across US School Districts," *Public Administration Review* 77, no. 6 (November/December 2017): 919-31, <https://doi.org/10.1111/puar.12789>; Diana Quintero, Michael Hansen, and Nicolas Zerbino, "Uncovering the Sources of Gender Wage Gaps among Teachers: The Role of Compensation off the Salary Schedule," EdWorkingPaper 23-737, Annenberg Institute at Brown University, <https://doi.org/10.26300/q8hd-ws21>.
51. Nicole S. Simon and Susan Moore Johnson, "Teacher Turnover in High-Poverty Schools: What We Know and Can Do," *Teachers College Record* 117, no. 3 (March 2015): 1-36, <https://doi.org/10.1177/016146811511700305>.
52. See the NBPTS website for more details about the certification process and cost: "National Board Certification," National Board for Professional Teaching Standards, <https://www.nbpts.org/certification> (accessed September 21, 2023)
53. Education Commission of the States, *National Board Certification and the Teaching Workforce: 2020 Update* (Denver: Education Commission of the States, 2020), <https://www.ecs.org/wp-content/uploads/SIR-NBCT.pdf>
54. Dan Goldhaber and Emily Anthony, "Can Teacher Quality Be Effectively Assessed? National Board Certification as a Signal of Effective Teaching," *Review of Economics and Statistics* 89, no. 1 (February 2007): 134-50, <https://doi.org/10.1162/rest.89.1.134>.
55. Dan Goldhaber, "National Board Teachers Are More Effective, But Are They in the Classrooms Where They're Needed the Most?," *Education Finance and Policy* 1, no. 3 (Summer 2006): 372-82, <https://doi.org/10.1162/edfp.2006.1.3.372>; Dan Goldhaber and Michael Hansen, "National Board Certification and Teachers' Career Paths: Does NBPTS Certification Influence How Long Teachers Remain in the Profession and Where They Teach?," *Education Finance and Policy* 4, no. 3 (Summer 2009): 229-62, <https://doi.org/10.1162/edfp.2009.4.3.229>.
56. James Cowan and Dan Goldhaber, "Do Bonuses Affect Teacher Staffing and Student Achievement in High Poverty Schools? Evidence from an Incentive for National Board Certified

Teachers in Washington State," *Economics of Education Review* 65 (August 2018): 138–52, <https://doi.org/10.1016/j.econedurev.2018.06.010>; Ana M. Elfers and Margaret L. Plecki, "Results of a State Incentive Program on the Supply and Distribution of National Board Certified Teachers," *Leadership and Policy in Schools* 13, no. 2 (2014): 147–68, <https://doi.org/10.1080/15700763.2014.901392>.

57. Matthew Ronfeldt, Dan Goldhaber, James Cowan, Emanuele Bardelli, Joy Johnson, and Christopher Daniel Tien, "Identifying Promising Clinical Placements Using Administrative Data: Preliminary Results from ISTI Placement Initiative Pilot," CALDER Working Paper no. 189 (April 2018), <https://caldercenter.org/publications/identifying-promising-clinical-placements-using-administrative-data-preliminary-results>.
58. Dan Goldhaber, John Krieg, and Roddy Theobald, "Effective Like Me? Does Having a More Productive Mentor Improve the Productivity of Mentees?," CALDER Working Paper no. 208-1118-1 (November 2018), <https://caldercenter.org/publications/effective-me-does-having-more-productive-mentor-improve-productivity-mentees>.
59. Dan Goldhaber, John Krieg, and Roddy Theobald, "Rethinking the Geography of Student-Teaching Placements in a Post-COVID-19 World," Brookings Institution, February 8, 2021, <https://www.brookings.edu/articles/re-thinking-the-geography-of-student-teaching-placements-in-a-post-covid-19-world>.
60. Julie Cohen and Vivian Wong, "Using Classroom Simulators to Transform Teacher Preparation," Brookings Institution, November 17, 2021, <https://www.brookings.edu/articles/using-classroom-simulators-to-transform-teacher-preparation>.
61. Tuan D. Nguyen, Lam D. Pham, Michael Crouch, and Matthew G. Springer, "The Correlates of Teacher Turnover: An Updated and Expanded Meta-Analysis of the Literature," *Educational Research Review* 31 (November 2020): 100355, <https://doi.org/10.1016/j.edurev.2020.100355>.
62. Richard M. Ingersoll and Michael Strong, "The Impact of Induction and Mentoring Programs for Beginning Teachers: A Critical Review of the Research," *Review of Educational Research* 81, no. 2 (June 2011): 201–33, <https://doi.org/10.3102/0034654311403323>.
63. Sog Yee Mok and Fritz C. Staub, "Does Coaching, Mentoring, and Supervision Matter for Pre-service Teachers' Planning Skills and Clarity of Instruction? A Meta-Analysis of (Quasi-) Experimental Studies," *Teaching and Teacher Education* 107 (November 2021): 103484, <https://doi.org/10.1016/j.tate.2021.103484>.
64. Andrew J. Hobson, Patricia Ashby, Angi Malderez, and Peter D. Tomlinson, "Mentoring Beginning Teachers: What We Know and What We Don't," *Teaching and Teacher Education* 25, no. 1 (January 2009): 207–16, <https://doi.org/10.1016/j.tate.2008.09.001>.
65. Matthew A. Kraft, David Blazar, and Dylan Hogan, "The Effect of Teacher Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence," *Review of Educational Research* 88, no. 4 (August 2018): 547–88, <https://doi.org/10.3102/0034654318759268>.
66. Michael Hansen and Nicolas Zerbino, "Does Teacher Sorting Contribute to Financial Inequalities?" chapter 2 in *Do School Districts Allocate More Resources to Economically Disadvantaged Students?*, Brookings Institution, December 16, 2022, <https://www.brookings.edu/articles/chapter-2-does-teacher-sorting-contribute-to-financial-inequalities>.
67. Eric Isenberg, Jeffrey Max, Philip Gleason, Liz Potamites, Robert Santillano, Heinrich Hock, and Michael Hansen, *Access to Effective Teaching for Disadvantaged Students* (Washington, DC: Institute of Education Sciences and US Department of Education, 2013), <https://files.eric.ed.gov/fulltext/ED544345.pdf>; Dan Goldhaber, Lesley Lavery, and Roddy Theobald, "Uneven Playing Field? Assessing the Teacher Quality Gap Between Advantaged and Disadvantaged Students," *Educational Researcher* 44, no. 5 (June/July 2015): 293–307, <https://doi.org/10.3102/0013189X15592622>.
68. Anne Podolsky and Leib Sutcher, "California Teacher Shortages: A Persistent Problem" (Washington, DC: Learning Policy Institute, 2016), <https://learningpolicyinstitute.org/product/ca-teacher-shortage-persistent-problem-brief>.

69. Leslie T. Fenwick, *Jim Crow's Pink Slip: The Untold Story of Black Principal and Teacher Leadership* (Cambridge, MA: Harvard Education Press, 2022).
70. See discussion of this history in Seth Gershenson, Michael Hansen, and Constance A. Lindsay, *Teacher Diversity and Student Success: Why Racial Representation Matters in the Classroom* (Cambridge, MA: Harvard Education Press, 2021), chapter 3.
71. Michael Hansen and Diana Quintero, "The Diversity Gap for Public School Teachers Is Actually Growing Across Generations," Brookings Institution, March 7, 2019, <https://www.brookings.edu/articles/the-diversity-gap-for-public-school-teachers-is-actually-growing-across-generations>.
72. Hannah Putman, Michael Hansen, Kate Walsh, and Diana Quintero, "High Hopes and Harsh Realities: The Real Challenges to Building a Diverse Workforce," Brookings Institution, August 2016, https://www.brookings.edu/wp-content/uploads/2016/08/browncenter_20160818_teacherdiversityreportpr_hansen.pdf.
73. Gershenson, Hansen, and Lindsay, *Teacher Diversity*.
74. Dan Goldhaber, Thomas J. Kane, Andrew McEachin, Emily Morton, Tyler Patterson, and Douglas O. Staiger, "The Educational Consequences of Remote and Hybrid Instruction during the Pandemic," *American Economic Review: Insights* 5, no. 3 (September 2023): 377-92, <https://www.aeaweb.org/articles?id=10.1257/aeri.20220180&from=f>.
75. Jon Valant, "It's Time to Start Voting in Your Local School Board Elections," Brookings Institution, November 1, 2021, <https://www.brookings.edu/articles/its-time-to-start-voting-in-your-local-school-board-elections>.
76. Mica Pollock and John Rogers, with Alexander Kwako, Andrew Matschiner, Reed Kendall, Cicely Bingener, Erika Reece, Benjamin Kennedy, and Jaleel Howard, *The Conflict Campaign: Exploring Local Experiences of the Campaign to Ban "Critical Race Theory" in Public K-12 Education in the US, 2020-2021* (Los Angeles: UCLA's Institute for Democracy, Education, and Access, 2022).
77. "A Policy Agenda to Address the Teacher Shortage in US Public Schools," Economic Policy Institute, October 15, 2020, <https://www.epi.org/publication/a-policy-agenda-to-address-the-teacher-shortage-in-u-s-public-schools>.
78. "Enrollment Projections Suggest Teacher Shortage in Late 1980s," *Education Week*, November 1982, <https://www.edweek.org/education/enrollment-projections-suggest-teacher-shortage-in-late-1980s/1982/11>; "2022 Title II Report: Academic Year 2020-21 Data," US Department of Education, <https://title2.ed.gov/Public/Home.aspx>.
79. Digest of Education Statistics 2021, Table 201.10, "Historical Summary of Public Elementary and Secondary School Statistics: Selected Years, 1869-70 through 2018-19," https://nces.ed.gov/programs/digest/d21/tables/dt21_201.10.asp.
80. Matthew A. Kraft and Melissa Arnold Lyon, "The Rise and Fall of the Teaching Profession: Prestige, Interest, Preparation, and Satisfaction over the Last Half Century," EdWorkingPaper 22-679, Annenberg Institute at Brown University, 2022, <https://doi.org/10.26300/7b1a-vk92>.
81. David S. Knight, "Are School Districts Allocating Resources Equitably? The Every Student Succeeds Act, Teacher Experience Gaps, and Equitable Resource Allocation," *Educational Policy* 33, no. 4 (June 2019): 615-49, <https://doi.org/10.1177/0895904817719523>; Michael Hansen and Nicolas Zerbino, "Does Teacher Sorting Contribute to Financial Inequalities?," chapter 2 in *Do School Districts Allocate More Resources to Economically Disadvantaged Students?*, Brookings Institution, December 16, 2022, <https://www.brookings.edu/articles/chapter-2-does-teacher-sorting-contribute-to-financial-inequalities>.
82. A similar logic applies to class size differences across subjects. Math and science subjects are more difficult to staff than English or social studies, and allowing for larger math and science classes (especially in secondary grades) would be a straightforward application of the logic described in this section. However, class sizes generally do not systematically vary within schools by subject.

83. Christopher Jepsen and Steven Rivkin, "Class Size Reduction and Student Achievement," *Journal of Human Resources* 44, no. 1 (January 2009): 223–50, <https://doi.org/10.3368/jhr.44.1.223>; Michael Hansen, "Right-Sizing the Classroom: Making the Most of Great Teachers," CALDER Working Paper no. 110 (January 2014).
84. Simon Calmar Andersen, Louise Beuchert, Helena Skyt Nielsen, and Mette Kjærgaard Thomsen, "The Effect of Teacher's Aides in the Classroom: Evidence from a Randomized Trial," *Journal of the European Economic Association* 18, no. 1 (February 2020): 469–505, <https://doi.org/10.1093/jeea/jvy048>; Steven W. Hemelt, Helen F. Ladd, and Calen R. Clifton, "Do Teacher Assistants Improve Student Outcomes? Evidence from School Funding Cutbacks in North Carolina," *Educational Evaluation and Policy Analysis* 43, no. 2 (June 2021): 280–304, <https://doi.org/10.3102/0162373721990361>.
85. See full methodological details and the findings from this project at National Council on Teacher Quality, "NCTQ Teacher Prep Review," <https://www.nctq.org/review> (accessed September 21, 2023).
86. See details on implementing the UTeach model at UTeach Institute, "UTeach Implementation," <https://institute.uteach.utexas.edu/uteach-implementation> (accessed September 21, 2023).
87. Michael Hansen and Li Feng, "Using Teacher Loan Forgiveness to Attract a Stronger, More Diverse Workforce," Brookings Institution (December 2019), <https://www.brookings.edu/articles/using-teacher-loan-forgiveness-to-attract-a-stronger-more-diverse-workforce>.



4. School Staffing and Teacher Quality

Thomas S. Dee

Executive Summary

Good teaching is deeply important for its immediate impact on both student learning and multiple, longer-run dimensions of educational and economic success. However, the effectiveness of individual teachers is highly variable and unevenly distributed across students. Fortunately, research over the last decade shows that strategies for improving the performance of in-service teachers have considerable promise. For example, focused training can significantly amplify the impact of teachers on student learning. Such professional development appears to be particularly effective when it emphasizes specific challenges of classroom practice. Similarly, performance-based teacher assessment systems can guide effective professional development and introduce high-powered incentives for teacher excellence, as well as establish informed procedures for directing chronically underperforming teachers out of the classroom.

However, the practical challenges to realizing such meaningful improvements in the effectiveness of the teaching workforce—and doing so consistently at scale—are considerable. For example, the exact ways to design and deliver consistently effective professional development for teachers are uncertain. This strongly indicates the need to embed new and ongoing professional development efforts within purposive cycles of design and evaluation. Similarly, there are several substantive logistical and political barriers to introducing effective and enduring systems of teacher performance assessment. These include the challenges of designing aligned and accurate data systems and assessment measures that reliably capture the variation in teacher performance and coupling them with the clear communication of reliably implemented incentives. The perception of political durability may also be key to the success of such teacher assessment reforms. However, recent research studies have identified several initiatives that serve as encouraging proof points for the promise of these reforms.

- *A Nation at Risk* placed a focus on teacher quality and anticipated some of the most dramatic education policy innovations of the past forty years. Creating large-scale, lasting changes related to teacher effectiveness has proved challenging, however.

- To move forward, we need a deeper understanding of how best to design reforms such as new teacher evaluation models or improved teacher professional development. Political barriers have also stood in the way of taking these reforms to scale.
- Policymakers should explore how to build broader coalitions around teacher effectiveness and perhaps use more incremental approaches to help build the evidence base for more lasting and scalable reform.

• • •

The publication of *A Nation at Risk (ANAR)* in 1983 was the defining moment of the “first wave” of education reform (NCEE 1983). It articulated improbably long-lived insights that continue to define education policy and discourse to this day. In particular, *ANAR* underscored, with uncommon rhetorical flourishes, the contrast between the ambitious ideals of a “Learning Society” and existing educational standards defined by modest minimum requirements, such as the low expectations embedded in high schools’ minimum competency tests and “cafeteria-style” curricula. Clearly, *ANAR*’s most prominent recommendation was the adoption of high school graduation requirements grounded in a “New Basics” curriculum that would feature four years of English; three years of science, math, and social studies; a half year of computer science; and, for college-bound students, two years of foreign-language instruction.

However, *ANAR* also commented on several other dimensions of the education system in the United States, including the state of the teaching profession. In particular, *ANAR* concluded that “too many teachers are being drawn from the bottom quarter of graduating high school and college students” (22). The report also underscored the inadequate subject-matter focus of teacher training, low pay, teachers’ limited influence on key professional decisions (e.g., textbooks), and the targeted character of teacher shortages. These findings—and the seven specific recommendations *ANAR* made regarding teaching—have been the focus of education research, commentary, and policymaking to this day.

In this chapter, I provide a compact overview of key insights from the research and policy-making that occurred in the wake of these recommendations. I focus specifically on the developments relevant to *in-service* teachers, while the important issues related to recruitment, induction, and mentoring in the teaching profession are addressed separately by Michael Hansen in chapter 3. *ANAR* made four specific recommendations relevant to in-service teachers. One is that teacher salaries should be “professionally competitive, market-sensitive, and performance-based” and linked to “an effective evaluation system” that rewards effective teachers and guides underperforming teachers toward improvement or termination. A related second recommendation advocates for collectively developed “career ladder” designations that distinguish beginning, experienced, and master teachers. *ANAR*’s remaining two recommendations for in-service teachers focus on supporting teacher improvement through funded time for professional development (30–31).

THEORIES OF ACTION

ANAR's recommendations for in-service teachers tacitly reflect two broad and complementary theories of action for improving teacher effectiveness and student outcomes. One involves improving the effectiveness of existing teachers. The intent is for this to occur through professional development activities and through the implementation of well-designed financial and professional incentives. Both of these intend to promote an understanding of high-quality classroom practices as well as their consistent use. The second theory of action focuses on selection—that is, performance assessment systems designed to retain and elevate the most effective teachers while ensuring that persistently ineffective teachers exit the classroom. Notably, these policy recommendations stand in sharp contrast to conventional efforts to promote teacher effectiveness through generic salary increases unrelated to performance or need and through reducing class sizes by hiring more teachers.

The motivations for *ANAR*'s theories of action rest upon several important stylized facts about teachers that have become increasingly well established since its publication. Arguably, the most foundational evidence concerns the variation in effectiveness across teachers. An older debate had questioned whether there are aspects specific to teaching that make it prohibitively difficult to measure teacher effectiveness in a valid and reliable manner (Murnane and Cohen 1986; Ballou 2001). However, richer data and methodological advances have led to a consensus about the general validity of teacher effectiveness measures while also acknowledging important evidence on the degree of noisiness common to such measures (Staiger and Rockoff 2010).

These studies indicate that the variation in teacher effectiveness is large, particularly relative to the effects of other promising education interventions. Specifically, a one-standard-deviation improvement in teacher effectiveness corresponds to a gain in student performance on standardized tests of roughly 0.1 to 0.2 standard deviations (e.g., Rivkin et al. 2005; Rockoff 2004; Aaronson et al. 2007; Staiger and Rockoff 2010). Critically, the manner in which teachers are currently assessed—that is, informal, “drive-by” evaluations—captures virtually none of this documented variation, rates the vast majority of teachers as satisfactory, and results in little performance-based attrition of low-performing teachers from the classroom.

Another important stylized fact is that, at the hiring stage, school leaders have little capacity to identify the teachers who will become more effective (Staiger and Rockoff 2010). This combination of facts—that teachers vary considerably in impact, but this impact can be observed much more easily after several years in the classroom than at the hiring stage—suggests the need for broader access to the teaching profession coupled with discerning assessment systems that guide subsequent personnel decisions. In particular, decisions to tenure rather than dismiss the lowest-performing teachers can have dramatic consequences given the length of teaching careers (Staiger and Rockoff 2010).

Over the past fifteen years, this evidence has motivated a number of ambitious public and philanthropic efforts to systematically improve the effectiveness of the teacher workforce through performance-based assessment systems. Recent research has also provided more credible evidence of direct initiatives designed to improve the performance of all in-service teachers through professional development. I discuss these policy innovations and the related research below.

IMPROVING TEACHER EFFECTIVENESS

ANAR recommended that teachers receive eleven-month contracts so that they could spend more time in professional development and provide additional instruction for students with special needs. While the eleven-month contract has not been widely adopted, broader efforts to improve the performance of in-service teachers through direct training and support involve a substantial expenditure of time and money. However, accurately identifying the magnitude of these outlays is not straightforward given the accounting challenges of categorizing such activities and their demands on time for both teachers and nonteaching staff. For example, a study by Alexander and Jang (2019) examined expenditure reports for Minnesota school districts and found that 1 percent of 2013–14 operational expenditures was spent on activities defined by the state as staff development. In contrast, a study by the New Teacher Project (2015) found that 2013–14 expenses related to teacher improvement constituted, on average, 8 percent of district budgets. This figure consisted of both direct expenditures on teacher improvement, such as professional development, coaching, and new-teacher support, as well as related indirect expenditures, such as the management, strategic, and operational expenses for these improvement efforts.

Focusing specifically on professional development, a study commissioned by the Gates Foundation (2014) found that the typical teacher spends sixty-eight hours per year on professional learning directed by districts, or eighty-nine hours when courses and self-guided professional learning are included. Most of the time spent by teachers in professional development occurs in workshops and professional learning communities conducted by district staff. The cost of this professional development was estimated at \$18 billion per year in 2014. Teacher perceptions of the quality of these investments have generally not been encouraging, nor do they appear to have clear links to teacher performance or improvement (TNTP 2015; Gates Foundation 2014). The Gates report also stresses the overwhelming use of district staff instead of market-tested external providers to provide professional development, as well as limited teacher voice in choosing their training.

Despite the considerable expense and prominence of teacher professional development, credible research on the impact of these investments has also been quite limited over much of the period since *ANAR*'s publication. For example, Yoon et al. (2007) reviewed more than 1,300 studies potentially addressing the impact of teacher professional development on student learning and found only nine studies that met the evidence standards in the federal What Works Clearinghouse: six randomized controlled trials and three quasi-experimental studies

conducted between 1986 and 2003. However, what these studies revealed suggests a striking proof of concept: teachers who received substantial professional development could boost the achievement of the average control-group student by 21 percentile points. Notably, these nine professional development initiatives focused on elementary grades but differed in their theories of action (Yoon et al. 2007).

However, other quasi-experimental studies serve as a reminder that implementing effective professional development consistently at scale is a serious challenge. Jacob and Lefgren (2004) examined the effect of teacher training in Chicago Public Schools using a credible natural experiment in which schools with low baseline test scores received additional resources for staff development. They found that this initiative had “no statistically or academically significant effect” on math or reading achievement of elementary students. Similarly, Harris and Sass (2011) examined student-level longitudinal data linked to teacher data for the state of Florida and did not find an overall impact of professional development on teacher productivity. However, they did find positive effects of content-focused math professional development on student outcomes at the elementary and middle-school levels.

Over the past decade, experimental studies of teacher professional development have proliferated. In general, they have provided mixed evidence of the learning impact of investments in professional development. For example, experimental studies by Garet et al. (2008, 2010) found that reading- and math-focused training changed teacher knowledge and practice but without clearly improving student achievement. However, meta-analytic summaries of such experimental professional development evaluations suggest that positive effects exist but vary considerably by program design. For example, Basma and Savage (2018) examined seventeen literacy-focused professional development studies and found an overall effect size for reading achievement of 0.225. Similarly, in a meta-analysis of ninety-five STEM-focused professional development studies with experimental and quasi-experimental designs, Lynch et al. (2019) report an average effect size of 0.21.

However, other multisubject meta-analyses suggest smaller but still positive effects on student learning. For example, Fletcher-Wood and Zuccollo (2020) identified fifty-three experimental evaluations of teacher professional development and found an overall effect size of 0.09. Similarly, Sims et al. (2021) reviewed 104 experimental evaluations and found an overall effect size of 0.05. Given the considerable financial expense of most training investments, effects of this size, though positive, raise serious questions about cost-effectiveness.

These reviews also note and seek to examine the considerable variation across professional development programs in terms of impact. Kennedy (2016) argues that the widely discussed design features of teacher professional development—namely program duration, emphasis on content knowledge, and use of professional learning communities—are far less relevant than whether the training addresses any of the four persistent challenges of teaching: portraying content, managing student behavior, enlisting student participation, and knowing what students understand. In a similar vein, Sims et al. (2021) characterize professional development programs by the more general ways they change teacher skills and behaviors. Specifically,

they characterize teacher professional development by four “IGTP” traits that indicate whether teachers are provided with new insights (I), goal-oriented behaviors (G), and techniques (T) that are embedded in practice (P). And they conclude that professional development programs with all four traits have an effect size on student learning of 0.17. However, these assessments may obscure the relevance of professional development initiatives that focus on the most effective elements of content and practice, such as an emphasis on “science of reading” approaches in literacy-focused training.

Overall, this evidence indicates that *ANAR* was prescient in emphasizing the need for ongoing training of in-service teachers. The available evidence suggests that such training can have substantial effects on student learning. However, realizing the increasingly well-established potential of this training is not straightforward. It involves the perennial challenge of translating research findings—that is, the critical design features of effective professional development—into genuine changes in high-impact practice at scale.

TEACHER EVALUATION AND PERFORMANCE-BASED INCENTIVES

ANAR also made prominent recommendations to dramatically change how we pay and evaluate public school teachers. In general, the status quo to this day compensates teachers according to single-salary schedules that rigidly structure pay according to years of experience and observed qualifications (e.g., a graduate degree) that do not consistently predict teacher effectiveness. This approach has historical origins in well-intentioned efforts to eliminate overt discrimination and capriciousness in teacher pay. Today, critics allege that this inflexible approach has led to low and undifferentiated salaries that do little to attract, motivate, and retain the most-effective teachers and to direct the least-effective teachers out of the classroom, particularly in hard-to-staff schools and high-need subjects. Furthermore, this approach to pay is coupled with low-stakes, “drive-by” teacher evaluations that capture little of the variation in teacher performance and do not provide reliable guidance for professional learning (Weisberg et al. 2009).

ANAR envisioned an alternative in which teacher compensation was substantially higher but also based on performance in a manner that would direct persistently underperforming teachers either to improve or to leave the profession. In the aftermath of *ANAR*’s publication, several states and districts experimented with providing teachers with extra pay and career-ladder recognitions for demonstrated merit (though, not generally, dismissing chronically underperforming teachers). These reforms tended to be short-lived despite encouraging results (Cornett and Gaines 1994). While the rollback of these reforms was clearly a policy choice, the underlying causes are debated. Ballou (2001) argued that it largely reflected the opposition of teachers’ unions. Murnane and Cohen (1986) contended that it reflected the distinctive character of teachers’ professional practice—that is, multidimensional and difficult to observe. However, random-assignment evidence from a comparatively well-implemented

career ladder program in Tennessee indicates that it was effective in identifying teachers who raised student achievement (Dee and Keys 2004).

The past two decades have witnessed a diverse variety of ambitious efforts, often encouraged by prominent philanthropic and federal initiatives, to measure teacher performance and to link it to improvement supports and incentives such as financial benefits, career-ladder designations, and dismissal threats. The research on these different reforms suggests their promise but also underscores the nontrivial challenges (e.g., design features, implementation, and political credibility) that make the consistent realization of this promise difficult. For example, the Obama administration's Race to the Top (RttT) initiative disbursed more than \$5 billion to states in a competition based in part on their commitment to developing systems for promoting teacher effectiveness. While RttT was effective in promoting state policy adoption (Howell and Magazinnik 2017), its effects on key design features and implementation are far less clear. In particular, while states were more likely to have multiple measures of teacher performance in the wake of RttT, the use of this data to inform salary and retention decisions remained uncommon (Hallgren, James-Burdumy, and Perez-Johnson 2014). The state reforms over this period were "rarely sustained over time," offered low bonuses, and rated fewer than 1 percent of teachers as unsatisfactory (Bleiberg et al. 2021).

A more granular focus on the available evidence from specific initiatives provides richer insights into these issues of design, implementation, and political durability. For example, several studies focused narrowly on simply providing teachers with incentives for improved performance. These studies often found null (or weak) effects that are likely to reflect the unique character of these programs. "Cash for test scores" experiments with individual incentives for teachers in Nashville (Springer et al. 2012) and group incentives for teachers in Round Rock, Texas (Springer et al. 2013), found little to no evidence of effects on teacher practices, attitudes, and the learning gains of their students. Similarly, studies of a group-based teacher-incentive experiment in New York City (Goodman and Turner 2013; Fryer 2013) found that they had no overall effects on key teacher or student outcomes.

Critics of teacher incentives suggest that these null findings reflect a misunderstanding of teacher motivations and the manner in which such incentives might debase intrinsic motivation (e.g., Murnane and Cohen 1986). However, three design features of these studies could also contribute to these null findings and have important implications for performance-based assessment and compensation. First, the fact that participants know that these experimental incentives have a short term (e.g., two years) can sharply attenuate the resulting motivation to undertake changes in professional practices. This same concern can also apply to the incentives embedded in at-scale policy reforms that are viewed as faddish and unlikely to endure politically. Second, these initiatives generally focused on student achievement as the incentivized outcome. This may weaken the impact of incentives if teachers do not see or understand how they should change everyday practice to realize these rewards. A related third point is that these incentive studies generally did little to support and guide teachers in how they could change their professional practices to earn these rewards.

Three other studies suggest the potential importance of other design features. A teacher-incentive study in Chicago Heights, Illinois, found positive effects on student achievement (but only in the first wave of the experiment) when the incentives were framed as the loss of an award rather than a gain (Fryer et al. 2022). Second, the Talent Transfer Initiative (TTI) found positive effects when offering high-performing teachers a high-powered incentive (\$20,000) linked to a distinctly clear, easily observed, and important behavior: working in a hard-to-staff school for two years (Glazerman et al. 2013). However, it is notable that these incentive-based gains were difficult to realize. More than 1,500 teachers had to be approached in order to fill only eighty-one vacancies. Third, the Accelerating Campus Excellence (ACE) program in Dallas similarly provided large incentives to highly effective teachers willing to work in hard-to-staff schools. Morgan et al. (2023) presented evidence that ACE produced dramatic gains in student performance: a 0.3 effect size in reading and 0.4 in math. This study also found that this success replicated as the program went to scale and that these gains were reversed when the program was eliminated.

Notably, these focused incentive programs all fall short of the more comprehensive system of assessments, supports, and incentives recommended by ANAR. TAP: The System for Teacher and Student Advancement (formerly known as the Teacher Advancement Program), which was introduced in 1999 and is currently active in “nearly twenty states and hundreds of school districts across the US” (Cohodes, Eren, and Ozturk 2023), is closer to ANAR’s vision. Specifically, the defining features of TAP include career ladder designations for teachers and job-embedded, professional learning led by master teachers. In support of this professional learning, TAP also provides teachers with comprehensive evaluations of their professional practice. However, it is not clear that this “instructionally focused accountability” articulates clear mechanisms for directing consistently low-performing teachers out of the classroom (the selection mechanism in ANAR’s theory of change). Finally, TAP includes performance pay typically linked to observations of teachers’ professional practice, such as classroom observation, portfolios, and interviews, as well as test scores.

The available evidence suggests that TAP is effective in improving teacher performance and student outcomes. Specifically, in a quasi-experimental study based on 1,200 schools from two states, Springer, Ballou, and Peng (2014) found that TAP increased student performance, particularly at the elementary school level, with effect sizes varying from 0.12 to 0.34 by grade. Similarly, Cohodes, Eren, and Ozturk (2023), leveraging the rollout of TAP across schools in South Carolina, found that it generated improvements in several long-run outcomes, including educational attainment, criminal activity, and the take-up of government assistance. However, a random-assignment evaluation of TAP in Chicago schools by Glazerman and Seifullah (2012) found that it did not improve student achievement and that it was also vexed by the challenges of implementing this reform with fidelity, such as teacher payouts being smaller than originally stated and no rewards based on value added because of inadequate data systems.

Two other high-profile studies provided further evidence of the serious challenges of implementing comprehensive reforms of teacher assessments and compensation as well as of credibly assessing their effects. The first example is the federal Teacher Incentive Fund (TIF). Congress established TIF in 2006 to provide grants to high-need schools implementing

performance-based compensation systems. The four required components of TIF reforms also resembled those suggested by ANAR: (1) measures of teacher performance, including observations of classroom practice; (2) large, differentiated, difficult-to-earn performance bonuses; (3) additional pay for career-ladder opportunities, such as becoming a master teacher and coach; and (4) professional development linked to the teacher assessments. A congressionally mandated study of TIF focused on the 2010 grant recipients in more than 130 school districts and found it led to student achievement of 1 to 2 percentile points higher in reading and math (Chiang et al. 2017).

However, there are two important caveats to this evidence of modest impact. First, the implementation of these reforms in the study districts was incomplete. Only about half of the participating districts reported implementing all four components of the reforms required by TIF. In particular, professional development was frequently not provided, and most teachers received bonuses, “a finding inconsistent with making bonuses challenging to earn” (Chiang et al. 2017). Second, the treatment-control contrast assessed in this random-assignment study did not examine the effect of TIF versus “business as usual.” Instead, the treatment schools in the study were intended to receive pay-for-performance bonuses while the control group received automatic bonuses. And all study participants, both treatment and control, were assigned access to the three other TIF components: career ladder responsibilities and rewards, evaluative feedback, and professional development. In this critical but often overlooked detail, the federal study of TIF more closely resembles the studies of teacher incentives noted above than a true evaluation of teacher assessment systems.

The Gates-funded Intensive Partnerships for Effective Teaching initiative is a second widely discussed example of implementing and evaluating teacher assessment systems. This initiative sought to introduce assessment reforms within three school districts and four charter management organizations. Similar to both TAP and TIF, this effort featured focused professional development and career ladder incentives along with performance pay and retention decisions based on direct, structured observation of teacher practice and value-added scores. A quasi-experimental study found that these reforms did not clearly improve the focal student outcomes of high school graduation and college attendance (Stecher et al. 2018). However, the implementation of the reforms appears to have been weak. The teacher evaluations flagged few teachers as poor performers, and in sites with available data, only 1 percent were dismissed for poor performance. As with the federal TIF evaluation, the treatment contrast that was studied was muted because the comparison schools in this study often adopted similar policies.

IMPACT, the highly controversial teacher assessment reforms introduced in the District of Columbia Public Schools (DCPS), is distinctive as a seminal and enduring effort to implement ANAR’s recommendations with fidelity. IMPACT evaluated DCPS teachers on multiple measures with a heavy emphasis on structured classroom observations, including some conducted by district staff, and linked professional development. These evaluations resulted in measures of teacher performance that exhibited variation rather than being largely uniform. IMPACT linked these measures to high-stakes consequences: substantial pay increases for “highly effective” teachers, particularly those in high-poverty schools; dismissal for a small

number of “ineffective” teachers; and a dismissal threat for “minimally effective” teachers who did not become effective within a year.

A quasi-experimental study of the incentive contrasts embedded in IMPACT found it had positive effects on teacher performance (Dee and Wyckoff 2015). This study’s design leveraged a feature of IMPACT in which teachers with performance scores just below a threshold value were deemed “minimally effective” and subject to a dismissal threat while those with scores at or above the threshold were not. A comparison of teachers just below and above this threshold found that the threat of dismissal caused minimally effective teachers either to leave the district or to improve their measured performance substantially. A powerful financial incentive for highly effective teachers to repeat their prior performance also appeared to have positive effects.

Three other aspects of IMPACT merit emphasis. First, the political credibility and resiliency of IMPACT appeared to be highly salient. In 2010, when the city (and district) leadership who championed IMPACT were forced out of office, the first “minimally effective” designations did not appear to change teacher behavior. However, the ratings reported in the summer of 2011, when it appeared that IMPACT would endure, did drive changes in teacher behavior.

Second, evidence indicates that IMPACT not only improved the performance of existing teachers but also replaced underperforming teachers who exited with substantially more effective instructors. Specifically, a quasi-experimental study by Adnot et al. (2017) finds that, when a low-performing teacher exited, their replacement raised student performance by 0.14 standard deviations in reading and 0.24 standard deviations in math. Third, the performance benefits of IMPACT’s incentives endured through subsequent revisions to the teacher supports and ratings structure (Dee, James, and Wyckoff 2021).

A second district reform of note (and one with strong parallels to IMPACT) began in the Dallas Independent School District in 2015. Specifically, like IMPACT, the Teacher Excellence Initiative (TEI) replaced a single-salary schedule with compensation based on multiple measures of teacher performance. Furthermore, like IMPACT, it also did so in the context of accountability for school principals. TEI also implemented a unique design feature to discourage inflated or arbitrary ratings of teachers. It fixed the overall distribution of ratings and penalized principals for subjective ratings that were highly misaligned with test-based ratings. A synthetic-control study by Hanushek et al. (2023) found that these reforms led to statistically significant increases in student achievement that grew over time to a roughly 0.2 standard deviation in math and a 0.1 standard deviation in reading.

CONCLUDING THOUGHTS

ANAR’s recommendations that focused on improving the effectiveness of in-service teachers were a harbinger of some of the most dramatic education policy innovations of the past forty years. And these innovations have provided us with several proofs of concept and new

insights that establish the potential to improve student learning through dramatic changes in teacher evaluation, in-service training, and compensation.

However, it must also be acknowledged that there has clearly not been large-scale, lasting change regarding ANAR's teacher-focused recommendations. Uninformative, low-stakes assessments of professional practice and rigid single-salary schedules are still the norm for the vast majority of teachers in US public schools. And while in-service teachers do engage in extensive professional development, the impact of these expensive and highly variable investments is uncertain at best.

Any serious effort to reimagine the assessment, training, and compensation of in-service teachers should begin by confronting the factors that have contributed to the long durability of the status quo. There appear to be three broad and interrelated impediments to substantive change. The first is the need to improve the knowledge base of how best to design the key features of these reforms. For example, efforts to improve teacher evaluation and introduce performance-based teacher pay rely critically on valid and reliable measures of teacher performance. Promising gains in measuring teacher effectiveness are likely to come from continued improvements to structured rubrics for classroom practices. Incentives can better guide the professional improvement of teachers when they are linked to the high-impact, everyday classroom practices teachers directly control and can enhance through complementary training.

Another important area where improved knowledge is critical to driving at-scale change concerns the design of teacher professional development. The typical professional development experience, workshops directed by internal district staff, is often criticized (e.g., the New Teacher Project 2015). At the same time, a recent and growing body of experimental studies indicates that purposively designed professional development can have substantial impact. This literature generally emphasizes the particular benefits of in-service training that focuses on meeting more general challenges of teacher practice (e.g., Kennedy 2016; Sims et al. 2021). While more can be learned about the design of professional development, the question of how to design its delivery is even more uncertain. A study from the Gates Foundation (2014) suggests that relying more on external providers of professional development will make it easier to move nimbly to market-tested and effective approaches. However, several of the teacher assessment reforms discussed here instead emphasize redesigning internally provided professional development to rely on master teachers who may be better positioned to serve as coaches providing embedded and relevant training. These issues underscore the need to build a complementary learning agenda around any new reforms (e.g., inquiry cycles, networked improvement communities).

A second impediment to realizing ANAR's vision concerns the multifaceted operational challenges of implementing meaningful reforms effectively at scale. The null findings from credibly identified studies of professional development in at-scale field settings suggest this issue (Jacob and Lefgren 2004; Harris and Sass 2011). However, more-direct and sobering evidence comes from several well-funded, high-profile efforts to introduce teacher assessment and

compensation reforms at some scale. These include (1) the failure to deliver value-added bonuses because of data-system inadequacies in TAP (Glazerman and Seifullah 2012); (2) the limited variation in teacher ratings and their infrequent use in personnel decisions in the Gates Foundation’s Intensive Partnership for Effective Teaching (Stecher et al. 2018); (3) the inconsistent delivery of professional development and the broad distribution of bonuses under the federal Teaching Incentive Fund (Chiang et al. 2017); and (4) the limited use of teacher evaluations to guide salary and retention decisions under the RttT initiative (Hallgren, James-Burdumy, and Perez-Johnson 2014).

A third and closely related impediment is political opposition. With regard to introducing performance-based pay, this most obviously refers to the opposition of teachers’ unions. However, it can also involve unresponsive public-sector bureaucracies. Furthermore, reform efforts can also fail when their success and durability rely on politically determined funding commitments. The political opposition to reform in the broader public also turns on misinformation about what the existing evidence discussed here actually indicates. Specifically, opponents of the types of reforms recommended by ANAR often argue that investments in professional development are effective while performance-based pay has failed.

Given these interlocking issues, a compelling way to achieve change at scale may involve forming political coalitions around compelling reforms that adopt some but not all of ANAR’s proposals. For example, it may be possible to move school districts toward more effective professional development delivered by a carefully curated set of outside vendors if their provision involved cost-sharing that saved district resources. Alternatively, it may be possible to achieve durable political support for a teacher evaluation system if that system focuses narrowly on identifying master teachers and providing them with training and extra pay to coach their peers but takes a more incremental approach toward dismissing underperforming teachers. Intentionally combining such efforts with careful evaluation could, over the longer term, seed further evidence-based change in this important domain.

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.

We are worse off than we were forty years ago relative to the state of the teaching profession. I wonder the extent to which our reforms have contributed to the decay. In the years following the publication of ANAR, my grandmother, who had graduated at the top of her class, retired from teaching earlier than she had planned because her state instituted a teacher test she was not interested in taking. Twenty years later, I started teaching in my home district. Had I not been certified at the time, I would not have been considered for an interview. Due to the teacher shortage today, states are creating a plethora of on-ramps into the profession, and

districts are forced to hire anyone they can find. While the teacher reform initiatives have been well intended, the application has come at high cost and low benefit. I applaud and wholeheartedly support Dee’s assertion that we should build political coalitions and pivot to focus teacher evaluations on identifying excellence. We need to also prioritize improvement-focused feedback in our support and evaluation systems.

—Holly Boffy, Louisiana State Board of Elementary and Secondary Education

Author Thomas Dee concludes: “There has clearly not been large-scale, lasting change regarding ANAR’s teacher-focused recommendations. Uninformative, low-stakes assessments of professional practice and rigid single-salary schedules are still the norm for the vast majority of teachers in US public schools. And while in-service teachers do engage in extensive professional development, the impact of these expensive and highly variable investments is uncertain at best.” His research review supports that conclusion, and I agree. However, I feel that we should not stop these efforts; rather, we should significantly improve them. While teacher evaluation systems may never be implemented into a performance pay system, the information (ratings) is critical in ensuring appropriate staffing, developing meaningful career ladders, and identifying targeted personal professional development.

In addition to Dee’s closing recommendations, which I think are helpful, a significant piece still missing from the overall analysis is the critical need for strong school leadership to address the effectiveness of in-service teachers. This has not been seriously addressed in the research designs nor in the implementation of teacher evaluation systems.

—Angelika Schroeder, Colorado State Board of Education

REFERENCES

- Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. “Teachers and Student Achievement in the Chicago Public High Schools.” *Journal of Labor Economics* 25, no. 1 (January): 95-135.
- Adnot, Melinda, Thomas Dee, Veronica Katz, and James Wyckoff. 2017. “Teacher Turnover, Teacher Quality, and Student Achievement in DCPS.” *Educational Evaluation and Policy Analysis* 39, no. 1 (March): 54-76.
- Alexander, Nicola A., and Sung Tae Jang. 2019. “Expenditures on the Professional Development of Teachers: The Case of Minnesota.” *Journal of Education Finance* 44, no. 4 (March): 385-404.
- Ballou, Dale. 2001. “Pay for Performance in Public and Private Schools.” *Economics of Education Review* 20, no. 1 (February): 51-61.
- Basma, Badriah, and Robert Savage. 2018. “Teacher Professional Development and Student Literacy Growth: A Systematic Review and Meta-Analysis.” *Educational Psychology Review* 30 (June): 457-81. <https://doi.org/10.1007/s10648-017-9416-4>.
- Bleiberg, Joshua, Eric Brunner, Erica Harbatkin, Matthew A. Kraft, and Matthew Springer. 2021. “The Effect of Teacher Evaluation on Achievement and Attainment: Evidence from Statewide Reforms.” EdWorkingPaper 21-496. Retrieved from Annenberg Institute at Brown University. <https://doi.org/10.26300/b1ak-r251>.
- Chiang, Hanley, Cecilia Speroni, Mariesa Herrmann, Kristin Hallgren, Paul Burkander, and Alison Wellington. 2017. *Evaluation of the Teacher Incentive Fund: Final Report on Implementation*

- and Impacts of Pay-for-Performance across Four Years*. NCEE 2018-4004. Washington, DC: National Center for Education Evaluation and Regional Assistance.
- Cohodes, Sarah, Ozkan Eren, and Orgul Ozturk. 2023. "Teacher Performance Pay, Coaching, and Long-Run Student Outcomes." National Bureau of Economic Research Working Paper Series no. w31056.
- Cornett, Lynn M., and Gale F. Gaines. 1994. "Reflecting on Ten Years of Incentive Programs: The 1993 SREB Career Ladder Clearinghouse Survey." Atlanta: Southern Regional Education Board.
- Dee, Thomas, and James Wyckoff. 2015. "Incentives, Selection, and Teacher Performance: Evidence from IMPACT." *Journal of Policy Analysis and Management* 34, no. 2 (October): 267–97.
- Dee, Thomas S., Jessalynn James, and James Wyckoff. 2021. "Is Effective Teacher Evaluation Sustainable? Evidence from District of Columbia Public Schools." *Education Finance and Policy* 16, no. 2 (Spring): 313–46.
- Dee, Thomas S., and Benjamin J. Keys. 2004. "Does Merit Pay Reward Good Teachers? Evidence from a Randomized Experiment." *Journal of Policy Analysis and Management* 23, no. 3 (Summer): 471–88.
- Fletcher-Wood, Harry, and James Zuccollo. 2020. *The Effects of High-Quality Professional Development on Teachers and Students: A Rapid Review and Meta-Analysis*. London: Education Policy Institute.
- Fryer, Roland G. 2013. "Teacher Incentives and Student Achievement: Evidence from New York City Public Schools." *Journal of Labor Economics* 31, no. 2 (April): 373–407.
- Fryer, Roland G., Jr. Steven D. Levitt, John List, and Sally Sadoff. 2022. "Enhancing the Efficacy of Teacher Incentives through Framing: A Field Experiment." *American Economic Journal: Economic Policy* 14, no. 4 (November): 269–99.
- Garet, Michael S., Stephanie Cronen, Marian Eaton, Anja Kurki, Meredith Ludwig, Wehmah Jones, Kazuaki Uekawa et al. 2008. *The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement*. NCEE 2008-4031. Washington, DC: National Center for Education Evaluation and Regional Assistance.
- Garet, Michael S., Andrew J. Wayne, Fran Stancavage, James Taylor, Kirk Walters, Mengli Song, Seth Brown et al. 2010. *Middle School Mathematics Professional Development Impact Study: Findings after the First Year of Implementation*. NCEE 2010-4009. Washington, DC: National Center for Education Evaluation and Regional Assistance.
- Gates Foundation. 2014. *Teachers Know Best: Teachers' Views on Professional Development*. Seattle: Bill & Melinda Gates Foundation.
- Glazerman, Steven, Ali Protik, Bing-ru Teh, Julie Bruch, and Jeffrey Max. 2013. "Executive Summary." In *Transfer Incentives for High-Performing Teachers: Final Results from a Multisite Randomized Experiment*. NCEE 2014-4004. Washington, DC: National Center for Education Evaluation and Regional Assistance.
- Glazerman, Steven, and Allison Seifullah. 2012. *An Evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after Four Years*. Princeton, NJ: Mathematica Policy Research.
- Goodman, Sarena F., and Lesley J. Turner. 2013. "The Design of Teacher Incentive Pay and Educational Outcomes: Evidence from the New York City Bonus Program." *Journal of Labor Economics* 31, no. 2 (April): 409–20.
- Hallgren, Kristin, Susanne James-Burdumy, and Irma Perez-Johnson. 2014. *State Requirements for Teacher Evaluation Policies Promoted by Race to the Top*. NCEE Evaluation Brief. NCEE 2014-4016. Washington, DC: National Center for Education Evaluation and Regional Assistance.
- Hanushek, Eric A., Jin Luo, Andrew J. Morgan, Minh Nguyen, Ben Ost, Steven G. Rivkin, and Ayman Shakeel. 2023. "The Effects of Comprehensive Educator Evaluation and Pay Reform on Achievement." National Bureau of Economic Research Working Paper Series no. w31073.

- Harris, Douglas N., and Tim R. Sass. 2011. "Teacher Training, Teacher Quality and Student Achievement." *Journal of Public Economics* 95, nos. 7-8 (August): 798-812.
- Howell, William G. 2015. "Results of President Obama's Race to the Top." *Education Next* 15, no. 4 (Fall): 58-67.
- Howell, William G., and Asya Magazinnik. 2017. "Presidential Prescriptions for State Policy: Obama's Race to the Top Initiative." *Journal of Policy Analysis and Management* 36, no. 3 (Summer): 502-31.
- Jacob, Brian A., and Lars Lefgren. 2004. "The Impact of Teacher Training on Student Achievement: Quasi-Experimental Evidence from School Reform Efforts in Chicago." *Journal of Human Resources* 39, no. 1 (Winter): 50-79.
- Kane, Thomas J., Daniel F. McCaffrey, Trey Miller, and Douglas O. Staiger. 2013. "Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment." MET Project Research Paper, Bill & Melinda Gates Foundation.
- Kennedy, Mary M. 2016. "How Does Professional Development Improve Teaching?" *Review of Educational Research* 86, no. 4 (February): 945-80.
- Lynch, Kathleen, Heather C. Hill, Kathryn E. Gonzalez, and Cynthia Pollard. 2019. "Strengthening the Research Base That Informs STEM Instructional Improvement Efforts: A Meta-Analysis." *Educational Evaluation and Policy Analysis* 41, no. 3 (September): 260-93. <https://doi.org/10.3102/0162373719849044>.
- Morgan, Andrew J., Minh Nguyen, Eric A. Hanushek, Ben Ost, and Steven G. Rivkin. 2023. "Attracting and Retaining Highly Effective Educators in Hard-to-Staff Schools." National Bureau of Economic Research Working Paper Series no. w31051.
- Murnane, Richard J., and David K. Cohen. 1986. "Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and a Few Survive." *Harvard Educational Review* 56, no. 1 (April): 1-18.
- NCEE (National Commission on Excellence in Education). 1983. *A Nation at Risk: The Imperative for Educational Reform*. Washington, DC: US Department of Education.
- Podgursky, Michael J., and Matthew G. Springer. 2007. "Teacher Performance Pay: A Review." *Journal of Policy Analysis and Management* 26, no. 4 (Autumn): 909-50.
- Rivkin, Steve G., Eric A. Hanushek, and John G. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73, no. 2 (March): 417-58.
- Rokoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review* 94, no. 2 (May): 247-52.
- Sims, Sam, Harry Fletcher-Wood, Alison O'Mara-Eves, Sarah Cottingham, Claire Stansfield, Jo Van Herwegen, and Jake Anders. 2021. *What Are the Characteristics of Teacher Professional Development That Increase Pupil Achievement? A Systematic Review and Meta-Analysis*. London: Education Endowment Foundation.
- Springer, Matthew G., Dale Ballou, Laura Hamilton, Vi-Nhuan Le, J. R. Lockwood, Daniel F. McCaffrey, Matthew Pepper, and Brian M. Stecher. 2013. *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching*. Nashville, TN: National Center on Performance Incentives.
- Springer, Matthew G., Dale Ballou, and Art Peng. 2014. "Estimated Effect of the Teacher Advancement Program on Student Test Score Gains." *Education Finance and Policy* 9, no. 2 (Spring): 193-230.
- Springer, Matthew G., John F. Pane, Vi-Nhuan Le, Daniel F. McCaffrey, Susan Freeman Burns, Laura S. Hamilton, and Brian Stecher. 2012. "Team Pay for Performance: Experimental Evidence from the Round Rock Pilot Project on Team Incentives." *Educational Evaluation and Policy Analysis* 34, no. 4 (December): 367-90.
- Staiger, Douglas O., and Jonah E. Rockoff. 2010. "Searching for Effective Teachers with Imperfect Information." *Journal of Economic Perspectives* 24, no. 3 (Summer): 97-118.

- Stecher, Brian M., Deborah J. Holtzman, Michael S. Garet, Laura S. Hamilton, John Engberg, Elizabeth D. Steiner, Abby Robyn, Matthew D. Baird, Italo A. Gutierrez, Evan D. Peet et al. 2018. *Improving Teaching Effectiveness. Final Report*. Santa Monica, CA: RAND Corporation.
- The New Teacher Project (TNTP). 2015. *The Mirage: Confronting the Hard Truth about Our Quest for Teacher Development*. New York: TNTP Inc. https://tntp.org/assets/documents/TNTP-Mirage_2015.pdf.
- US Department of Education, Office of Elementary and Secondary Education, *State and District Use of Title II, Part A Funds in 2020–21*, Washington, DC, 2022.
- Weisberg, Daniel, Susan Sexton, Jennifer Mulhern, and David Keeling. 2009. *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*. New York: The New Teacher Project.
- Yoon, Kwang Suk, Teresa Duncan, Silvia Wen-Yu Lee, Beth Scarloss, and Kathy L. Shapley. 2007. *Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement*. Issues & Answers Report, REL 2007-No. 033. Washington, DC: US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.



5. The Case for Curriculum

Since A Nation at Risk, Education Reform Efforts Have Mostly Stopped at the Classroom Door

Robert Pondiscio

Executive Summary

Decades of education reform have left policymakers, educators, and students alike fatigued and unimpressed. From standardized testing to accountability measures and smaller classroom sizes, almost every idea under the sun has been tried and tried again, except for one: curriculum reform is the black sheep of education, long thought to be the sole domain of the individual teacher, with even the most prominent education reformers unwilling to take up its banner. Inspired by the work of E. D. Hirsch Jr. and the emerging science of reading movement, policymakers and educators should come together and work to reform the slapdash and unchallenging curriculum that defines many American classrooms and craft new knowledge-rich materials that align with high standards. When research indicates that a strong curriculum leads to greater results for students than replacing a 50th-percentile teacher with a 75th-percentile teacher, it is unthinkable that we would not make every effort to introduce high-quality instructional materials into the classrooms around the nation. The education reform movement of the past few decades has underperformed, exhausting its energy and spending its moral capital on “structural” reforms like standardized testing, accountability, and programs to transform the American teacher workforce. Placing curriculum at the center of reform efforts holds the promise to not only raise academic outcomes, but also get better results from the teachers we have—not the teachers we wish we had.

- American educators have tried many reforms to raise academic outcomes without finding one reform that works consistently.
- Curriculum reform is the one approach that hasn’t been given a fair trial.
- Rather than blaming the teacher education system, let’s try an approach that offers the promise of boosting outcomes with the existing workforce.



Several years ago, fresh from my South Bronx fifth-grade classroom and still feeling my way through the unfamiliar world of public policy and education reform, I found myself in a New York City ballroom where Michelle Rhee was receiving the Manhattan Institute’s Urban Innovator Award. She was, at that time and by a considerable margin, the most prominent and celebrated figure of the education reform movement in the United States, which was itself at the zenith of its power, prestige, and moral authority. As then chancellor of Washington, DC, schools, Rhee instituted a series of high-profile reforms, including closing chronically failing schools and tying teacher pay to performance. A few months earlier, she had been on the cover of *Time* magazine scowling and wielding a broom—a symbol of her determination to sweep underperforming teachers out of the city’s classrooms.

Rhee had recently announced the launch of a new initiative, Students First, and a goal to raise \$1 billion to support political candidates to advocate for “real change,” which she defined as putting students’ needs before those of adult interests such as teachers’ unions or wasteful bureaucracies like the one she’d been waging war upon in the nation’s capital. From my vantage point as a classroom teacher and curriculum reform advocate, focusing attention, energy, and resources on political campaigns and legislative races seemed even less likely to improve student outcomes than the work she’d been doing running Washington’s school system.

After the event, I had the opportunity to talk with Rhee about her reform work as chancellor and the work she was now envisioning in political advocacy. I wondered aloud whether it made sense to reach conclusions about the effectiveness of individual teachers who are poorly trained, have little say over their curriculum, and as often as not have no curriculum at all. As she moved from leading a major metropolitan school district to a position of influence over state-level education decisions, perhaps she might keep curriculum in mind? “The last thing we’re going to do,” she replied with a chuckle, “is get wrapped up in curriculum battles.”

I was taken aback by the dismissal. Readers may recall that Rhee was the embodiment of an ed reform movement that embraced a confrontational, even pugilistic style. In her talk that afternoon before a number of prominent figures in education and policymaking, she’d urged her listeners not to shrink from conflict on behalf of children, yet she herself had no stomach for a debate over what America’s children should learn in school.

More than a decade later, the encounter still stands out in my memory. Confronting the teachers’ unions on pay and tenure is worth a fight. So, too, is flipping a state legislative seat. Yet it was too heavy a lift to say what third-graders should know about American history, geography, or science, or whether they needed to know anything at all.

In fairness, this mindset was not unique to Rhee, who was merely the most vocal and visible representative of a theory of change common to education reformers of the time, which saw the external structures of education and the exercise of political power as the most important levers for improving student outcomes. The logic of this brand of reform assumed, at least

tacitly, that schools and teachers know what to do, have the capacity for improvement, and need mostly to be properly incentivized—or threatened—in order to be made to do it. The cri de coeur that launched the modern education reform movement, the 1983 report *A Nation at Risk*, appropriately sounded an alarm over school curricula that “have been homogenized, diluted, and diffused to the point that they no longer have a central purpose” and have been replaced with “a cafeteria style curriculum in which the appetizers and desserts can easily be mistaken for the main courses.” But concern over the substance of K-12 education in the United States had become an afterthought among the “structural reform” leaders like Rhee who came to prominence a generation later.¹

With the benefit of hindsight, it is no longer controversial to say that the structural reform theory of change has underperformed, its assumptions found wanting. To be sure, while twelfth-grade scores on the National Assessment of Educational Progress (NAEP) have not budged in two generations, there was evidence prior to the COVID-19 pandemic of gains in the earlier years of the twenty-first century, particularly in fourth-grade reading and math.² Nor has the education reform movement been without significant accomplishments. Urban charter schools, the movement’s clearest victory, have transformed urban education for low-income students of color in many major US cities, creating unprecedented opportunities for students and bringing competitive pressure to bear on local school districts.³ Graduates of schools founded and run by charter management networks continue to graduate from high school and attend college at much higher rates than they would have had those schools never existed.⁴ But the fact remains that if the classic ed reform playbook of higher academic standards, high-stakes testing, and muscular accountability was going to bear fruit, drive watershed improvement in student outcomes, or appreciably narrow racial achievement gaps, we’d have clear evidence of it by now.

Worse, as the education reform movement evolved from the do-gooder earnestness of its early days to a punitive technocratic regime, it overspent its moral capital and contributed to unmistakable reform fatigue.⁵ This led a significant number of public education stakeholders—parents, teachers, and taxpayers—to regard its policies and practices with skepticism, even cynicism, particularly as education spending continued to rise while student achievement stagnated and even declined.⁶

The lingering effects of COVID-related disruptions have shifted much of the attention in US education away from long-running debates over testing and accountability to more urgent discussions about learning loss, student mental health issues, and declining school attendance. It seems unlikely that the bipartisan ed reform coalition whose agenda dominated America’s K-12 agenda in the first decades of the twenty-first century will be returning to prominence anytime soon, if ever. The appetite for reform has waned considerably. The movement is what advertising and marketing professionals call a tainted brand. Indeed, ed reform “is now considered to be a loaded term that is no longer spoken in polite company,” former Massachusetts secretary of education James Peyser recently observed, “without risking a heated argument or losing the friendship of former allies.”⁷

EDUCATION REFORM'S NEXT FRONTIER: INSTRUCTIONAL REFORM

A simple fact about education in the United States trumps all others, yet has been largely overlooked in the education reform era and contributed to its disappointing underperformance: it takes nearly four million women and men to staff America's K-12 classrooms. A number that large, by definition, means that teachers will be people of average abilities and sentience—not saints, not superstars, and, more pertinently, not the cognitive elite, who do not exist in sufficient numbers to staff more than a fraction of America's classrooms. At the same time, it is not an overstatement to say that the ever-increasing demands placed on the average teacher make the job nearly impossible to do well, consistently, or sustainably. The reluctance even to cite, let alone address, this mismatch of expectations and abilities contributes to the mediocre performance of education in the United States, which has not changed significantly or satisfactorily since *A Nation at Risk* warned that “the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people.”⁸

If teacher capacity is unlikely to change, then what must change is the teacher's job. If the education reform movement is to regain its momentum and moral authority, becoming not merely a disruptive force but an *effective* one, it must reinvent itself as a practice-based movement that is clear-eyed and candid about human capital and system capacity, committed not to transforming the teacher workforce but to making teaching doable by the *existing* workforce and those likely to enter the profession in the future.

At the same time, candor requires acknowledging that this transformation can't be speedily or satisfactorily addressed, even if taken up with urgency. Education policy is a weak lever to change classroom practice. Enduring change requires shifts in the culture of teaching, which is notoriously slow to evolve, resistant to change, and skeptical—even cynical—about reform. For a 2017 study published by the National Council of Professors of Educational Administration, Wartburg College professor Richard Snyder interviewed a series of teachers with more than twenty years in the field to understand their perspectives on reforms and change. One teacher, who was aggravated with changes to class schedules, often ignored new directives from administrators. “In fact, Mr. Booker—a social studies teacher with over 30 years of experience—acknowledged giving ‘lip service’ to numerous initiatives, then returning to his own classroom intent on accomplishing intellectual discourse through interactive lecture.”⁹ Other teachers criticized increased top-down control, whether it be from Common Core State Standards or area education agency (AEA) consultants. Concerned about her loss of autonomy, Mrs. Rittmeyer stated:

So now the AEA is teaching us how to teach because we don't know how to teach kids how to read, and learn letters and sounds, things like that . . . never have darkened the doors of our classroom, but they can meet with us once a week and tell us what to do. That's very frustrating.¹⁰

The cyclical change and rebranding of old ideas common to education reform frustrated teachers the most. A high school teacher told Snyder, “I've become more frustrated,

especially when I started hearing things I've heard before and [being] spun as new. . . . I don't like the way we, we put brand new wrappers on things . . . and I sit through a pile of meetings and hear the same things I heard 15 years ago."¹¹

There is one arrow in the policymaker's quiver that remains mostly unutilized and can be used to effect positive change: curriculum reform. The adoption and implementation of high-quality instructional materials (HQIM) and making curriculum and its implementation central to school improvement efforts contribute to working conditions that allow teachers to focus on lesson *delivery*—not lesson *design*—while lending greater consistency to the student experience regardless of school setting and raising outcomes at something closer to scale. A systematic 2017 review of the effects of curricular choices in K-12 education conducted by the Johns Hopkins Center for Research and Reform in Education concluded that curriculum is “a critical factor in student academic success” and that a “comprehensive, content-rich curriculum is a common feature of academically high-performing countries.”¹² However, to a degree largely lost upon policymakers and other stakeholders in education, the curricula in most US schools and districts are not “comprehensive and content-rich” at all, but teacher driven, often improvisational and incoherent.

We have known for some time (or at least have had ample reason for curiosity and further study) that curriculum could be a richer vein of ore to mine than many of the more commonly pulled “structural” reform levers. In 2012, Grover J. “Russ” Whitehurst and Matthew M. Chingos noted in a widely read Brookings Institution paper that the effect size of choosing a better second-grade math curriculum was larger than replacing a 50th-percentile teacher with a 75th-percentile teacher. Clearly, it is easier to give children access to a strong curriculum than it is to dramatically increase the effectiveness of their teachers. At the same time, Whitehurst and Chingos lamented that “little research exists on the effectiveness of most instructional materials, and very little systematic information has been collected on which materials are being used in which schools.”¹³

It is hard to account for or excuse entirely the long-standing indifference to curriculum in education reform, which has long seemed to assume that differences in student outcomes are attributable mostly to who the teacher is, not what is being taught. To be sure, researchers have amply demonstrated that some teachers are more effective than others.¹⁴ But identifying what makes them so has proven elusive. No consistent or clear relationship has been found, for example, between teacher credentialing or certification exams and classroom effectiveness.¹⁵ However, education reformers of the late 1990s and early 2000s were enamored with “alternative certification” routes and high-profile initiatives like Teach for America to lure the “best and the brightest” graduates of elite universities to spend at least a few years in the classroom. Results have been mixed.¹⁶ But even if they were stellar, more than 80 percent of full-time teachers still enter the classroom via traditional training and certification routes, limiting this strategy's potential for transformational change.¹⁷ Neither is it likely that simply paying teachers more will make the profession as attractive to the best and brightest as technology, engineering, medicine, or law.

“What we teach isn't some sidebar issue in American education: it *is* American education,” David Steiner, executive director of the Institute for Education Policy at Johns Hopkins, has

noted. “The track record of top-performing countries, early evidence of positive effects from the faithful implementation of high-quality curricula here in the United States, and the persistent evidence that our classrooms are underchallenging our students at every level compel us to put the materials that we use to teach at the core of serious education reform.”¹⁸ If the assumption is that curriculum either is settled or doesn’t matter to student outcomes, it is a demonstrably incorrect assumption. A 2016 study by RAND Corporation revealed that virtually every English language arts (ELA) teacher in America—99 percent of elementary teachers; 96 percent of secondary school teachers—routinely use “materials they created or selected themselves.” Among elementary school teachers, 94 percent reported turning to Google to find ELA lesson plans and instructional materials; 87 percent searched Pinterest. The numbers are virtually the same for math.¹⁹ Survey data by the research firm MDR found that teachers spend seven hours per week searching for instructional resources and another five hours per week creating their own classroom materials.²⁰ The open question is whether this is an effective use of teacher time, and what might they use those hours for instead.

In theory, curating, customizing, or creating lessons from scratch allows teachers to tailor their instruction to meet the specific needs, interests, and abilities of their students. By designing their own curriculum, either in whole or in part, teachers can ostensibly adapt and differentiate class content, instructional methods, and assessments, resulting in a more personalized and engaging learning experience for students. If it could be demonstrated that the vast number of hours teachers spent doing curriculum design work paid dividends in raising student achievement, then no further comment would be necessary. The available evidence does not suggest a richer academic experience for students, however.

A 2019 study authored by University of Southern California Associate Professor Morgan Polikoff and education consultant Jennifer Dean explored the quality of supplemental materials teachers downloaded from popular websites, revealing “a major mismatch between what content experts think educators should (and shouldn’t) use in classrooms and what teachers, hungry for instructional resources, are choosing to download.”²¹ Polikoff and Dean rated most of the materials teachers chose themselves from popular websites such as Share My Lesson and Teachers Pay Teachers as “mediocre” or “probably not worth using.” Similarly, a 2018 report from the New Teacher Project (TNTP) based on one thousand observations found that students “spent more than 500 hours per school year on assignments that weren’t appropriate for their grade and with instruction that didn’t ask enough of them—the equivalent of six months of wasted class time in each core subject.”²² Disadvantaged students were the hardest hit. “Classrooms that served predominantly students from higher income backgrounds spent twice as much time on grade-appropriate assignments and five times as much time with strong instruction, compared to classrooms with predominantly students from low-income backgrounds,” the study found.

A more comprehensive and rigorous study, conducted by Thomas J. Kane of Harvard University and David Blazar of the University of Maryland, examined data on student achievement and math textbook adoptions in six states over the course of three school years and found “little evidence of differences in average math achievement growth in schools using different elementary math curricula.”²³ However, the pair also reported that while the vast majority

of teachers used their school’s official curriculum in more than half of their lessons, few used it exclusively. Even more pertinently—perhaps ominously—they found only “modest” amounts of teacher professional development on the adopted textbooks and curriculum.

“Some may interpret our findings as implying that curriculum choice does not matter. We believe that would be an overstatement,” wrote Kane, Blazar, and their colleagues. “It is true that, at current levels of classroom implementation, we do not see that schools using different textbooks or curriculum materials differed in terms of average student achievement growth on the CCSS-aligned assessments. Yet, it is possible that, with greater support for classroom implementation, the advantages of specific curricula would emerge and we would see larger differences” (31).

In a commentary on the report, Kane and David Steiner compared teachers’ use of newly adopted curricula to a smartphone with “mysterious functions” that most consumers never use, preferring instead to dabble with the unfamiliar device. “With the new curricula, we have handed teachers a tool much more complicated than any smartphone, one that holds great promise but requires complex behavior changes. And we have largely left them to figure it out on their own,” the pair observed. “The average teacher received only 1.1 days of professional development devoted to their curriculum during the 2016–17 school year and 3.4 days when including prior years.” Rather than conclude that curriculum has no effect on student outcomes, the pair wisely prescribed a “call to action,” stating the following:

Education policymakers can no longer simply exhort schools and districts to implement curricula more thoroughly. We need to provide clearer guidance on an effective transition to more-rigorous curricula. States, districts, and the national philanthropies who have been supporting the better materials should test different packages of supports—with different combinations of professional development for teachers, training for principals on what to look for during classroom observations, classroom coaching, videotaped practice sessions with teachers—and identify the suite of supports necessary to generate closer adherence to the curricula and to boost student achievement.”²⁴

In sum, spending hours on customizing curriculum or creating units and lessons from scratch is burdensome, results in lessons of low rigor and quality, and is almost certainly a less valuable use of teacher time than studying student work, giving feedback, developing subject matter expertise, and building relationships with students and their families. The bespoke nature of instructional planning is a standard feature of a teacher’s job in the United States but one that lacks evidence of efficacy.

Change does not have to come in the form of a centralized, top-down curriculum imposed by distant state governments, but as the guarantor of public education, states have a clear interest in ensuring that the best curricula are adopted and implemented. In the early 2010s, under the tenure of then superintendent of education John White, the state of Louisiana pioneered an approach whereby the state department of education, in partnership with teachers across the state, evaluated dozens of math and ELA curriculum programs, sorting them into three quality tiers and publishing the results online. Districts that adopted the top “Tier 1” programs

were given financial incentives and state-provided professional development to encourage adoption, thus, in the words of one state education official, “making the best choice the easy choice.” Initial results were encouraging: the 2015 NAEP tests showing that “Louisiana 4th graders showed the highest growth among all states . . . and the second-highest in math” provide suggestive evidence that curriculum reform holds promise as an effective weapon for raising test scores.²⁵

Crucially, there is broad support among teachers for the adoption and implementation of HQIM. In a May 2022 report on the availability of HQIM, EdReports, the leading reviewer of instructional materials in the United States, surveyed teachers and asked how important it was for the materials they use to be aligned with standards. Nearly three-fourths (73.3 percent) of teachers said it was “extremely important,” and another 20.9 percent said it was “somewhat” important.²⁶ Even so, the actual use of HQIM in classrooms is shockingly low. EdReports also found that only 25.6 percent of teachers used at least one ELA standards-aligned material per week, with that percentage rising to 39.7 percent for math standards materials.²⁷ This reinforces the impression that teachers lack the discernment when choosing or customizing curricular materials to exercise their judgment appropriately. Reducing this gap between the belief in using HQIM and the actual use of those materials is a key priority in making sure students get the best possible education.

THE BENEFITS OF A CORE CURRICULUM

For decades, the renowned education theorist E. D. Hirsch Jr. has argued for the adoption of a core curriculum across academic disciplines and from the first days of school based on a belief, firmly grounded in cognitive science, that a shared body of knowledge is an essential building block of literacy in an economically and socially diverse country.²⁸ Hirsch’s argument emphasizes that a core curriculum would help bridge gaps in knowledge among students from diverse backgrounds, promote social cohesion, and equip individuals with the foundational knowledge necessary for success in various academic and professional pursuits, while contributing to the broadly embraced goal of building students’ critical thinking and problem-solving skills. Hirsch’s vision suggests that there should be far more similarities than differences in the student experience regardless of where a child attends school and regardless of their socioeconomic status. Specifying the knowledge and skills that children should share and which their education would give them fair and equal access to would promote educational excellence and equity, allowing all to engage meaningfully in society, understand complex texts, and communicate effectively.

It is beyond the reach of even the most determined policymaker to impose a single curriculum on America’s K-12 education system, however, simply because no such system exists. The word “education” famously appears nowhere in the US Constitution, thus devolving the responsibility to the individual states. The political unpopularity of Common Core State Standards, commonly mistaken for a national curriculum, demonstrates the hostility to anything approaching federal control of classroom content. Our tradition of local control of education makes it unlikely in the

extreme that the United States will ever adopt the kind of national curriculum common to many other nations, including those whose academic performance easily outpaces our own. Any significant momentum to make curricular content more consistent from state to state, from district to district, and even across the hall in the same school will have to be driven by reforms from within the field of education itself. The strongest argument for this is that it is simply too much to ask of teachers to be effective at both curriculum design and delivery. It places undue burden on teachers with no evidence of effectiveness for students.

In his 2016 book *Leadership for Teacher Learning*, Dylan Wiliam observes that when teachers are asked to identify something that they will stop doing or do less of to create time and space for them to explore improvements to their teaching, they fail miserably. “They go through the list of their current tasks and duties and conclude that there is nothing they can stop doing or do less of, because everything that they are doing contributes to student learning,” he writes.²⁹ “In my experience, it is hardly ever the case that teachers are doing things that are unproductive. This is why leadership in education is so challenging. The essence of effective leadership is stopping people from doing good things to give them time to do even better things.”³⁰

What might those “better things” include? For my own 2019 book, *How the Other Half Learns*, I spent a year embedded in a high-performing Success Academy charter school in New York City’s South Bronx, a few blocks from where I’d taught fifth grade in a struggling elementary school run by the New York City Department of Education some years earlier. Success Academy, a network of approximately fifty charter schools based in New York City, has distinguished itself with an outstanding track record of high achievement among the predominantly low-income, minority students it serves. If it were a stand-alone public school district, its test scores would make it New York State’s best-performing district, despite the high concentration of poverty in most of the neighborhoods in which its schools are located.

The purpose of the book was to see what lessons could be gleaned from Success Academy’s instructional model and exported to K-12 public education at large. In the course of my reporting, I interviewed a pair of hedge fund managers who had written the charter school application for what was originally called Gotham Charter School. Joel Greenblatt and John Petry focused their efforts on creating a school model that was replicable, a conviction reinforced by visits to high-performing New York City charters like KIPP Infinity Middle School in Harlem. “We were watching this amazing English lesson. I elbowed the assistant teacher and said, ‘Just for my interest, what’s the background of that teacher?’” Greenblatt told me in an interview for the book. “‘He wrote for [the television sitcom] *Frasier* for five years,’” she replied. “It would be wonderful if everyone could have that teacher, but everyone can’t,” Greenblatt explained.³¹

This insight was out of step with the education orthodoxy of the time and the orthodoxy of the “no excuses” charter school movement in particular. But it found its way under CEO Eva Moskowitz into the model of Success Academy, which implements a single curriculum across its network, creating ripple effects in the classroom. Like teachers everywhere, Success Academy staff spend a significant amount of their time preparing lessons for students. However, where lesson planning for many, even most, US teachers involves lesson creation

or customization, at Success Academy it's referred to as "intellectual preparation." As the name implies, it means preparing to teach a lesson or unit, not creating one from scratch. This appears to greatly contribute to one of Success Academy's most significant accomplishments: its ability to get uniformly good results from relatively inexperienced teachers. It is important to note that Success Academy's curriculum is not "scripted." Teachers work from an established curriculum—units and lesson sequences prepared by network-level staff. Success Academy's model asks teachers to focus on *teaching*, not on creating curriculum or gathering instructional resources. Compared to the more common practice of lesson planning as lesson design, Success Academy's model functionally creates hours of capacity for teachers to look at data, study student work, diagnose, act, adapt, and intervene quickly when students are struggling or falling behind. As Moskowitz told me at the time, "You can't be successful in our model without studying student work."³²

One of the promises of the US charter school movement is that those schools might serve as laboratories for innovative instructional practices and strategies.³³ Adopting a core curriculum and asking teachers to focus on lesson delivery, not on lesson design, and asking them to study student work and become effective diagnosticians and interventionists is one such practice that appears worth emulating.

CONCLUSION

Gifted musicians in an orchestra generally do not write the symphonies they play. We do not think less of a talented actor who merely performs but did not write a Shakespearean tragedy. Great chefs need not be farmers, butchers, or fishermen. Teachers, by noteworthy contrast, are expected to be both expert lesson deliverers and instruction designers. Significantly, this expectation is something the profession tends to valorize; it didn't drift into it. Teachers College at Columbia University, for example, among the premier teacher preparation programs in America, maintains as a core tenet of its Master of Arts in Curriculum and Teaching program that "teachers are necessarily and rightly adapters and designers of curriculum."³⁴ There is no evidence that this practice benefits students.

Making curriculum an afterthought in our efforts to improve student outcomes and giving insufficient professional development to properly implement those materials when they are adopted makes a hard job nearly impossible and virtually enshrines poor performance and teacher burnout as policy. As Marcy Stein, an education professor with expertise in evaluating instructional design at the University of Washington-Tacoma, puts it, "Few teachers ever take coursework on instructional design and, therefore, have little knowledge of the role it plays in student learning." Further reflecting on the many extraneous burdens we place on teachers, Stein notes:

Even if teachers were taught about instructional design, they would likely not have the time to prepare instructional materials, field test those materials to determine if they are effective, and modify the materials before using them to teach students. An iterative process is crucial for the development of effective materials.³⁵

Readers might be tempted to see in between the lines of the preceding quotation an argument for the elimination of teacher autonomy or even a case for “McSchool,” a basic education deliverable by teachers of minimal competence and cognition who must be spoon-fed a scripted curriculum. Having anticipated this argument, let me put it to rest. An idea that is common to teacher training and professional development is that there should be a “why” behind everything a teacher does in the classroom, from classroom management to instructional decisions. The same principle applies here: the point is not for school districts to adopt a curriculum and for teachers to deliver it robotically. Well-prepared teachers should acquire through their training and professional development a sophisticated understanding of their subject matter and pedagogy and have it *operationalized* for them in the form of a curriculum or program.

The one thing we cannot give teachers is more time. It is too much to ask of teachers to create their own programs and curricula; the opportunity costs are profound. The twofold challenge for policymakers is to privilege the adoption of an effective curriculum but also for teachers to understand *why* it’s effective. It cannot be denied that the current culture of teaching largely expects teachers to both deliver and design curricula, making it less likely they do either expertly. This de facto demand tends to result in a student experience that is incoherent, marked by curriculum gaps and repetitions and by lessons that are lacking in rigor. The principal point here is to return to teachers time spent needlessly or excessively planning units and lessons from scratch so that they can spend more time on higher-yielding activities: studying student work, giving feedback, deepening their subject expertise, and building relationships with students and families.

The ed reform era, now receding in power and prestige, overestimated the impact of education policy and “structural reform” to improve student outcomes. Renewed efforts to improve school performance must focus on instructional reform. This does not imply, however, that there is no role for policymakers, only that they must be clear-eyed about their leverage and its limits, keeping in mind these key takeaways:

- Efforts to improve student outcomes by changing the composition of the teacher workforce or dramatically raising their level of sophistication and skill are unlikely to be successful due to the large number of teachers needed to staff our schools. What must change is the job: teaching must be doable by women and men of ordinary talents and sentience.
- While the evidence base is insufficiently robust to say with certainty, there is ample reason to suggest it is easier, less expensive, and more effective to change curricula than to change teachers.
- The soul of effective teaching is studying student work, giving effective feedback, and developing relationships with students. Teacher time spent on curating and customizing lessons, however valuable, takes time away from these more impactful uses of teacher time. The adoption of a high-quality curriculum and training on its effective implementation is the first, most critical step toward transforming the teacher’s job.

- Education occurs in a public context; there will always be a role for policymakers to ensure accountability. However, improvements at scale will not be wrested from rewards and punishments, nor from other “structural” reforms.

There are, finally, encouraging signs in both education policy and practice that suggest a new willingness to question long-standing classroom practices and that curriculum and instruction are gaining traction as reform levers, at least in early childhood education. In the past few years, the science of reading (SoR) movement has emerged as one such welcome development in literacy education.

Drawing upon a deep research base in a variety of fields, including cognitive psychology, linguistics, and neuroscience, the movement has sought to reorient teacher training, curriculum adoption, and teacher professional development around evidence-based instructional practices, particularly foundational skills such as phonemic awareness, phonics, vocabulary, fluency, and comprehension. Critically, the SoR movement seeks to address the shortcomings in teacher preparation and professional development.

Indeed, much of the movement’s energy and focus have been directed at colleges of education, which have come under harsh scrutiny for their failure to adequately prepare teacher candidates to successfully teach reading. In a 2023 report by the National Council on Teacher Quality (NCTQ), which monitors and reports on teacher preparation programs, Dr. Heather Peske, NCTQ president, stated, “We’re in the midst of a long overdue revolution on the science of reading, but teacher prep programs haven’t fully caught up.”³⁶ The report studied nearly seven hundred teacher preparation programs across the country, looking for evidence that coursework for future elementary school teachers included “core components of scientifically based reading instruction.” NCTQ determined that only 25 percent of the programs it evaluated adequately cover all five core components of scientifically based reading instruction: phonemic awareness, phonics, fluency, vocabulary, and comprehension. “Perhaps more alarmingly, another 25% of programs do not adequately cover even a single component,” the report noted.³⁷

What is true of teacher training can be said with equal certainty about curricula. It is simply too much to expect teachers to be able to effectively operationalize the science of reading in the absence of effective curriculum and professional development on its implementation. As many expert observers have noted, the SoR movement will ultimately succeed or fail not just on early literacy skills but through the adoption and implementation of the kind of “knowledge-rich” curriculum E. D. Hirsch Jr. and others have championed for decades. This requires collaboration and coordination by teachers within and across grades and over many years of schooling. It is literally impossible to accomplish if teachers simply close their doors and teach what they like. Thus, the SoR movement offers both a test case and a potential proof point for many of the arguments made above.

A strong, knowledge-based curriculum not only leads to smarter teachers and students but also has beneficial downstream effects, such as giving teachers the opportunity to develop subject matter expertise, help struggling students, and make connections with families rather

than spending hours each week scouring the internet and creating lesson plans from scratch. Let's not ask what more teachers can do. Let's ask instead what are the things that only a teacher can do. Everything else should be a job for someone else.

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.

The use of HQIM in every school district (not lip service to use, but real use) would be a strong, positive step in education practices in the United States. However, some of the difficulties with making this a reality are not mentioned in Pondiscio's very fine piece. Many teachers simply don't believe that their students can handle the newer, more rigorous materials—and in many cases, where their students may be two or more years behind grade level, the teachers are right. Much stronger pre-K-12 schooling is indispensable to making widespread use of HQIM a reality. A second issue is principal buy-in: in many cases, principals simply don't understand why curriculum reform is so important and act accordingly—for example, failing to put curriculum-specific elements into their classroom observation rubrics. Finally, a point that Pondiscio does make: professional development for teachers can't be a one-day-a-year business. If we are going to change how teachers go about selecting their instructional materials, we have to change their entire professional mindsets, and that takes multiple interventions over a long period.

—Dr. David Steiner, executive director of the Johns Hopkins Institute for Education Policy

Pondiscio's paper is a history of reform and a challenge for the future. As a person who has been in the education reform movement for more than thirty years, I recognize that system change is difficult. Clearly, the quality of the teacher does have an impact on the success of the student, but is that impact as great as that of the quality of the curriculum? In my career, I have advocated for teacher evaluations and the categorizing of the teachers based on their ability to deliver academic results. In retrospect, the teacher was the wrong target, and that reform ultimately did not yield the significant change we had hoped.

Pondiscio's focus on curriculum ultimately is a very logical approach to improve academic outcomes. We know that curriculum drives results, but curriculum without system change and professional development will likely not yield significant academic improvements. The author's reference to the science of reading evidence that is sweeping the nation is an example of how we need quality curriculum and support to change systems and teaching. Change will not occur without the entire suite of actions.

—Rep. Bob Behning, Indiana House of Representatives

NOTES

1. National Commission on Excellence in Education (NCEE), *A Nation at Risk: The Imperative for Educational Reform* (Washington, DC: US Department of Education, 1983).
2. National Assessment of Educational Progress, "NAEP Reading: Reading Highlights 2022," The Nation's Report Card, 2022, <https://www.nationsreportcard.gov/highlights/reading/2022>; National Assessment of Educational Progress, "NAEP Mathematics: Mathematics Results," The Nation's Report Card, 2022, <https://www.nationsreportcard.gov/mathematics/?grade=4>.
3. Center for Research on Education Outcomes (CREDO), "As a Matter of Fact: The National Charter School Study III 2023," June 16, 2023.
4. Emmeline Zhao, "Raw Numbers: Charter Students Are Graduating from College at Three to Five Times the National Average," *74 Million*, July 30, 2017, <https://www.the74million.org/article/raw-numbers-charter-students-are-graduating-college-at-three-to-five-times-the-national-average>.
5. Matt Collette, "A Painful Decade of School Reform," *Slate*, March 5, 2015, <https://slate.com/human-interest/2015/03/reform-fatigue-how-constant-change-demoralizes-teachers.html>.
6. Andrew Coulson, "State Education Trends: Academic Performance and Spending over the Past 40 Years," Cato Institute Policy Analysis 746 (2014), <https://www.cato.org/sites/cato.org/files/pubs/pdf/pa746.pdf>.
7. Robert Pondiscio, "Exit Interview: James Peyser in Defense of Education Reform," Thomas B. Fordham Institute, July 20, 2023, <https://fordhaminstitute.org/national/commentary/exit-interview-james-peyser-defense-education-reform>.
8. NCEE, *Nation at Risk*.
9. Richard Snyder, "Resistance to Change among Veteran Teachers: Providing Voice for More Effective Engagement," *NCPEA International Journal of Educational Leadership Preparation* 12, no. 1 (2017): 6.
10. Snyder, "Resistance to Change," 7-8.
11. Snyder, 8.
12. David Steiner, "Curriculum Research: What We Know and Where We Need to Go," *StandardsWork*, March 2017, <https://standardswork.org/wp-content/uploads/2017/03/sw-curriculum-research-report-fnl.pdf>.
13. Matthew M. Chingos and Grover J. "Russ" Whitehurst, "Choosing Blindly: Instructional Materials, Teacher Effectiveness, and the Common Core," Brookings Institution, April 10, 2012, <https://www.brookings.edu/articles/choosing-blindly-instructional-materials-teacher-effectiveness-and-the-common-core>.
14. Eric A. Hanushek, "The Difference Is Teacher Quality," in *Waiting for "Superman": How We Can Save America's Failing Public Schools*, ed. Karl Weber (New York: Public Affairs, 2010), 81-100.
15. Marcus Winters, "Measuring Teacher Effectiveness: Credentials Unrelated to Student Achievement," Manhattan Institute, September 7, 2011, <https://manhattan.institute/article/measuring-teacher-effectiveness-credentials-unrelated-to-student-achievement>.
16. "National Evaluation of Teach for America, 2001-2004," *Mathematica*, <https://www.mathematica.org/projects/teach-for-america>.
17. National Center for Education Statistics, "Characteristics of Public School Teachers Who Completed Alternative Route to Certification Programs," Institute for Education Sciences, May 2018, <https://nces.ed.gov/programs/coe/indicator/tlc#suggested-citation>.
18. Steiner, "Curriculum Research."
19. V. Darleen Opfer, Julia H. Kaufman, and Lindsey E. Thompson, "How Are Teachers Perceiving and Addressing State Standards?," RAND Corporation, January 1, 2016, https://www.rand.org/pubs/research_reports/RR1529-1.html.
20. Nicole Gorman, "Survey Finds Teachers Spend 7 Hours per Week Searching for Instructional Materials," *Education World*, February 7, 2017, https://www.educationworld.com/a_news/survey-finds-teachers-spend-7-hours-week-searching-instructional-materials-490526015.

21. Morgan Polikoff and Jennifer Dean, "The Supplemental Curriculum Bazaar: Is What's Online Any Good?," Thomas B. Fordham Institute, December 2019, <https://files.eric.ed.gov/fulltext/ED601253.pdf>.
22. "The Opportunity Myth," New Teacher Project, <https://opportunitymyth.tntp.org>.
23. Thomas Kane and David Blazar, "Learning by the Book," Center for Education Policy Research at Harvard University, September 9, 2019, <https://cepr.harvard.edu/curriculum>.
24. Thomas J. Kane and David M. Steiner, "Don't Give Up on Curriculum Reform Just Yet (Opinion)," *Education Week*, April 2, 2019, <https://www.edweek.org/leadership/opinion-dont-give-up-on-curriculum-reform-just-yet/2019/04>.
25. Robert Pondiscio, "Louisiana Threads the Needle on Ed Reform," *Education Next*, July 25, 2017, <https://www.educationnext.org/louisiana-threads-the-needle-ed-reform-launching-coherent-curriculum-local-control>.
26. EdReports, "State of the Instructional Materials Market 2021: The Availability and Use of Aligned Materials," May 25, 2022, <https://www.edreports.org/resources/article/state-of-the-instructional-materials-market-2021-the-availability-and-use-of-aligned-materials>.
27. EdReports, "Instructional Materials Market 2021."
28. Daniel Willingham, "How Knowledge Helps," American Federation of Teachers, July 18, 2023, <https://www.aft.org/ae/spring2006/willingham>.
29. Dylan Wiliam, *Leadership for Teacher Learning: Creating a Culture Where All Teachers Improve So That All Students Succeed* (Bloomington, IN: Solution Tree Press, 2016), 180.
30. Wiliam, *Leadership*, 180.
31. Robert Pondiscio, *How the Other Half Learns: Equality, Excellence, and the Battle over School Choice* (New York: Penguin, 2019), 75.
32. Eva Moskowitz, interview by author, New York City, 2018.
33. Richard Kahlenberg and Halley Potter, "Restoring Shanker's Vision for Charter Schools," American Federation of Teachers, July 17, 2023, https://www.aft.org/ae/winter2014-2015/kahlenberg_potter.
34. "Master of Arts in Curriculum and Teaching," Teachers College, Columbia University, accessed August 28, 2023, <https://www.tc.columbia.edu/curriculum-and-teaching/curriculum-and-teaching/degrees--requirements/curriculum-and-teaching-ma>.
35. Robert Pondiscio, "How We Make Teaching Too Hard for Mere Mortals," *Education Next*, May 16, 2016, <https://www.educationnext.org/how-we-make-teaching-too-hard-for-mere-mortals>.
36. National Council on Teacher Quality (NCTQ), "New Data Finds Major Gaps in Science of Reading Education for Future Elementary Teachers," 2023, <https://www.nctq.org/publications/New-Data-Finds-Major-Gaps-in-Science-of-Reading-Education-for-Future-Elementary-Teachers>.
37. NCTQ, "New Data."



6. The World Is a Lab

Innovation in Schools after A Nation at Risk

Eric Bettinger

Executive Summary

Education needs innovation. *A Nation at Risk* applauded the notion of schools and districts as local laboratories, where they could develop, scale, and eventually disseminate innovative practices. This chapter reviews how innovation in the organization and delivery of education, both inside and outside of the classroom, has affected the quality of education. No review can cover the full range of innovative practices, and I focus attention on districtwide innovations such as the small schools initiative, magnet schools, superstar superintendents, and innovation zones. On in-school innovations, I review evidence on class size and instructional time. In each case, I demonstrate how local experimentation led to adoption and expansion of particular practices. For example, success with small schools led to massive philanthropic and government investment across states and municipalities.

However, while many of the innovations show promising results in local experiments, scaling and expansion have not always replicated local success. While the conditions of local experimentation focus on questions of efficacy in terms of improving educational outcomes, policymakers have to consider cost-effectiveness. A major reason for the unsuccessful expansion or replication of local results is that the cost of the innovative practices can be prohibitively large. Schools and districts must often sacrifice other programs to transfer funds to these innovative practices. Moreover, local labor markets cannot always provide the teachers, administrators, and infrastructure needed to replicate the results from local experimentation.

While *A Nation at Risk* was correct that local school systems are laboratories, the process of adopting, expanding, and scaling local experiments has proven more difficult because of the conditions required to replicate the innovation. While administrators can look for solutions in

experiments conducted throughout educational systems globally, adoption of these practices should consider local contexts. Moreover, administrators should verify that the new programs can replicate the prior results when scaled and transferred.

- Innovations that promise to revitalize American education gain attention and often are promoted elsewhere with massive investment.
- Seldom do these innovations accomplish as much elsewhere as they did in the original experimentation, partly because the cost to implement is often prohibitively large.
- Finding innovations that can be replicated at scale is the next challenge to be solved.

• • •

Education thrives on innovation. While core practices in education have slowly and incrementally changed over time, the innovation and creativity of teachers and administrators have facilitated improvements by modifying the organization and delivery of education. In calling for solutions, *A Nation at Risk (ANAR)* repeatedly called on local “political and educational leaders to search for solutions” and to use the “ingenuity of our policymakers, scientists, State and local educators, and scholars in formulating solutions” (NCEE 1983, 12, 15). Indeed, *ANAR* lauded the “local laboratory” model that decentralized schooling affords to test new models that could scale and disseminate throughout the country. This chapter takes aim at understanding how innovation in the organization and delivery of education both within and outside the classroom has affected the quality of education.

I start by looking at large-scale initiatives that sought to modify the ways school districts organized themselves. For example, the Bill and Melinda Gates Foundation provided significant funding for the small schools or the “school-within-a-school” initiative. The initiative sought to divide large high schools into smaller sub-high schools that might provide a greater cohort experience and allow for greater interaction between teachers and groups of students. After reviewing small schools and other out-of-the-classroom innovations, I turn my attention to innovation within the classroom. Since *ANAR*, major evidence on class size and the timing of schooling, among other innovations, has influenced policy and practice. This has affected how states and local districts organize and deliver education to students inside schools and classrooms. I review evidence on these and other innovative practices.

It is impossible to really track forty years of innovation across thousands of school districts. Many innovations were never disseminated, scaled, or evaluated. Other innovations became so popular and widespread that the editors of this series chose to dedicate entire chapters to them. For example, the rapid changes in technology with computers and the integration of standards and accountability systems are two innovations that have altered every classroom across the United States. In this series, Tom Vander Ark’s chapter on technological innovations and Michael J. Petrilli’s chapter on standards and accountability review these innovations and how they have shaped school organization, classroom practice, and teacher and principal

accountability (see chapters 7 and 11, respectively). This chapter provides a useful companion to those chapters.

In this chapter, I focus on four sources of system innovation and two sources of classroom innovation. For system innovation, I review evidence on small schools, magnet schools, superstar superintendents, and innovation zones. For classroom innovation, I discuss evidence on class size and the duration of schooling. In choosing the specific innovations, I have conveniently sampled practices that have expanded beyond a single district or school and have thereby shaped and influenced education policy and practice. I also try to focus on innovations where rigorous research has provided some hint as to the causal impacts of these policies.

While many of the innovations seem to have strong evidentiary bases, they remain seemingly underutilized. In some cases, the costs of the interventions remain prohibitively large. In other cases, the modifications needed to establish the interventions at scale compromise the capacity of the interventions to affect outcomes. I discuss other impediments to expansion, and more generally I discuss obstacles that inhibit the use of the “school laboratory” model.

SYSTEMS INNOVATION

“Systems innovation” refers to new policies and practices that modify governance and the structure of school systems. The specific innovations upon which I focus are those in school organization, creation of specialized schools, school leadership, and innovation zones, which provide schools and districts latitude to implement new practices and policies.

SCHOOL SIZE AND ORGANIZATION

Perhaps the most notable of the systemwide changes was the small schools movement. In the early 2000s, nearly twenty-six hundred small schools were created nationwide (Ravitch 2008). One of the major forces behind this investment was the Gates Foundation, which began in the early 2000s to fund the decomposition of large high schools. Based on its interpretation of existing literature, the Gates Foundation encouraged schools to maintain a size of four hundred students. As Tom Vander Ark, who at that time was executive director of education at the Gates Foundation, said, “Young people who attend smaller schools that provide a rigorous, personalized education and enable close relationships with adults are more likely to graduate and continue their education” (Bergstein 2003). The basic theory was that personalized education and deeper relationships with teachers would improve the quality of education and provide better role models and coaching as students began considering and then pursued subsequent education.

Unfortunately, the short-term results did not generate the anticipated results. Over nine years, the Gates Foundation invested more than \$2 billion in creating small schools, and in 2009, it “refined” its strategy. As Bill Gates wrote: “[M]any schools had higher attendance and

graduation rates than their peers. While we were pleased with these improvements, we are trying to raise college-ready graduation rates, and in most cases, we fell short” (Gates 2009).

Indeed, the short-term evidence on small schools was a bit pessimistic (Ravitch 2008). At the time that Gates was making allocative decisions, the evidence was not strong. However, the tide of positive evidence was soon to come. Studies by Bloom et al. (2010), Schwartz et al. (2013), Barrow et al. (2013), and Abdulkadiroğlu et al. (2013) found positive, long-run impacts. The study by Abdulkadiroğlu et al. (2013), the only one that took advantage of randomized admission lotteries, reported that college attendance rates in New York improved by seven percentage points, with additional improvements in math and English scores on the state’s High School Regents Examinations. Additional work by Schwartz et al. (2013) argued that small schools improved the performance of most New York high schools, *including* schools that were not small schools.

Yet despite the positive, long-run evidence on small schools, they are no longer receiving the same public and philanthropic support they did in the past. Strong evidence has not revitalized the initiative or the funding streams. While existing small schools have remained, few additional small schools have been added. Why the lack of continued or renewed support?

First, there was a perception that lessons could be applied and scaled up in other settings that did not require small schools. For example, one theory as to the success of small schools relied on the notion that students developed rich, meaningful, personalized relationships with faculty and counselors. However, such relationships may be possible in other settings as well. As Robert Hughes, director of K-12 Education at the Gates Foundation, explained, “[W]ith some work, you can really build structures that enable kids to be known and to get the kind of support they need to be successful [even] in larger schools” (Kolodner 2015). Also, small schools cost more per student (Schwartz et al. 2013) than traditional schools. If traditional schools can replicate the counseling and other relationship-based mechanisms, then expanding small schools, a far more expensive and involved intervention, may not be necessary. Indeed, the fact that all schools in a school district, including large schools, benefit from the presence of small schools suggests that small schools increase the visibility of some mechanisms that can be transferred to larger schools.

Second, there have been concerns about the capacity of institutions to scale small schools. One concern is that the cost of staffing might be prohibitively large (Schwartz et al. 2013). Another concern is that small schools raised the demand for both teachers and principals, and it is unclear that the supply of new teachers and qualified principals can satisfy the demand.

While small schools remain somewhat stalled, there are several aspects of the small schools movement that have had a lasting impact. For example, it was one of the largest, highest-profile experiments in innovating the structure and design of schools. This gave some momentum to districtwide interventions and experimentation, expanding the scope of the laboratories envisioned in *ANAR*. Also, as mentioned above, it provided significant information about

reforming underperforming schools that could be applied in other settings. And there was also important heterogeneity in the impacts that provided additional policy lessons about the implications of certain types of schools—namely charter schools, which generated significant impacts. For example, the KIPP schools were cited by Gates in the 2009 announcement that the Gates Foundation was refining its investment strategy. Subsequent work such as Abdulkadiroğlu et al. (2013) demonstrated that there were specific strategies (e.g., high accountability) employed by charter schools that may have led to the greater impacts observed in some small schools as opposed to others. Each of these lessons reinforced the notion that school districts could be local laboratories for innovation.

SPECIALIZED SCHOOLS

A second strategy that altered the ways districts organized schools had to do with school choice and the underlying supply of schools. In this series, the chapter by John D. Singleton focuses extensively on school choice (see chapter 10), so I defer any discussion of charter schools and vouchers to that chapter. My discussion here centers on magnet schools. While many charter schools behave like magnet schools and vice versa, the administration of magnet schools typically continues under the direction of school district offices, whereas the administration of charter schools often moves outside the district’s purview.

Magnet schools existed long before *ANAR*. As districts grappled with how to desegregate schools, many created specialist or alternative schools to give parents additional options. The first magnet schools began in the late 1960s. The first large-scale experiment occurred in 1970 in Minneapolis (Waldrip 2023), and the first specialized high school, focused on career themes, opened in Dallas in 1971.

As of 2016, there were 4,340 magnet schools across the United States, with the most common theming centered on science, technology, engineering, and mathematics (STEM); fine and performing arts; international baccalaureate; career and technical education; and world languages. Whereas before *ANAR*, magnet schools were often considered as a way to encourage desegregation, the major expansion and specialization of magnet schools took place after *ANAR* with greater emphasis on school choice.

The magnet school expansion happened for a variety of reasons. Some of it was driven by parents as they tried to find ways to improve their children’s educational performance by building on specific skills. Some of it came as magnet schools, particularly vocation-oriented magnet schools, demonstrated that they could have positive impacts on students who were struggling in mainstream classrooms.

To date, the evidence on magnet schools is largely positive. Gamoran (1996) uses the National Education Longitudinal Study to compare test scores of students at magnet, public, and private schools. Gamoran found that students at magnet schools score higher in science, reading, and social sciences. Crain et al. (1998) use oversubscription lotteries to measure the impact of career-oriented magnet schools. The researchers provide mixed evidence. On the one hand, career magnet schools had lower graduation rates than comprehensive schools.

This was the result of greater emphasis on career and vocational curricula. On the other hand, students who attended magnet schools reported fewer “reckless adolescence behavior” at age twenty. Kemple and Snipes (2000) provide evidence that career magnet schools are an effective way to reduce dropout rates among those at the highest risk not to graduate. A synthesis of the literature suggests that impacts are “generally positive” (Wang et al. 2018).

Magnet schools were refined in the laboratory of public schooling, and new iterations of magnet schools build upon the principles of school choice discussed by Singleton (in chapter 10) and on principles of innovation. Career education in particular has become much more central to education policy. As the returns to high school education have stagnated, emphasis on employability and skills has fueled much of the advance in career education. This concern not only has been present in the United States but also has become increasingly popular as a policy tool in developing countries (see review in MacDonald and Dunbar 2015). The evidence from magnet schools provided significant lessons that shaped early attempts to strengthen vocational education.

Magnet schools also raise the question as to whether education should be similar across students. In many ways, the traditional school model presents very little variety across the types of skills that students develop; however, magnets exist to allow some students to specialize beyond what a traditional school might allow. Heterogeneous students might need more heterogeneous offerings than traditional schools can provide. Magnet schools might be a way to improve the efficacy of education for a subset of students, and there may be limits to the degree of differentiation among students and schools that are possible. Hence, the gains could be large but diminishing as differentiation expands.

In the short run, magnet schools’ enrollment and presence will continue to expand. Whereas they were at one point an answer to desegregation and integration of schools, they are increasingly a means for parents to express their preferences in terms of students’ education opportunities. Their growth continues, and while charter school enrollment remains larger, magnet schools remain a viable way to provide differentiated education opportunities. Moreover, as charter schools become more specialized (e.g., STEM or vocation focused), the line dividing magnet and charter schools will continue to blur.

SCHOOL LEADERSHIP

Another trend that has taken place since *ANAR* is the increased emphasis on high-profile school district superintendents. In many large school districts, superintendents have become chief executive officers with greater power and salaries. For example, Barbara Byrd-Bennett served as the CEO of both Cleveland Public Schools and, later, Chicago Public Schools. She previously served in a leadership capacity for New York City schools. Her compensation in both Cleveland and Chicago was controversial, given the size of the packages (O’Donnell 2012).¹ Other CEOs grabbed headlines and made national news as well, such as Michelle Rhee in Washington, DC, and Arne Duncan in Chicago Public Schools. In districts with more than twenty-five thousand students enrolled, superintendent compensation ranges from \$140,000

to almost \$400,000 (McCord and Finnan 2019). To put this in perspective, the median base salary for a beginning teacher in districts of the same size is \$44,150.

Prioritizing hiring high-profile CEOs with extensive experience and increasing their compensation has been a prevailing societal trend (Bivens and Kandra 2022). Just as CEOs' track records were believed to have an impact on a company's performance, schools sought to enhance their quality by appointing elite superintendents who would bring about substantial improvements and elevate the institutions they served. Hence, an increasing number of school districts applied corporate strategies to superintendents. In the corporate world, there was a perception that the supply of such leaders was finite, leading to bidding wars and large compensation. The emphasis on superstar superintendents faced the same competition and compensation.

However, as Chingos et al. (2014) demonstrated, superintendents who bring about significant, statistically reliable changes in student achievement within their districts, while controlling for other factors that affect academic performance, are indeed rare. They found that superintendents account for only a tiny fraction (0.3 percent) of student differences in achievement, significantly less than other factors such as student characteristics, teachers, schools, and districts. They further indicated that student achievement does not improve with longer superintendent service, and hiring a new superintendent does not lead to immediate gains in student achievement.

Increased emphasis on superintendents may not directly yield higher test scores, but nevertheless it remains an area of continual research. For instance, Hart et al. (2019) proved that encouraging superintendent longevity can support student achievement, as those with more in-state experience possess a comprehensive understanding of the state's curriculum, testing programs, and the organizational stability required for effective leadership. Mitigating superintendent turnover, as suggested by Grissom and Mitani (2016), could involve considering salary increases, particularly in smaller and rural districts and those with lower student achievement, as this would help retain superintendents who are often lured by higher-paying positions in larger, more urban districts with better academic performance.

The hiring and reliance on superstar superintendents is very much an experiment in progress. While some districts have moved away from the strategy of hiring CEO-like superintendents in favor of other approaches, there are still districts that continue to explore this path. Ongoing research and the findings regarding superintendent longevity and compensation emphasize the importance of considering contextual circumstances being faced or the necessity of exploring alternative strategies.

INNOVATION ZONES

While the principles of innovation zones may have been part of the policies dating back to the early 1990s, states and school districts began implementing legislation in the mid-2000s. Innovation zones are schools or districts to which states or districts grant greater

autonomy over curriculum, budgeting, and staffing. Typically, states and districts grant innovation zones additional relief from other state and local regulations. While schools are free to enact policies and practices that differ from the norm, the schools are held accountable for improvements in student outcomes. After some early experimentation, innovation zones began expanding, and by 2017, they covered more than 108 schools and 63,000 students (Iyengar et al. 2017).

Innovation zones presuppose that regulation and centralization impair the ability of a district to try new and innovative practices. By providing greater autonomy, schools and districts can explore new practices in finance, governance, curriculum, and staffing. This greater autonomy comes at the cost of higher accountability for student outcomes, and schools and districts can lose the autonomy if student outcomes do not improve. More generally, innovation zones are just one category of school turnaround.² Under the Race to the Top (RttT) legislation, the federal government funded school turnaround strategies that included other variants such as school improvement grants or No Child Left Behind (NCLB) waivers. These school turnaround programs also allowed school districts to have more autonomy in some aspect of staffing, management, and curriculum (see, e.g., the cases of New Orleans in Ruble 2015 and Philadelphia in Gill et al. 2007).

Many studies of innovation zones are emerging. Zimmer et al. (2017), for example, examine the innovation zones established in Memphis, Nashville, and Chattanooga, Tennessee. They showed that innovation zones significantly outperformed other public schools and other alternative methods of changing governance. Math scores, for example, increased by 0.20 standard deviations in innovation zones relative to other schools. Science and reading scores also increased. Zimmer et al. (2017) argue that one of the largest mediating factors in the innovation zone was the retention of experienced, successful teachers. Innovation zones in Tennessee generally offered significant raises for teachers who transferred to innovation zone schools. Teachers who previously had significant value added in the classroom were more likely to shift into these innovation zone schools. While the competition for high-achieving teachers may be a zero-sum game in the short run, the responsiveness to compensation incentives alongside the added autonomy may strengthen the overall workforce by inducing the retention and recruitment of top teachers.

The use of innovation zones and other strategies aimed at strengthening school autonomy remains a hot topic. In this series, Michael T. Hartney, for example, explores other innovations in governance and how they have played out (see chapter 9). While the current scale of innovation zones is low, the case of innovation zones is interesting as the initial results have encouraged continued expansion, with at least twenty-five states having adopted policies encouraging innovation zones in districts that were previously classified as failing and more considering legislation to allow innovation zones (see Jones and Chambers 2021).

The expansion of innovation zones raises questions about teacher supply. If, as Zimmer et al. (2017) argue, the mechanism by which innovation zones improve outcomes is through attracting top teachers at the cost of having other schools lose top teachers, then innovation zones

might lead to a continued division between high-value-added teachers and others. If the higher wages and reduced legislation in innovation zones serve to attract more (and better) teachers to the profession, then innovation zones might generate momentum toward improving the overall teaching pool. However, if this does not happen, then the competition for teachers is a zero-sum game in which the available teachers for underprivileged schools will be disadvantaged relative to those who want to attend innovation zone schools.

Finally, the continued expansion of innovation zones has two implications for the future. First, the continued expansion of legislation allowing innovation zones suggests that these zones will become increasingly visible in the future. Second, given that the emphasis on innovation zones is both deregulation and expanded accountability, it also suggests a growing discontent with the existing regulations in traditional school districts. Innovation zones are a means of circumventing some regulations. If innovation zones eventually create momentum around deregulation, then deregulation might displace (or potentially devalue) innovation zones.

CLASSROOM INNOVATION

I next turn to classroom innovation. I focus on two separate innovations—class size and the timing of schooling. As before, these are only a fraction of the possible innovations that I could use; however, these are two areas where significant experimentation and subsequent policy implementation have happened since *ANAR*.

CLASS SIZE

Scholars from all disciplines have long postulated that class size affects student outcomes. The underlying theory suggested that teachers can give more attention to students in smaller classes and that this extra attention might provide a boost in students' education outcomes. Lazear (1999), for example, uses a model to demonstrate that the probability of classroom disruptions likely increases as class size goes up.

However, in the mid-1980s, some doubt emerged on the relationship. In a series of papers, Hanushek (e.g., 1986, 1999) showed that estimates of the effects of class size were ambiguous. Perhaps students were not as sensitive to class size as they might have been to other inputs, or perhaps teachers used different technologies as class size changed. Nonetheless, the relationship between class size and academic achievement has been hotly contested in the education literature (see, e.g., Mishel and Rothstein 2002; Hoxby 2000; Angrist and Lavy 1999).

In 1985, then governor Lamar Alexander sponsored the Tennessee STAR experiment. The experiment created small classes in kindergarten through third grade and implemented the intervention with a school-based randomized controlled trial. The results of the experiment were stunning. Education test scores improved by roughly 0.25 standard deviations, roughly one grade level higher than students in regular classrooms (Mosteller 1995). Subsequent

research (see review by Schanzenbach 2014) suggested that the impacts endured through primary, secondary, and tertiary schooling.

Some criticisms have been made of the Tennessee STAR experiment. For example, Hoxby (2000) discusses the possibility that the results are exaggerated. She argues that they may be the result of the Hawthorne effect arising from teachers performing differently than they would have otherwise as a result of participating in a high-profile experiment. Others have refuted this characterization, calling the Tennessee STAR experiment the “Barbary steed” of the class size literature (Krueger 2003). Nonetheless, Tennessee STAR influenced policymakers. States including California, Florida, and Texas established class size limits. The policy in California in particular provided extensive financial incentives to schools that implemented class size limits (Schanzenbach 2014).

While the policy debate has leaned heavily in recent years toward reductions in class size, there have not been significant studies to date documenting whether state policies around class size have generated close to the same effects as observed in Tennessee STAR. In fact, there is some evidence that the emphasis on class size has come at the expense of other inputs. For example, Sims (2008) shows that California schools largely achieved class size reductions up to grade three by increasing class sizes in subsequent grades. He shows that the increase in test scores after grade three may have reversed potential positive impacts of class size.

The class size debates are not over and will likely continue for the foreseeable future. Until evidence can definitely show that the expansion of reduced class size through state policies leads to sustained improvements in student achievement, the debate over class size will continue. Even if Tennessee STAR’s evidence shows improvements in student outcomes as a result of class size reductions, it does not mean that class size reductions can produce impacts in scaled-up policies. In the Tennessee STAR experiment, Tennessee allocated additional funds. In scaled-up versions at the state level, the cost is likely prohibitive. States have to reallocate funds toward increasing the number of teachers and away from other inputs. In the case of California, it allocated funds to increase the number of early elementary school teachers, yet the cost of this was a decrease in the funds to hire teachers in other grades, and hence, higher class sizes resulted in those other grades. This could counteract any positive impacts from class size in early grades. Indeed, there is no evidence to date that California’s aggressive class size policy has led to any improvement in outcomes. The literature on class size largely focuses only on class size, but in a scaled-up policy, the improvements from class size must be weighed against the costs of reduced educational inputs elsewhere. As long as costs remain prohibitive, it is unclear whether any state can produce a class size policy that can replicate the gains from the Tennessee STAR experiment.

While *ANAR* did not necessarily take on the issue of class size, its call for local experiments to identify promising solutions resonates with the issue of class size. In considering changes in classroom practice, the debates on class size have led to significant investigations throughout the United States and beyond—not just in primary schooling but also in higher education

(e.g., Bettinger and Long 2018). However, a limitation of experimentation can be its ability to understand how the impacts would change as scaled-up versions of the policy reverberated throughout the education landscape. The formulation of policy around class size has largely proceeded without finding a solution for the costs of reduced class size, and states have sacrificed other inputs in order to accommodate class size. While innovation is present in the case of class size, pushing innovation forward without considering the costs of scaling may never generate the promised impacts.

TIMING OF SCHOOLING

One input that was specifically mentioned in *ANAR* was the length of the school day and year. *ANAR*'s authors lamented that the United States had shorter school days and school years than its competitors. The *ANAR* authors strongly recommended a seven-hour school day and a school year of two hundred to two hundred twenty days.

For at least a decade after *ANAR*, there was very little movement or experimentation with the length of the school day. In 1997, Arizona became the first state to increase the length of the school year, requiring at least two hundred days of instruction rather than one hundred eighty. By 1998, fourteen states were considering changes to the school calendar (National Center on Time and Learning 2017); however, outside of a few districts in Arizona, few changes were happening at scale.

Since 2000, though, there have been significant changes in the time allocated for schooling. Some of these have come in response to the charter school movement. For example, from 2000 to 2012, the average length of the school day nationally increased by 0.2 hours; by contrast, the average length of the school day in charter schools increased by 0.4 hours (Farbman 2015). As Farbman (2015) noted, multiple studies of charter schools and other school turnaround efforts have attributed the impacts of charter schools, in part, to the length of the school day (e.g., Dobbie and Fryer 2013; Fryer and Dobbie 2011).

Additional evidence has come from outside the United States. For example, Germany increased weekly education instruction by two hours, thereby improving outcomes, particularly for high-achieving students (Huebener et al., 2017). Studies in Chile (Bellei 2009), Israel (Lavy 2012), Italy (Battistin and Meroni 2016), Brazil (Rosa et al. 2022), and Latin America more generally (Alfaro et al. 2015) have shown similarly positive impacts of increasing instructional time. These other studies have found greater benefits for both low- and high-performing students.

Within the United States, RttT grants for school improvement often targeted limited experiments in the length of the school day and evidence to date. Some schools implemented changes in the length of the school day in response to these grants. More generally, the largest policy shifts have been in Chicago and Boston. In 2012, Chicago moved from a 5.75-hour school day to a seven-hour school day, and in 2015, Boston Public Schools approved a forty-minute extension of the school day.

While the length of the school day has been the subject of both policy changes and experimentation, there are few studies on lengthening the school year. The average number of school days has shown almost no change nationally and remains around one hundred eighty days.

What does the future hold? The mounting evidence on the impact of increased instructional time will likely increase pressure to consider policy options, particularly for students who are struggling. The continued expansion of charter schools, which have longer school days on average, will continue to put upward pressure on the length of the school day. Not only do they contribute to the increased average school day, but they also put pressure on districts to examine the amount of instruction time they offer. Areas where charter schools provide greater competition to local schools are likely to face greater pressures to increase instruction time. In terms of increasing the average length of the school year, there appears to be little momentum.

OTHER SYSTEMIC SHIFTS

While I have highlighted systematic changes that have focused on improving education quality, there are other systemic shifts that have occurred. Many of these have less to do with school inputs and more to do with the changing context of education. I briefly consider three examples.

LEARNING DISABILITIES

Learning disabilities have become more prevalent over time. Zablotsky et al. (2019), for example, report a significant increase from 2009 to 2017 in the percentage of children diagnosed with any developmental disorder, attention deficit disorder, and autism. Special education enrollment rates continue to rise (Diament 2022). Moreover, the COVID-19 pandemic increased attention on issues of mental health among students.

These changes in health have impacts on classrooms. Students with disabilities have renewed protection and have increased access to accommodations as a result of the Americans with Disabilities Act and the expansion of “504” plans. Students with disabilities are often more expensive to educate, costing as much as thirteen times that of the average student (Griffith 2008), and the increased incidence of documented disabilities puts financial pressures on schools. While schools receive additional funds for students with disabilities, the marginal cost of educating a student with disabilities is likely higher than the increased allotment. Indeed, Bergman and McFarlin (2020) showed that charter schools actively discriminate against students with disabilities in the way that they encourage (or discourage) enrollment. The reason for this discrimination is likely the disparity between the cost and revenue associated with a student with disabilities.

While the increase in disability diagnosis and treatment will certainly improve education quality for those with disabilities, the additional education expenditure required to teach students

with disabilities inevitably leads to reductions in expenditures elsewhere. Increased expansion of charter schools, if indeed charter schools discriminate against students with disabilities, could exacerbate existing inequalities by segregating students by costs. Improvements in our ability to diagnose and treat learning disabilities can reduce the costs of educating students with disabilities and reduce the fiduciary burden.

SCHOOL SAFETY

School shootings have become more commonplace, and a frequent motivation for students to pursue charter or private schools is often school safety. Since *ANAR*, the presence of police, metal detectors, and other security enhancements has shifted the ways schools behave. While the prevalence, particularly in the wake of violent shootings, seems high, in truth there has been a decline in the rate of victimization and threats to teachers from 1994 to 2016 (US Department of Education 2020).

School safety, including policies and procedures to ensure safety, continues to be a hot topic in state and federal legislation. Each school must maintain a plan for ensuring safety and for dealing with school violence.³ As in the case of increased disability rates, a focus on school safety requires resources and attention. Governments have been reluctant to increase funding to fully cover the costs of such expenditures. The resulting policies create more pressure on schools to cut expenditures in other ways. Moreover, to date, there has been little experimentation in ways that can help identify cost-effective strategies for improving school safety. Using schools as laboratories across the United States could provide greater opportunities to learn best practices.

PARENTAL INPUTS

Education scholars have often posited that parental inputs are a significant part of students' academic achievement. While the correlation between parental characteristics is extremely strong, particularly in the case of the mother's education, few papers have established causal relationships between parental inputs and student outcomes. Many localized experiments have attempted to increase parental involvement; however, none of these efforts have scaled in any meaningful way.

There are a couple of notable exceptions in more recent years. While not occurring in the United States, the *Oportunidades* (i.e., *Progresa*) conditional cash transfer program was a major randomized controlled experiment in Mexico that targeted parents and students. Parents received a subsidy conditional on student attendance and student health visits. These programs had a demonstrable impact on student attendance and attainment (see overview in Skoufias 2001). While parents are clearly involved in the treatment, it is not clear if the effects came because of their vigilance or other factors (attendance, health, or increased family income).

Recent randomized controlled experiments have aimed at a more novel approach to encouraging parental involvement. With the expansion of texting capabilities, researchers have used

text messages to try to engage parents. York and Loeb (2014) did this for literacy among low-income parents. Through a series of text messages, York and Loeb coached parents of young children how to teach literacy skills. They found that students arrived at kindergarten with improved literacy as a result of parental engagement.

Bergman (2021) and Bettinger et al. (2022) tested interventions that focused on communicating with parents about students' academic behaviors. Parents received notes about students' truancy and assignment completion. These notes reshaped parents' beliefs and led to improvements in attendance and academic achievement. Bettinger et al. (2022) demonstrate that the saliency of the messages in informing parents of important behaviors that they should be monitoring was likely the mechanism by which this impacted student achievement. In both cases, the cost of the intervention was small relative to the benefits.

New and innovative research designs are extending more and more interventions to parents, and this remains fertile ground for the laboratory of public schools. The recent text message interventions are particularly cost-effective and may have more potential to scale.

CONCLUSION AND DISCUSSION

When I was invited to write this chapter, the charge was to document how education systems changed after *ANAR*. How does one capture forty years of trial and error, of innovation and failure? I have chosen to identify a handful of innovations within districts and within schools. These examples—using small schools, specialized schools, school leadership, and innovation zones as cases of education system innovations and class size and using the time spent in schools as cases of classroom innovation—are just a sampling of innovations that have changed, at least incrementally, the way in which education is delivered. One only needs to browse the research-related web pages of organizations such as the American Institute for Research, MDRC, Mathematica, and RAND Corporation to learn about the breadth of continued experimentation and innovation in schools.

Perhaps the most disappointing aspect of many of these interventions is their failure to be scaled up or to generate impacts when scaled up. The key problem in many cases is the cost of scaling. Oftentimes, scaled versions lack the same features as the original laboratory experiment, and in many cases, funding the scaled version requires sacrifices in other areas. Moreover, in many cases, we lack the supply of personnel or funding to move forward. Perhaps the great challenge of the next forty years will be learning how to create cost-effective versions of the innovations that laboratories produce.

ANAR envisioned an education ecosystem where experimentation and learning from the laboratories of local schooling provided lessons and accelerated the process of change. While one can debate the relative quality of education over time, education systems of experimentation and learning across organizations have greatly increased, especially with the advent of the internet and the role of social media in drawing attention to innovation and to evidence.

The What Works Clearinghouse and other formal and informal collections of evidence on innovative practice and policy only accelerate the role of education institutions as laboratories in identifying promising practices and moving them to scale. However, finding and scaling the products of these laboratories remains the next challenge to be solved if the vision of schools as laboratories is to yield long-run improvements in the quality of education.

HESI PRACTITIONER COUNCIL RESPONSE

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.

What Bettinger shows convincingly is that specific reforms that restructure the arrangement of schooling—its inputs and our fundamental assumptions about the operational model of schools—can meaningfully move the needle in student outcomes, even if we don’t change the fundamental incentives around those outcomes. This is an important thing to remember, because through innovation zones and the like, we can give individual localities the power to make changes to the resources and processes underlying schools, even if they can’t change the system’s priorities and incentives themselves. Would it be better if we changed the underlying incentives, priorities, and value proposition of schooling itself? You bet. In my estimation, but there’s much we can and should be doing in the interim. Freeing up the inputs on the ground to allow educators more autonomy, educating them about what changes in the arrangement of schooling have proved helpful, and holding student outcomes dear can make an important and positive difference.

—Michael Horn, cofounder and distinguished fellow,
Clayton Christensen Institute for Disruptive Innovation

The imagery of education as an ecosystem resonates as we think about the future of education and what it could be. A true learning ecosystem would reflect a more unifying approach when crafting pre-K-12 education policy, adapt more quickly to best educational practices and emerging technologies, be shaped by learners’ experiences, and foster agency among learners in meeting their goals and aspirations. While some of these elements surfaced as a result of *A Nation at Risk*, our continued insistence on traditional learning approaches, limitations in scaling innovative assessment practices, and political polarization in an area that was once unifying preclude us from scaling the innovations we know are successful and that all learners are due.

The adage “when you know better, do better” should drive education policy considerations. A continued focus on what works in education—with a greater emphasis on personalized competency-based learning experiences in which learners are immersed in real-world learning—is the most promising educational initiative underway. Many states are looking at how best to reimagine assessment and accreditation practices toward a more learner-centered approach that

is meaningful to students, teachers, parents, schools, and communities. Successful districts and state agencies have worked collaboratively with their communities and policymakers in building upon many of the successes that came about as a result of *A Nation at Risk*, envisioning a future of learning that moves us closer to creating the learning ecosystem we all desire.

—Dr. M. Jeremy Tucker, superintendent, Liberty Public Schools, Missouri

REFERENCES

- Abdulkadiroğlu, Atila, Weiwei Hu, and Parag A. Pathak. 2013. "Small High Schools and Student Achievement: Lottery-Based Evidence from New York City." National Bureau of Economic Research, NBER Working Paper No. 19576.
- Alfaro, P., D. K. Evans, and P. Holland. 2015. "Extending the School Day in Latin America and the Caribbean." Policy Research Working Paper; No. 7309. World Bank, Washington, DC. <http://hdl.handle.net/10986/22183>.
- Angrist, Joshua D., and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114, no. 2:533–75. <http://www.jstor.org/stable/2587016>.
- Barrow, Lisa, Amy Claessens, and Diane Whitmore Schanzenbach. 2013. "The Impact of Chicago's Small High School Initiative." National Bureau of Economic Research, NBER Working Paper No. 18889.
- Battistin, E., and E. C. Meroni. 2016. "Should We Increase Instruction Time in Low Achieving Schools? Evidence from Southern Italy." *Economics of Education Review* 55:39–56.
- Bellei, Cristián. 2009. "Does Lengthening the School Day Increase Students' Academic Achievement? Results from a Natural Experiment in Chile." *Economics of Education Review* 28, no. 5:629–40.
- Bergman, Peter. 2021. "Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment." *Journal of Political Economy* 129, no. 1:199–222.
- Bergman, Peter, and Isaac McFarlin Jr. 2020. "Education for All: A Nationwide Audit Study of School Choice." National Bureau of Economic Research, NBER Working Paper No. 25396.
- Bergstein, Betsy. 2003. "Stronger, Smaller High Schools to Help Thousands of Oregon Students." Gates Foundation Press Release. <https://www.gatesfoundation.org/ideas/media-center/press-releases/2003/04/oregon-small-schools-initiative>.
- Bettinger, Eric, Nina Cunha, Guilherme Lichand, and Ricardo Madeira. 2020. "When the Effects of Informational Interventions Are Driven by Salience—Evidence from School Parents in Brazil." August 2020. Last revised January 24, 2022. Available via SSRN. <http://dx.doi.org/10.2139/ssrn.3644124>.
- Bettinger, Eric, and Bridget Long. 2018. "Mass Instruction or Higher Learning? The Impact of College Class Size on Student Retention and Graduation." *Education Finance and Policy* 13, no. 1:97–118.
- Bivens, Josh, and Jori Kandra. 2022. "CEO Pay Has Skyrocketed 1,460% Since 1978: CEOs Were Paid 399 Times as Much as a Typical Worker in 2021." Economic Policy Institute, October 4. <https://www.epi.org/publication/ceo-pay-in-2021>.
- Bloom, Howard, Saskia-Levy Thompson, and Rebecca Unterman. 2010. "Transforming the High School Experience: How New York City's New Small Schools Are Boosting Student Achievement and Graduation Rates." MDRC, June 2010.
- Chingos, Matthew M., Katherine M. Lindquist, and Grover J. "Russ" Whitehurst. 2014. "School Superintendents: Vital or Irrelevant?" Brookings Institution, September. <https://www.brookings.edu/articles/school-superintendents-vital-or-irrelevant>.

- Crain, Robert L., Anna Allen, Judith Warren Little, Debora Sullivan, Robert Thaler, Denise Quigley, and Gail Zellman. 1998. "The Effects of Career Magnet Schools." Institute on Education and the Economy; Brief No. 22, December. <https://www.tc.columbia.edu/centers/iee/BRIEFS/Brief22.pdf>.
- Diament, Michelle. 2022. "Across US, Special Education Enrollment on the Rise." *Disability Scoop*, June 24, 2022. <https://www.disabilityscoop.com/2022/06/24/across-us-special-education-enrollment-on-the-rise/29920>.
- Dobbie, Will, and Roland G. Fryer. 2013. "Getting beneath the Veil of Effective Schools: Evidence from New York City." *American Economic Journal: Applied Economics* 5, no. 4: 28–60.
- Farbman, David A. 2015. "The Case for Improving and Expanding Time in School: A Review of Key Research and Practice." Updated and revised February 2015. <https://www.timeandlearning.org/sites/default/files/resources/caseformorelearningtime.pdf>.
- Fryer, R., and W. Dobbie. 2011. "Are High-Quality Schools Enough to Increase Achievement among the Poor? Evidence from the Harlem Children's Zone." *American Economic Journal: Applied Economics* 3, no. 3: 158–87.
- Gamoran, A. 1996. "Student Achievement in Public Magnet, Public Comprehensive, and Private City High Schools." *Educational Evaluation and Policy Analysis* 18, no. 1: 1–18.
- Gates, Bill. Annual Letter 2009. <https://www.gatesfoundation.org/ideas/annual-letters/annual-letter-2009>.
- Gill, B., R. Zimmer, J. B. Christman, and S. Blanc. 2007. "State Takeover, School Restructuring, Private Management, and Student Achievement in Philadelphia." Santa Monica, CA: RAND Corporation.
- Griffith, Michael. 2008. "State Funding Programs for High-Cost Special Education Students." Education Commission of the States: State Notes, May 2008. <https://www.ecs.org/clearinghouse/78/10/7810.pdf>.
- Grissom, Jason A., and Hajime Mitani. 2016. "Salary, Performance, and Superintendent Turnover." *Educational Administration Quarterly* 52, no. 3: 351–91. <https://doi.org/10.1177/0013161x15627677>.
- Hanushek, Eric A. 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24, no. 3:1141–77.
- . 1999. "The Evidence on Class Size." In *Earning and Learning: How Schools Matter*, 131–68. Edited by Susan E. Mayer and Paul E. Peterson. Washington, DC: Brookings Institution.
- Hart, Walter H., Megan Schramm-Possinger, and Sherry Hoyle. 2019. "Superintendent Longevity and Student Achievement in North Carolina Public Schools." Editorial Review Board, 2019.
- Hoxby, Caroline M. 2000. "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." *Quarterly Journal of Economics* 115, no. 4: 1239–85. <http://www.jstor.org/stable/2586924>.
- Huebener, M., A. Kuger, and J. Marcus. 2017. "Increased Instruction Hours and the Widening Gap in Student Performance." *Labour Economics* 47: 15–34. <https://doi.org/10.1016/j.labeco.2017.04.007>.
- Iyengar, Nithin, Kate Lewis-LaMonica, and Mike Perigo. 2017. "School District Innovation Zones: A New Wave of District-Led Efforts to Improve Economic Mobility." Bridgespan Group, October 2017. <https://www.bridgespan.org/getmedia/cf61a0c2-0519-49be-878d-7bd055a1d8ab/school-district-innovation-zones-a-new-wave-of-district-led-efforts-to-improve-economic-mobility.pdf>.
- Jones, F., and A. Chambers. 2021. "Innovation Zones: Policy Flexibility to Reimagine and Modernize K-12 Education Post-COVID-19." Vienna, VA: Aurora Institute Issue Brief, June 2021.
- Kemple, J. J., and J. C. Snipes. 2020. "Career Academies: Impact on Students' Engagement and Performance in High School." New York: MDRC, March 2000. https://www.mdrc.org/sites/default/files/Career_Academies_Impacts_on_Students.pdf.

- Kolodner, Meridith. 2015. "Once Sold as the Solution, Small High Schools Are Now on the Back Burner." *Hechinger Report*, September 29, 2015. <https://hechingerreport.org/once-sold-as-the-solution-small-high-schools-are-now-on-the-back-burner>.
- Krueger, Alan B. 2003. "Economic Considerations and Class Size." *Economic Journal* 113, no. 485: F34–63. <http://www.jstor.org/stable/3590138>.
- Lavy, Victor. 2012. "Expanding School Resources and Increasing Time on Task: Effects of a Policy Experiment in Israel on Student Academic Achievement and Behavior." Working Paper No. w18369. Cambridge, MA: National Bureau of Economic Research.
- Lazear, Edward. 1999. "Educational Production." Working Paper No. 7349. Cambridge, MA: National Bureau of Economic Research.
- Lichand, Guilherme, Eric Bettinger, Nina Cunha, and Ricardo Madeira. 2020. "The Psychological Effects of Poverty on Investments in Children's Human Capital." University of Zurich, Department of Economics, Working Paper No. 349, July 16, 2020. Last revised April 27, 2021. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3633815.
- Macdonald, Mott, and Muriel Dunbar. 2015. "Secondary Vocational Education: International Experience." World Bank Group, Final Report, April 2015. <https://documents1.worldbank.org/curated/en/470241468000264353/pdf/103893-WP-P145371-PUBLIC-Vocational-Education-Report-Final.pdf>.
- McCord, Robert S., and Leslie A. Finnan. 2019. "AASA Superintendent Salary & Benefits Study: Non-Member Version." ERIC, January 1, 2019, <https://eric.ed.gov/?id=ED619865>.
- Mishel, Lawrence, and Richard Rothstein. 2002. *The Class Size Debate*. Washington, DC: Economic Policy Institute, 2002.
- Mosteller, Frederick. 1995. "The Tennessee Study of Class Size in the Early School Grades." *Future of Children* 5, no. 2: 113–27. <https://doi.org/10.2307/1602360>.
- National Center on Time and Learning. 2017. "History of Expanded Learning Time." <https://timeandlearning.org/about/history>.
- NCEE (National Commission on Excellence in Education). 1983. *A Nation at Risk: The Imperative for Educational Reform*. Washington, DC: US Department of Education.
- O'Donnell, Patrick. 2012. "Former Cleveland Schools CEO Barbara Byrd-Bennett Becomes CEO of Chicago Public Schools." *Plain Dealer*, October 12, 2012. https://www.cleveland.com/metro/2012/10/former_cleveland_schools_ceo_b_1.html.
- Ravitch, Diane. 2008. "Bill Gates and His Silver Bullet." *Forbes*, November 19, 2008. https://www.forbes.com/2008/11/18/gates-foundation-schools-oped-cx_dr_1119ravitch.html?sh=79b3c7f44410.
- Rosa, Leonardo, Eric Bettinger, Martin Carnoy, and Pedro Dantas. 2022. "The Effects of Public High School Subsidies on Student Test Scores: The Case of a Full-Day High School in Pernambuco, Brazil." *Economics of Education Review* 87: 102201. <https://doi.org/10.1016/j.econedurev.2021.102201>.
- Ruble, W. 2015. "The Effect of Contracting Out Low Performing Schools on Student Performance." Working Papers 1521, Tulane University, Department of Economics. <https://ideas.repec.org/p/tul/wpaper/1521.html>.
- Schanzenbach, Diane Whitmore. 2014. "Does Class Size Matter?" National Education Policy Center, 2014. <http://www.jstor.org/stable/resrep42079>.
- Schwartz, Amy E., L. Stiefel, and Matthew Wiswall. 2013. "Do Small Schools Improve Performance in a Large, Urban District? Causal Evidence from New York City." *Journal of Urban Economics* 77: 27–40.
- Sims, David P. 2008. "A Strategic Response to Class Size Reduction: Combination Classes and Student Achievement in California." *Journal of Policy Analysis and Management* 27: 457–78.
- Skoufias, Emmanuel. 2001. "Progesa and Its Impacts on the Human Capital and Welfare of Children and Adults in Rural Mexico: A Synthesis of the Results of an Evaluation by IFPRI." Washington, DC: International Food Policy Research Institute, 2001.

- Stiefel, L., A. E. Schwartz, and M. Wiswall. 2015. "Does Small High School Reform Lift Urban Districts? Evidence from New York City." *Educational Researcher* 44, no. 3: 161–72. <https://doi.org/10.3102/0013189X15579187>.
- Toch, Thomas. 2010. "Small Is Still Beautiful." *Washington Monthly*, July/August 2010.
- US Department of Education. 2020. Digest of Educational Statistics (Particularly Tables 228.70 and 228.20).
- Waldrip, Donald. 2023. "A Brief History of Magnet Schools." <https://magnet.edu/brief-history-of-magnets>.
- Wang, Jia, Joan L. Herman, and Daniel Dockterman. 2018. "A Research Synthesis of Magnet School Effect on Student Outcomes: Beyond Descriptive Studies." *Journal of School Choice* 12, no. 2: 157–80. <https://doi.org/10.1080/15582159.2018.1440100>.
- York, Benjamin N., and Susanna Loeb. 2014. "One Step at a Time: The Effects of an Early Literacy Text Messaging Program for Parents of Preschoolers." NBER Working Paper; No. 20659. November 2014. Revised March 2018. <https://www.nber.org/papers/w20659>.
- Zablotsky Benjamin, Lindsey I. Black, Matthew J. Maenner, Laura A. Schieve, Melissa L. Danielson, Rebecca H. Bitsko, Stephen J. Blumberg, Michael D. Kogan, and Coleen A. Boyle. 2019. "Prevalence and Trends of Developmental Disabilities among Children in the United States: 2009–2017." *Pediatrics* 144, no. 4: e20190811.
- Zimmer, R., G.T. Henry, and A. Kho. 2017. "The Effects of School Turnaround in Tennessee's Achievement School District and Innovation Zones." *Educational Evaluation and Policy Analysis* 39, no. 4: 670–96. <https://doi.org/10.3102/0162373717705729>.

NOTES

1. Byrd-Bennett later pled guilty to kickback payments on textbook contracts.
2. While not all innovation zones were started as a means of creating school turnaround, many districts turn to innovation zones as a means of facilitating school turnaround.
3. See, for example, California's requirement for Comprehensive School Safety Plans at California Department of Education, "Comprehensive School Safety Plans," last reviewed Friday, August 19, 2022, <https://www.cde.ca.gov/ls/ss/vp/cssp.asp>.



7. Unfulfilled Promise

The Forty-Year Shift from Print to Digital and Why It Failed to Transform Learning

Tom Vander Ark

Executive Summary

The shift from print to digital is the most significant change in how human beings learn since the printing press. It marked a shift from information scarcity to abundance, a shift from searching to sorting. It changed what, how, and where people learn, and not always for the better. Social media feeds displaced newspapers. YouTube videos became the world's instruction manual. In American schools, the shift to digital learning was gradual, uneven, often chaotic, expensive, and even while there was observed improvement in engagement, largely ineffective at boosting traditional outcomes.

By 2022, American elementary and secondary schools were spending about \$44 billion annually on education technology (edtech) and related training, adding \$750 per student and more than 4 percent to school budgets. Over the forty-year shift, there was limited growth in basic skills (which was all lost during the pandemic) with some observed benefits to learner experience. New schools, particularly those in managed networks, appear to be the best examples of scaled technology-enhanced learning models implemented with fidelity and achieving strong results.

The shift to digital learning was marked by four phases: computers in the back of the room, the introduction of the World Wide Web, the rise of blended learning, and remote learning. The rise of generative AI in 2022 marks a new era of human-computer interaction.

This chapter identifies challenges in the adoption of technology-enhanced learning and makes recommendations for system leaders and policy makers.

- The shift from print to digital remade society, including the educational system.

- Hopes that tech would revolutionize public education have failed to be fulfilled.
- But there are ways to use technology to enhance learning that educators would be wise to adopt.

• • •

FOUR DECADES OF TECHNOLOGY ADOPTION

The historic shift from print to digital learning in US schools happened gradually over four distinct decade-long periods. These were marked by important developments in hardware and software, curricula and assessment, connectivity, public policy, and new learning models.

1983-1993: BACK OF THE CLASSROOM

A Nation at Risk, the 1983 report by the National Commission on Excellence in Education, was a clarion call for education reform and a driver of the standards movement, but it was silent on technology-enhanced learning.¹ Ironically, 1983 was also the year the Apple II catalyzed the first wave of classroom integration and adoption, despite being largely relegated to a supplemental role in the back of classrooms. Word processing and spreadsheets became widely used in business and could occasionally be accessed by secondary students in computer labs. Web browsers accessed static information but gave a few teachers the sense that we were moving from information scarcity to a new age of information abundance.

1994-2004: WELCOME TO THE WEB

The World Wide Web appeared in 1994, giving rise to dynamic, user-generated content and interactive web applications. Schools began developing web pages, educators started blogging, and school systems began communicating with email and web applications.

The US Department of Education established the Office of Educational Technology in 1994 and two years later published the National Educational Technology Plan, which outlined a vision for technology-enhanced learning. Congress created the \$2 billion Technology Literacy Challenge to promote the integration of technology into teaching and learning and the E-Rate program to provide discounted internet access and other telecommunications services.²

The first one-to-one laptop programs started in Australia in the early 1990s and inspired the Microsoft-sponsored Anytime Anywhere Learning initiative, which in 1995 helped ten school districts implement the blended and personalized learning model and inspired Maine (2002) and Michigan (2003) to sponsor one-to-one initiatives.³ The advent of learning management systems and web-accessible curricula, meanwhile, paved the way for statewide online schools including the Internet Academy in Washington State (1996) and the Florida Virtual

School (1997). A few years later, K12 and Connections Academy began supporting virtual schools in most states.

In 1998, responding to the explosion in digital content and communication, Congress updated copyright laws and passed the Children’s Online Privacy Protection Act, requiring websites and online services designed for children younger than thirteen to obtain parental consent.⁴ This was followed by the Children’s Internet Protection Act, which added federal funding for libraries.⁵ The Collaborative for Academic, Social, and Emotional Learning (CASEL) published guidelines for social and emotional learning in 1997.

Two project-based networks launched in California in 2000: High Tech High and New Tech Network. In addition, the NewSchools Venture Fund began accelerating the development of high-quality charter management organizations incorporating early blended learning models. And in 2001, the No Child Left Behind Act (NCLB) expanded federal oversight in public education, including annual testing provisions focused on grade-level proficiency. NCLB had the unintended consequence of reinforcing a time-based system of age cohorts and dampening developments in personalized and competency-based learning.

2005-2015: THE RISE OF BLENDED LEARNING

Expanding access to computers, digital curricula, and broadband connectivity accelerated the adoption of blended learning models that combined face-to-face instruction with online learning to create more flexible and personalized learning. In 2008, Clayton Christensen, Curtis W. Johnson, and Michael B. Horn published the blended learning treatise *Disrupting Class*, and a trio of North Carolina school districts—Henrico, Mooresville, and Charlotte-Mecklenburg—showcased technology-enhanced learning. In 2009, the International Association for K-12 Online Learning (iNACOL) shifted its focus to blended and competency-based learning.

Google Apps for Education launched in 2007, and the free suite of productivity tools quickly displaced several categories of premium products. In 2011, Google introduced Chromebook, an inexpensive, cloud-first mobile computer ideal for secondary school use, and four years later it commanded a 40 percent market share.⁶ In 2014, Google added Classroom, a free assignment manager, and it quickly became the most widely used learning platform. These tools accelerated the adoption of blended learning but dampened private investment in developing premium tools.

In response to the Great Recession (2007-2009), the federal government passed the 2009 American Recovery and Reinvestment Act, which included several education technology programs and more than \$10 billion to increase connectivity and innovation. State adoption of the Common Core State Standards gave rise to the development of two testing consortiums in 2010. High-stakes testing monopolized the next decade of US public education. The focus on grade-level proficiency was so strong that adaptive curriculum providers, which had focused on personalized growth, repackaged their applications into grade-level bundles to improve test scores.

Thirty years after venture investments began powering Silicon Valley innovation, edtech venture investment began expanding with the formation of six impact-oriented edtech venture funds between 2008 and 2015. Notable venture-backed startups included ClassDojo, Clever, Edmodo, Schoology, Kahoot, Duolingo, DreamBox, and EVERFI. In postsecondary learning, 2012 was the year of Massive Open Online Courses (MOOCs) with the launch of Coursera, Udacity, Udemy, and the nonprofit edX.

Blended learning matured in the 2010s with large urban school districts deploying a portfolio strategy and new and improving schools scaling in networks. Led by Dr. Terry Greer, Houston's PowerUp initiative was a frequently cited example of coordinated deployment of blended learning models, connectivity, hardware, and professional learning. In Denver, then superintendent Tom Boasberg scaled innovation networks and charter networks.

2016-2023: REMOTE LEARNING AND BEYOND

The modern whole-child movement was codified by the Science of Learning and Development (SoLD) Alliance beginning in 2016 with an outcome framework and design principles. Battelle for Kids began encouraging school districts to create a Portrait of a Graduate, embracing broader learning goals.

Private equity became a visible force in edtech in 2015 when Vista Equity acquired PowerSchool. Hundreds of edtech startups were acquired in the next five years, sometimes leading to integration but also encouraging more startups. By 2023, school districts were accessing almost twenty-six hundred different edtech applications.⁷ In the school-adjacent space, Microsoft launched Minecraft Education, and FIRST Robotics, VEX Robotics, LEGO, and Tynker all scaled.

Around the edges of the economy and education, Web 3.0 gave rise to distributed ledger technology including blockchain and decentralized autonomous organizations (DAOs) and token-based economics. Learning and Employment Records (LERs) are the only distributed ledger technology with traction and the likelihood of scale in education.

By 2019, connectivity advocates declared victory as nearly every school in America had broadband access, and there was close to one device for every learner.⁸ However, a year later, the quick shift to remote learning during the COVID-19 pandemic laid bare that at least 15 percent of US households with school-age children did not have a high-speed internet connection at home.⁹ The pandemic provoked several rounds of massive federal investment including roughly \$150 billion for supporting safe distance learning and connectivity.

The sudden shift to remote learning that occurred in most US school systems in 2020 illustrated that only a small percentage of public systems had well-developed digital learning models with one-to-one take-home technology. Those systems flipped quickly and efficiently to remote learning and experienced limited learning loss. The majority of school systems struggled to equip learners with devices and digital-first instruction. Remote learning also brought on a decline in the mental health of students as a result of COVID restrictions,

addictive use of social media, pressures to perform on traditional measures at school, as well as a context of violence and climate crisis.

As bad as the pandemic was for teaching, learning, and mental health, remote delivery may have been a tipping point for anywhere, anytime learning and teaching in teams on common platforms with shared resources and strategies. The return to physical school buildings in the 2022–23 school year may also have been a tipping point for a whole-learner focus with broader measures of success and stronger systems of support.

Most recently, in the fall of 2022, free generative artificial intelligence (AI) applications were released to the public. Global use exploded, as did speculation about implications for work and learning. High school and college students quickly began using AI chatbots to write essays and get help with homework. Teachers went back to school in the fall of 2023 with many unanswered questions about how and when generative AI would be used and how it would change expectations, assignments, and assessments.

LIMITED MEASURED BENEFIT FROM TECHNOLOGY ADOPTION

The forty-year shift from print to digital learning was gradual, chaotic, expensive, and largely ineffective at transforming learning experiences and outcomes. In most cases, computers were added slowly and opportunistically to schools with a mixture of teacher adoption and phased systemwide initiatives. Computers were purchased with surplus funds, grants, and periodic bursts of public funding, often without a plan or professional development. There was tremendous variation in the use of technology, even within schools, with few models of transformed learning at scale and a great deal of digitized twentieth-century pedagogy. When the pandemic forced systems to switch to digital learning, it was an unmitigated academic disaster for all but a few well-prepared systems.

Despite this, total K-12 edtech spending in the United States is about \$40 billion per year.¹⁰ Technology-related professional learning adds approximately \$4 billion more annually. About \$38 billion of that is spent by public schools, with a breakdown of roughly 4.4 percent of total expenditures (capital and operating) and about \$750 per student.

With nearly every school in the country connected to the internet and 94 percent reporting that they provide a laptop or tablet for learners, almost every school has some degree of blended learning.¹¹ Despite the near ubiquity of technology, there has been limited aggregate improvement in basic skills over the past thirty years, and most of the incremental gains were lost during technology-based remote learning.¹²

An evidence review of technology-enhanced learning found that technology largely failed to improve learning outcomes with the exception of some positive results for computer-aided learning in math, including trials with ASSISTments, Cognitive Tutor, and DreamBox Learning.¹³ Most elementary schools have subscriptions to personalized reading and math software, and

all of the vendors tout research backing and evidence of results. However, after decades of investment, the lack of improved reading and math scores at scale is complex (as discussed below), but the core problem appears to be a lack of use. A pre-pandemic survey showed that of the annual \$12 billion investment in learning software, only 16 percent of the licenses had high-fidelity use. Another study showed that 67 percent of software products went unused.¹⁴

In addition to boosting topline performance, technology adoption in most sectors aims to improve labor productivity. While blended learning models have the potential to improve productivity, labor savings were seldom a stated goal in education. In fact, student-to-teacher ratios have declined since 1990, adding the equivalent of four teachers to a typical elementary school.¹⁵ Staffing increases were likely not caused by the adoption of technology, but technology has added at least 5 percent to the cost model with no observed improvement in labor productivity.

WIDELY APPRECIATED BENEFITS

While technology has not transformed most systems, educators have come to believe that digital learning tools are integral to teaching and learning. A pre-pandemic Gallup study showed at least eight in ten teachers and nine in ten administrators strongly agree or agree that they see great value in using digital learning tools in the classroom.¹⁶ Despite some signs of a post-pandemic technology backlash, a 2023 study showed the same level of educator support for edtech.¹⁷

There are at least six areas of widely appreciated benefit of technology-enhanced learning:

- **Engagement:** Students may be more engaged and motivated when learning with technology.¹⁸ Opportunities include learning games, simulations, virtual reality, and maker tools.
- **Personalized learning:** Edtech can provide personalized learning experiences that give students voice and choice and allow students to control the pace and place of their learning by supporting both synchronous and asynchronous learning experiences.¹⁹
- **Project-based learning:** When students explore real-world problems through individual and group projects, it allows them to make sense of why content is useful and how it might be applied. Rigorous project-based learning has a strong effect on student achievement.²⁰ Inquiry and project-based learning in higher education is often called “active learning,” a field of study with more than twenty years of supporting research.²¹
- **Employment skills:** Technology-enhanced learning can develop job skills including spatial reasoning, abstraction, data visualization, content creation, and project management. It is central to delivering computer science and information technology

pathways including software development, cloud computing, network engineering, and machine learning.

- **Pathway access:** Online learning extends career exploration opportunities and the availability of career-related courses and college credit courses.
- **Persistence:** The combination of more engaging learning experiences, more learning options, and data-informed support systems appears to have contributed to improved persistence and graduation rates. These have improved by 20 percent in the past twenty years.

NETWORKS OF NEW SCHOOLS SHOW LEARNING AND STAFFING PRODUCTIVITY

In 2010, Project RED observed that “effective technology implementation in schools is complex, with hundreds of interrelated factors playing a part.” The study concluded that blended learning works best when each student has a personal portable device and when teachers base instructional practice on digital resources. They identified nine implementation factors and said system and school leadership were key to developing a strong plan with aligned components. With strong implementation, Project RED observed improvement in eleven success measures.²²

The most promising developments in technology-enhanced learning are networks of schools that were developed as blended learning models with shared platform resources, a mobile device for each student, and high-fidelity deployment. The largest and best examples of high-quality school development are charter management organizations including Alpha, Aspire, ASU Prep, Green Dot, Harmony, IDEA, KIPP, Summit, and Uplift. They have been recognized for their academic performance and operate with less funding than traditional public schools.²³

New York City is another example of high-quality new-school development. Between 2002 and 2008, the school district closed twenty-three large failing high schools where fewer than half the students graduated and opened more than two hundred new small schools in networks including Urban Assembly, Outward Bound, New Visions, and Internationals. The new schools featured personalized and project-based learning and showed dramatically higher achievement levels and graduation rates than both the schools they replaced and the schools serving comparable populations.

During the pandemic, a dozen or so quality online schools served full- and part-time learners, often in partnership with local school districts, with better-than-state-average achievement levels. These online schools include Florida Virtual, ASU Prep Digital and Khan World School in Arizona, GEM Prep in Idaho, and VLACS in New Hampshire. New schools create an

opportunity for the coherence of goals, learning models, tools and materials, staffing and professional learning, schedule, structure, and systems.

WHY TECHNOLOGY ADOPTION WAS CHAOTIC AND INEFFECTIVE

Education is not the only sector that struggles with effective technology adoption. Global business investment in information technology is nearly \$5 trillion annually. An EY study showed that three out of five companies do not know how much they spend on technology or what value it yields. To address the problem, the majority of business leaders are centralizing their investment in technology and are adding programs to measure outcomes.²⁴ An Accenture analysis concluded that the majority of businesses do not see a return on their technology investments, and just 14 percent of businesses achieve the intended impact of their investments. Businesses that were most successful with their investments were the ones investing in bold moves rather than incremental shifts.²⁵

Breakthrough results are achieved in business by new organizations formed around new technologies (similar to the new-school opportunity in education) and by organizations transformed by best practices from what EY called digital leaders, who do the following:²⁶

- Establish a strong digital foundation and culture of innovation first
- Focus digital investments on innovation and new products and services
- Align technology deployment, process innovation, and skill building
- Carefully allocate spending tied to milestone achievements

MORE OPPORTUNISTIC THAN COHERENT CAPITAL DEPLOYMENT

Unlike in every other developed nation, public education in the United States is primarily a local affair governed by more than thirteen thousand elected school boards and about five thousand nonprofit boards. Local funding is augmented by a maze of state and federal programs that come with compliance requirements.

Effective technology adoption requires the development of new learning models and coordinated investment in hardware, software, and professional development; however, public schools have limited ability to manage capital expenditures. School districts have taxing authority and may be able to pass a bond for new or remodeled facilities, but (with the rare exception of a technology levy) they rely on surplus funds, special programs, or grants to purchase assets like computers and network equipment. This, along with revolving-door leadership, makes it challenging to plan and execute a multiyear transformation investment strategy.

In addition to local control, education has historically been pedagogically decentralized. In the 1980s and 1990s, most teachers had some degree of pedagogical autonomy. As a result, the first two decades of technology adoption followed consumer rather than enterprise adoption patterns. The rise of blended learning (2005–2015) came with the growth of school networks and leading school districts attempting coordinated systemwide change initiatives. The pandemic flipped the remaining systems to an enterprise approach with common resources on a shared platform and more teaching in teams.

INEFFICIENT EDTECH MARKET

A highly decentralized education system with sporadic funding and a chaotic combination of teacher adoption and systemwide initiatives contributed to weak demand signaling and an inefficient market. Compared to other large sectors of the economy, education research and development has had limited public and private investment. Before the Great Recession, there was almost no venture capital investment in education. It wasn't until 2015 that six active impact-focused edtech funds led to the investment of \$3 billion in startups, which is roughly equivalent to the first banner year of technology venture funding in 1978 that launched the information age.²⁷

When Google (and later Microsoft) began offering free productivity tools and then a free learning platform, it displaced several existing applications and slowed investment in new platforms. The availability of free tools and weak demand signaling are two of the reasons the sector lacks sophisticated personalized learning platforms. Another reason the market is inefficient is that research is not yet a strong signal to venture investment. Studies on edtech and technology-enhanced learning are in the early stages of building evidence, and most studies are underpowered and not replicated.

Lastly, the marketplace struggles due to limited interoperability and edtech business models that are based on ownership of item-level data. When businesses own student data, they have control over how it is used and shared. This leads to a lack of transparency and accountability, creates conflicts of interest, and dampens innovation. Interoperability and data transferability will be key components of a future fueled by lifelong learning and portable learning records.

LIMITED CHANGE MANAGEMENT CAPACITY

Technology-enhanced learning has added to the complexity of developing and operating modern school models. Small school districts cannot afford or attract the talent necessary to develop or adopt a K-12 digital curriculum, technology stack, and associated professional learning experiences. There are about nine thousand school districts with fewer than twenty-five hundred students, which is too few to manage a technology infrastructure. Solving this will require consolidation or working in shared service networks.

Before 2002, there was little consulting capacity in the sector to support the development of new models or to guide system transformation efforts. By 2006, with philanthropic support, there were half a dozen national consultants, a dozen regional intermediaries, and two dozen scaled, managed, and voluntary school networks that contributed to change capacity.

THE NEW ERA OF AI-ENHANCED LEARNING

While machine learning applications were ubiquitous in corporate computing by 2017, it was the introduction of generative AI tools to the consumer market in the fall of 2022 that marked the beginning of a new era of human-computer interaction, an era that will change how we work and learn and more specifically how we create and share content. It is hard to know what the next forty years will bring, but edtech developments of the next four years should be anticipated and influenced for equity.

As students went back to school in the fall of 2023, generative AI applications (e.g., ChatGPT from Open AI, Bing from Microsoft, Bard from Google, Pi from Inflection AI) were widely available, but few schools had a plan in place for how to use them. Teachers should now assume that content creation (e.g., writing, coding, visual art) and problem-solving outside of class will include AI collaboration. This won't always be the case, as some students will have moral concerns or fears of getting caught, will lack access, or will just find it unhelpful for an assignment. Teachers should ask for disclosure of use and inform students of limitations. The AI detection tools are not worth using.

Built on top of the large language models that power generative AI, personal learning assistants like Khanmigo from Khan Academy provide smart tutoring and pathway guidance. Learning assistants like Project Leo from DaVinci Schools help students build inspired projects while gaining feedback from teachers, professionals, and peers throughout the process. AI-powered teaching and learning assistants will be widely adopted by 2025.

AI changes what we can expect young people to do, the quality of work they produce, and the value they can create. It further reduces the value of hand calculation and increases the need for computational thinking. Conrad Wolfram argues that computational thinking is required in all fields and in everyday living and suggests that “[this approach] is built on actual problems solved by real people in the real world with today’s technology. A computer-based maths curriculum should be built around real-world requirements such as data science, information theory, and modeling.”²⁸

The rise of AI-generated content requires new levels of digital discernment and civic literacy, stressing curation, application, and creation skills. Navigating this flood of synthetic visual and audio content will require applied ethics and a commitment to shared values at the level of the community and the planet. Decisions will become less clear, more frequent, and more impactful.

In addition to changing learning goals and experiences, AI applications will inform student services, improve resource allocation, and benefit talent acquisition and development. AI applications will help make facilities and transportation systems more efficient. They may (as we once believed social media would) improve community learning and help facilitate civic and education agreements. Exponential technologies will also, however, inevitably exacerbate income and wealth inequality.

CREDENTIALS AND LEARNER RECORDS

The post-pandemic labor market is increasingly focused on skills. These are best communicated via verifiable digital credentials. Eight in ten human resources professionals affirm that skill assessments hold equal or greater significance than traditional factors such as degrees and years of experience in their hiring decisions.²⁹ By 2025, workplace credentialing systems will add durable skills like critical thinking, communication, collaboration, and creativity as well as character skills like fortitude, growth mindset, and leadership.³⁰

The three million high school graduates in 2023 earned about two million credentials of varying value. The growth in skills-based hiring and the rise of skills credentialing will result in more high schools, colleges, and workforce preparation programs issuing credentials and supporting the credentialing process for durable skills as well as technical skills. In 2024, hundreds of high schools will begin to credential valuable experiences such as client projects and internships. In the second half of this decade, these new forms of evidence of demonstrated capabilities will begin to replace courses and grades as the primary signaling mechanism for talent transactions, and they will be securely shared via LERs.

In 2022, North Dakota was the first state to adopt a blockchain-based digital credential wallet and provide high school and college students ownership of their records. Although some higher education institutions are adopting LERs, they will only reach their full potential if states make them a public utility. This would allow lifelong learners to capture and communicate capabilities developed within and outside of school and to access offers of scholarship and employment more fully and efficiently.

NEW SCHOOLS

While the largest public school systems continue to decline in enrollment and close schools, more than one thousand new schools will be created by 2027. More than half will be charter schools, and more than half of those will be microschoools. More than half will incorporate aspects of new learning models that combine AI-powered personalized skill building and project-based learning. Of the high schools, more than half will focus on new career pathways and will include work-based learning, entrepreneurial experiences, and accelerated credentials achievement. More than half of the high school learners will unlock opportunity with a curated learner record.

RECOMMENDATIONS

Technology-enhanced learning largely failed to live up to its perceived promise as a result of chaotic decentralized adoption, uneven use, and weak teacher supports. Following are five recommendations for system leaders and policymakers to address these challenges and take advantage of the new era of AI-enhanced learning.

1. Encourage school networks. Encourage schools to work together in networks or systems that share a coherent learning model including shared outcomes, experiences, assessments, supports, and professional learning opportunities for educators. Expand high-quality managed networks. Encourage district schools to join voluntary networks like New Tech Network, NAF, Linked Learning, Summit Learning, and CAPS Network. Similarly, schools can join curriculum networks that provide some of the benefits of a whole-school model like EL Education and Project Lead the Way. Small districts can work together in networks like Collegiate Education in Texas and SparkNC in North Carolina. Small districts could also consolidate to create enterprise scale.

2. Leverage formative assessment. Streamline performance monitoring and reduce end-of-year testing by aggregating samples of formative assessment data at the state level. This would require schools to belong to an authorized network or district with an assessment system that had demonstrated accurate and consistent tracking of growth as well as proficiency. This lightweight progress monitoring solution could be extended to schools serving students funded through vouchers and education savings accounts.

3. Invest in R&D. Encourage disclosed use of generative AI applications in project authoring and content creation. Pilot AI learning assistants and support systems. Sponsor the formation of AI-forward school networks.

4. Update public learning infrastructure. Ensure that all learners and workers have access to broadband connectivity, learning devices, and statewide learning opportunities. Provision a Learning and Employment Record to ensure learners and workers a lifetime of portability, security, interoperability, and opportunity. School systems, with state and federal assistance, should enact interoperability and privacy standards in vendor relationships. Learner data should be portable within and between education systems.

5. Update learning goals. Host a regional conversation about what young people can do with exponential technology. Adopt learning goals that include creativity, collaboration, critical thinking, and entrepreneurship. Help learners identify strengths, interests, and community needs and build pathways to contribution through a series of learning experiences that are engaging and intentional, community connected, supported, and accelerated.

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.

To realize the promise of educational technology in transforming learning, it is imperative that policymakers and practitioners take a systems-change approach to leading in the future. By emphasizing purpose, creating coherence with goals and outcomes, and producing authentic evidence of learning, technology can support new learning designs. The key is to start with the broader ends in mind and backward-design toward future learning goals—not just layer technology on old reforms of the past forty years or on traditional models of ranking and sorting students.

It's critical to first ask the question: What is the appropriate role of technology in redesigning K-12 education? What resonates is being clearer about how technology meets the goals and aims, identifying what professional development is needed and what products and tools are required to meet the needs, and having a strong focus on instruction with contemporary pedagogy, authentic assessment, and curriculum redesign, all aligned to equip students for building the knowledge and skills to be prepared for lifelong learning.

As the educational landscape continues to evolve, research remains a touchstone for progress to unlock the full potential of educational technology. For policymakers and practitioners, modernizing research and development (R&D) is an important next step. Reorganization and investments are required to build a more robust federal, state, and local R&D infrastructure. A collective call to action for reorganizing the Institute of Education Sciences to future-proof education is needed in a rapidly changing, technology-driven world.

—Susan Patrick, independent consultant and advisor, and former president and CEO of the Aurora Institute

NOTES

1. National Commission on Excellence in Education (NCEE), *A Nation at Risk: The Imperative for Educational Reform* (Washington DC: US Department of Education, 1983).
2. The Technology Literacy Challenge is outlined in an archived White House document, "Educational Technology: Making It Happen," February 15, 1996, <https://clintonwhitehouse4.archives.gov/WH/New/edtech/2pager.html>.
3. S. Mahoney, "Laptops Win Over the Skeptics, Even in Maine," *New York Times*, March 5, 2003, <https://www.nytimes.com/2003/03/05/nyregion/laptops-win-over-the-skeptics-even-in-maine.html>.
4. Children's Online Privacy Protection Act of 1998, S.2326, 105th Congress (1998), <https://www.congress.gov/bill/105th-congress/senate-bill/2326>.

5. Children’s Internet Protection Act, H.R.4600, 106th Congress (2000), <https://www.congress.gov/bill/106th-congress/house-bill/4600>.
6. Anick Jesdanun, “How Google Chromebooks Conquered Schools,” Associated Press, February 10, 2017, <https://apnews.com/general-news-41817339703440a49d8916c0f67d28a6>.
7. Instructure, “The EdTech Top 40: K-12 EdTech Engagement,” June 2023, <https://www.instructure.com/edtech-top40?>
8. Emily Tate Sullivan, “Mission (Almost) Accomplished: Nonprofit EducationSuperHighway Prepares to Sunset,” EdSurge, April 4, 2019, <https://www.edsurge.com/news/2019-04-04-mission-almost-accomplished-nonprofit-educationsuperhighway-prepares-to-sunset>.
9. Brooke Auxier and Monica Anderson, “As Schools Close due to the Coronavirus, Some US Students Face a Digital ‘Homework Gap,’” Pew Research Center, March 16, 2020, <https://www.pewresearch.org/short-reads/2020/03/16/as-schools-close-due-to-the-coronavirus-some-u-s-students-face-a-digital-homework-gap>.
10. Instructure, “EdTech Top 40.”
11. Kristal Kuykendall, “New Data: Nearly Half of Schools Providing Home Internet Access to Students Who Need It This School Year,” THE Journal, September 27, 2022, <https://thejournal.com/articles/2022/09/27/nearly-half-of-schools-providing-home-internet-access-to-students-who-need-it-this-school-year.aspx#>.
12. NAEP Long-Term Trend Assessment Results: Reading and Mathematics, accessed October 9, 2023, <https://www.nationsreportcard.gov/highlights/ltr/2022>.
13. Maya Escueta et al., “Education Technology: An Evidence-Based Review,” National Bureau of Economic Research, accessed August 2017, <https://www.nber.org/papers/w23744>.
14. Michele Molnar, “More Than \$1 Billion in K-12 Ed-Tech Licensing Fees Go to Waste,” Education Week, EdWeek Market Brief, November 19, 2019, <https://marketbrief.edweek.org/marketplace-k-12/latest-ed-tech-usage-report-shows-improvements-still-1-billion-waste-k-12>.
15. National Center for Educational Statistics, “Public Elementary and Secondary Pupil/Teacher Ratios, by Selected School Characteristics: Selected Years, Fall 1990 through Fall 2017,” accessed October 9, 2023, https://nces.ed.gov/programs/digest/d19/tables/dt19_208.10.asp.
16. Gallup, “Educators Agree on the Value of Ed Tech,” September 12, 2019, <https://www.gallup.com/education/266564/educators-agree-value-tech.aspx>.
17. Classroom of the Future Report 2023, Clever, August 15, 2023, https://www.clever.com/wp-content/uploads/2023/08/CLEVER_CLASSROOM_OF_THE_FUTURE_REPORT_2023.pdf.
18. Niaz Makhdam Muhammad, “The Role of EdTech in Enhancing Learners’ Motivation,” *Selangor Science & Technology Review* 5, no. 5 (2021).
19. Kris Astle, “EdTech and the Classroom: How EdTech Has Transformed the Way Teachers Teach and Students Learn,” Learning Council, March 20, 2023, <https://thelearningcounsel.com/articles/edtech-and-the-classroom-how-edtech-has-transformed-the-way-teachers-teach-and-students-learn>.
20. Kristin De Vivo, “A New Research Base for Rigorous Project-Based Learning,” Kappan, January 24, 2022, <https://kappanonline.org/research-project-based-learning-de-vivo>.
21. Cynthia J. Brame (n.d.), “Active Learning,” Ohio State University, <https://www.oaa.osu.edu/sites/default/files/uploads/nfo/2019/Active-Learning-article.pdf>.
22. The Greaves Group, The Hayes Connection, and One-to-One Institute, *The Technology Factor: Nine Keys to Student Achievement and Cost-Effectiveness* (Jacksonville, FL: MDR, 2010), <https://www.k12blueprint.com/sites/default/files/Project-RED-Technology-Factor.pdf>.
23. Margaret E. Raymond, James L. Woodworth, Won Fy Lee, and Sally Bachofer, “As a Matter of Fact: The National Charter School Study III 2023,” CREDO, June 19, 2023, <https://ncss3.stanford.edu/wp-content/uploads/2023/06/Credo-NCSS3-Report.pdf>.

24. Sri Prabhakaran, Laura McGarrity, and Ankur Sharma, "The Digital Investment Index Reveals Companies Are Doubling Down on Tech Investments but Struggle to Clearly Define Their Digital Strategy, EY-Parthenon," April 5, 2022, https://www.ey.com/en_us/strategy/digital-investment-report#3.
25. Neil Franklin, "Majority of Businesses Fail to See a Return on Their Technology Investments," Workplace Insight, February 6, 2019, <https://workplaceinsight.net/majority-of-businesses-dont-see-a-return-on-technology-investments>.
26. Prabhakaran, McGarrity, and Sharma, "Digital Investment Index."
27. CB Insights, "Mega-Rounds Boost Global Ed Tech Funding to New Record," January 19, 2016, <https://www.cbinsights.com/research/2015-global-ed-tech-funding>.
28. Conrad Wolfram, *The Math(s) Fix: An Education Blueprint for the AI Age* (Champaign, IL: Wolfram Media, 2020).
29. SHRM press release, "New SHRM Research Makes the Case for Skills-Based Hiring," August 15, 2022, <https://www.shrm.org/about-shrm/press-room/press-releases/pages/new-shrm-research-makes-the-case-for-skills-based-hiring.aspx>.
30. America Succeeds, "Durable Skills," accessed October 9, 2023, <https://durablekills.org/about>.



8. Fixing Schools through Finance

Eric A. Hanushek

Executive Summary

Spending on schools is frequently used as an indicator of the quality of schools. In discussions of how to make schools better and more equitable, the first argument one hears is the necessity of increasing our spending levels. Unfortunately, history has not been very supportive of this strategy. Research suggests that adding resources to schools is likely to have a positive effect on student outcomes, but the estimated impact of resources is highly variable and depends on the context and constraints on the spending.

In simplest terms, it appears that *how* funds are spent is crucial—and generally more important than *how much* is spent. Divorcing decisions on “how much” from “how” has not been successful within the current structure of school decision-making.

The trouble from an overall policy perspective is that it has not been possible to describe when funds will be particularly effective or ineffective. To date, little headway has been made in describing the features of the particular contexts or the particular uses of funds that yield significant learning gains. Thus, policy solutions based on prescribing specific programs or spending patterns generally lack a scientific foundation.

An effective funding policy would set up incentives so that the decision-makers take actions that lead to better student outcomes. The central elements of such a system, building on what has previously been successful, include a strong accountability system with incentives and direct rewards for successful performance, empowered local decision-making by both schools and parents, and an ongoing information and evaluation system. This approach capitalizes on local knowledge of both educational demands and capacities.

The key idea is recognizing and rewarding success. If the objective is improving outcomes, the system should focus on outcomes and should reward those who contribute to success—that is, those who bring about high achievement.

- How funds are spent is even more important than how much money is spent when it comes to improving school quality.

- Funding policies should include incentives that lead to academic success.
- The key is providing schools and parents with a strong accountability system with a scientific foundation.

• • •

Strangely, the subject of revenues and expenditures is never addressed in *A Nation at Risk* (ANAR). That omission makes the cascade of calls to increase funding for schools, often justified by reference to the message of urgency in ANAR while disregarding use of funds, ironic. By ignoring the role of funding and budgeting, the recommendations from the US National Commission on Excellence in Education are untethered from any grounding in choices and trade-offs that all public policy required. On the other side, the calls for funding that are divorced from ideas of how the funds are to be used are equally problematic.

Spending on schools is frequently used as a summary statistic of the quality of schools. And in discussions of how to make schools better and more equitable, the first order of business is frequently the necessity of increasing our investment in schools—in other words, our spending on schools. Unfortunately, history has not been very supportive of this strategy.

A conventional perspective is that legislatures and school districts should decide how much to spend based on the trade-off between the expected benefits of school spending and the taxes required for any given revenue. Once the revenue is determined, school districts would make budget allocations in order to produce the best student outcomes.

However, this picture is complicated in the case of schools, since states—which have primary responsibility for schools—are concerned not only with overall student outcomes but also with the equity of public provision. Two factors enter into the equity discussions. First, education is not entirely a function of schools but has components of families and circumstances that enter into student outcomes. Thus, children from more educationally disadvantaged households, English language learners, and children with various handicaps need more from the schools if they are to pull even with students not facing such difficulties. Second, because local funding is heavily reliant on local property taxes, the size of the district tax base will directly influence the ability of a school district to raise revenues. Students who happen to reside in districts where the value of residential property and the presence of commercial and industrial properties are high have an advantage in raising revenues for their schools.

The legislature in each state is charged with making political decisions about both the level of spending and how statewide education and funding differences are addressed. How to reach decisions that weigh the underlying trade-offs is vigorously debated, and every state has its own solution to this.

Legislatures are not the only actors in these discussions. Various parties who have not liked the legislative outcomes have gone to the courts to try to change the legislative decisions. Starting in California in 1968, courts in all but two states (Hawaii and Utah) have had litigation

about school funding. The early cases in the 1970s and 1980s focused on equity issues largely related to differences in property tax bases and spending differentials across school districts, but the cases evolved throughout subsequent decades to focus more on the overall adequacy of funding to meet educational objectives.

The largest difficulty with the pattern and outcomes of revenue decisions is, however, the lack of a clear relationship between added spending and student outcomes. In simplest terms, the division of decisions between how much to spend and how to spend it has historically led to highly variable and quite disappointing results in terms of student outcomes. Specifically, it appears that how funds are spent is crucial—and generally more important than how much is spent. This does not say that more resources are never important for student outcomes. Nor does it say that more resources cannot be important for improved student outcomes. It does say that divorcing decisions on “how much” from “how” has not been successful within the current structure of school decision-making.

This chapter documents these overall conclusions. It then discusses alternative perspectives on funding for schools.

A SHORT HISTORY OF FUNDING

In order to frame the school finance discussion, it is helpful to describe briefly the nature of financing of schools in the United States.¹ The overall picture of enrollments, structure of the schools, and funding shows significant changes over time. Further, the aggregate picture hides significant variation across the states. The variety provides an important backdrop both for the analysis of school finance issues and for decision-making in the schools.

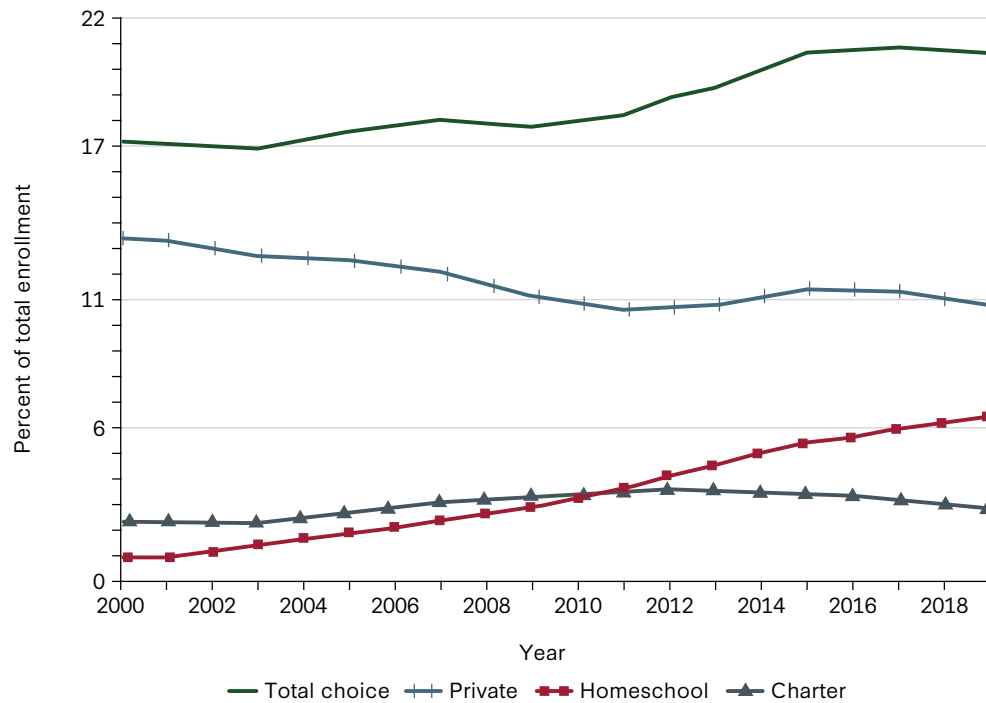
AN OVERVIEW OF US SCHOOLING

Public school enrollment in the United States, while rising during the 1990s, reached fifty million students in 2013 and stabilized there until the COVID-19 pandemic hit in 2020.² The full impact of the pandemic is not yet known, but public school enrollment fell by 3 percent from fall 2020 to fall 2021 and remained at the lower level through fall 2022.

These students are spread very unevenly across states and, within states, across separate local school districts. At the state level, Vermont had a total of 82,000 students while California had six million. The prime operating level is the school district, of which there were 13,452 in 2019, down from 117,408 in 1940. Moreover, the states are broken up into widely varying numbers of local districts. While Hawaii and the District of Columbia each have only one school district, five states had more than one thousand districts.

But even these aggregate variations understate the degree of heterogeneity in the schools, because the growing importance of school choice leads to even more decentralized operation of education. The public school district is the prime operating unit, but it does not cover the full

FIGURE 1 Enrollment patterns of US children, 2000-19



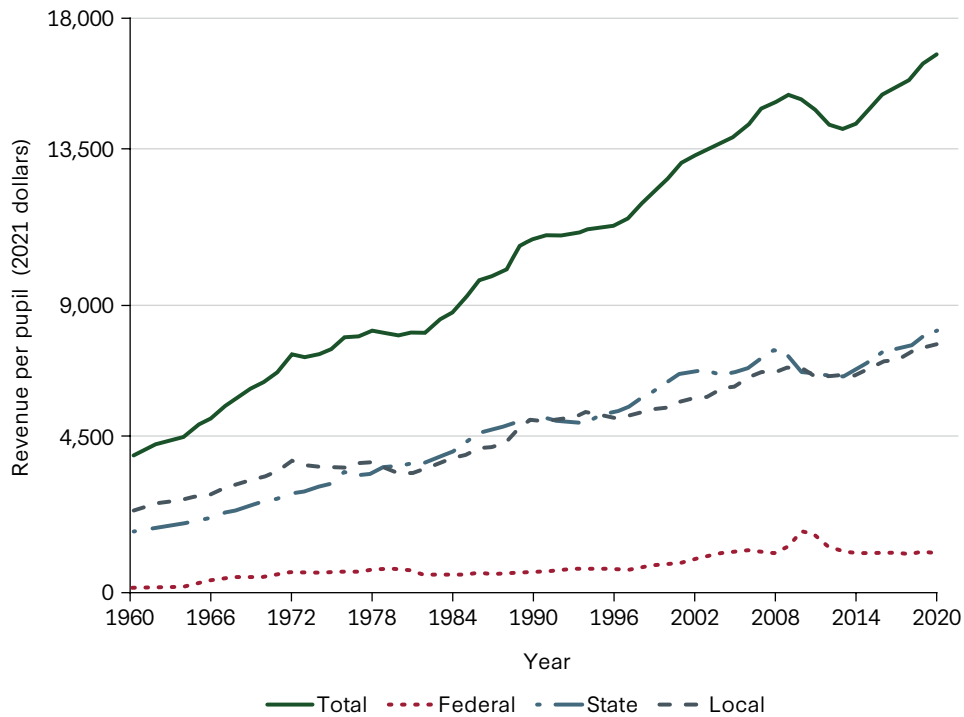
Source: Author's calculations using National Center for Education Statistics (NCES), Digest of Education Statistics 2021, https://nces.ed.gov/programs/digest/2021menu_tables.asp (tables 205.10, 206.10, 203.20, and 216.20); and 2014, https://nces.ed.gov/programs/digest/2014menu_tables.asp (Table 216.20) (Washington, DC: National Center for Education).

provision of education services. Charter schools were first established in Minnesota in 1991, and the model spread across the country. Charter schools are public schools that operate with varying degrees of autonomy, depending on the state. Typically, they are free to operate outside of many of the education regulations in a state, and importantly, they can set their own requirements for teacher preparation, salary schedules, and personnel rules independent of local teachers' unions. They receive public funding, and they are almost always required to take all applying students or to randomize admissions if more students apply than they can accommodate. They are also required to participate in the state student assessment systems.

Students can also attend private schools or be homeschooled. While this is changing, private schools almost always receive no direct public funding, as is the case for homeschooling. These parts of the system are generally very unregulated, and they can set their own curricula, standards, and hiring rules. They generally do not participate in state student assessment systems, and little is known about their performance except as indicated by parental choices.

Figure 1 shows the substantial changes in the structure of US schools in the twenty-first century in terms of parental choices that interact with school finance.³ There has been a steady rise in charter school attendance with relatively stable homeschool attendance and some decline in private schooling. The private school attendance is one-quarter nonsectarian and three-quarters religious based, with the religious component evenly split between Catholic and other denominations.

FIGURE 2 Per-pupil revenue of US public schools, by source, 1960-2019



Source: NCES, Digest of Education Statistics 2021, table 235.10.

Note, however, that these data are pre-pandemic. With the pandemic, traditional public school attendance fell, while the other choice options increased. Within the public school sector there was also a shift from the traditional public schools to charter schools. The long-run distribution of students remains unclear at this time.

REVENUES FOR US EDUCATION

The structure of the education sector and the attendance patterns that were highlighted relate directly to school finances. Because private schools and homeschooling are not publicly supported (to any significant degree), any increased attendance in these sectors relieves state and local governments of resource demands.⁴

Figure 2 traces revenues for the public schools from 1960 to 2019. The bulk of funding comes from state and local revenues, which each correspond to roughly 45 percent of per-pupil funding. The federal share, which began rising in the 1960s as the federal government assumed a larger role in financing schools for disadvantaged students and subsequently for special education students, rose around the 2008 recession and then returned to its historic levels. While not shown, the federal government also contributed large additional amounts of temporary funds with the onset of the pandemic in 2020.

TABLE 1 DISTRIBUTION OF FUNDING SOURCE MAKEUP WITH REPRESENTATIVE STATES, 2019 (PERCENT)

Revenue source	Mean	Minimum	Maximum
Local	42.3	2.1 (Hawaii)	92.0 (Washington, DC)
State	50.1	26.6 (Illinois)	90.3 (Vermont)
Federal	8.6	4.1 (New Jersey)	15.4 (Alaska)

Source: NCES, Digest of Education Statistics 2021, table 235.20.

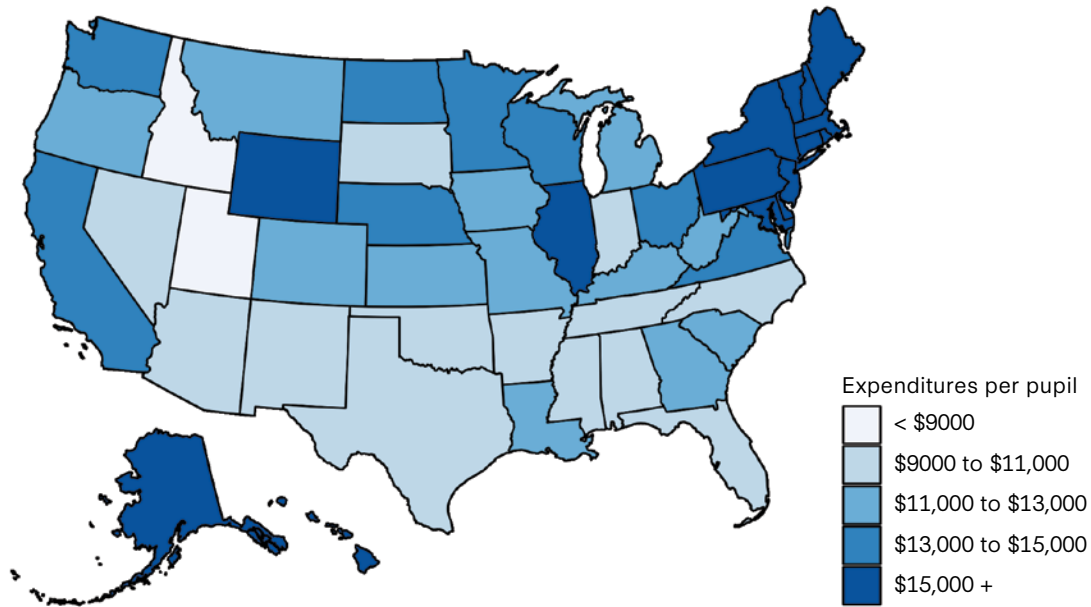
The steady increase in per-pupil funding over the entire period puts public school revenues per student in 2019 at more than four times that in 1960 in real terms. In fact, except for the dip in school revenues after the end of federal support for the 2008 recession, real per-pupil spending (i.e., adjusted for inflation) has risen continuously for more than one hundred years.

State revenues come from a variety of sources that differ across the fiscal structures of the different states. At the same time, with few exceptions, property taxes are the dominant source of local revenues.

Public school spending incorporates both traditional public schools and charter schools. For a variety of political and institutional reasons, charter school spending is systematically below that for traditional public schools, although there is debate about the exact magnitude of differences.

The aggregate revenue data hides the wide variation that is seen at the state level. States differ significantly in how revenues are raised and in the level of spending. Table 1 shows the extent of compositional differences in school funding. Typically, most of the revenue is derived from state and local sources with the federal government contributing a smaller portion, but the federal share across states differs from 4 percent to 15 percent of funding because the federal revenues are driven largely by poverty rates and special education classifications that differ across states. States like Hawaii, with its one district, and Vermont provide almost all funding at the state level, while funding for schools in Washington, DC, is provided almost entirely at the local level. For Alaska schools, 15 percent of the funding comes from the federal government, the highest percentage of all states.

FIGURE 3 Per-pupil expenditure by state, 2019



Source: NCES, Digest of Education Statistics 2021, table 236.65.

Figure 3 illustrates the distribution of state per-pupil spending levels in the 2018-19 academic year. Northeastern states spend more than \$15,000 per student, significantly higher than the \$9,000 to \$11,000 per pupil spent by the majority of southern states.

STUDENT PERFORMANCE

The United States has reliably assessed student performance with the National Assessment of Educational Progress (NAEP), otherwise known as the Nation's Report Card. The long-term trend (LTT) assessment of NAEP makes it possible to get representative national data for math and reading performance of students aged nine, thirteen, and seventeen since the 1970s. Beginning in 1992, a second version of NAEP, called Main NAEP, was started with testing of math and reading in grades four and eight.⁵

Table 2 provides data on NAEP testing results both in terms of changes in standard deviations (SD) and in terms of these changes relative to school expenditure.⁶ The pre-pandemic results fall into two distinct clusters. There are strong gains in the level of math performance for younger students—age nine (grade four) and to a lesser extent age thirteen (grade eight).⁷ But there are much more modest gains for age seventeen math and for reading at all ages.

TABLE 2 NAEP AND SPENDING TRENDS

Exam	Start year	End year	Δ score (SDs)	Δ score (SDs) per 10% spend inc.
<i>Long-term reading</i>				
Age 9	1971	2012	0.3134	0.0266
Age 13	1971	2012	0.2135	0.0181
Age 17	1971	2012	0.0373	0.0032
<i>Long-term math</i>				
Age 9	1978	2012	0.7049	0.0985
Age 13	1978	2012	0.5354	0.0748
Age 17	1978	2012	0.1705	0.0238
<i>Reading</i>				
Grade 4	1992	2019	0.1050	0.0247
Grade 8	1992	2019	0.0867	0.0204
<i>Math</i>				
Grade 4	1990	2019	0.8639	0.2028
Grade 8	1990	2019	0.5399	0.1268

Source: Handel and Hanushek (2023b)

Notes: Δ score (SDs) reports the change in test scores in each respective exam over the period from start year to end year in terms of the individual standard deviation of the exam in start year. The next column reports this value for each 10 percent increase in national per-pupil expenditure (from the base level in start year).

The scores cover different periods of time, so it is also useful within this discussion to place them in comparison to the spending on schools. When normalized by spending over the relevant time periods, the younger cohort math gains are all greater than 0.07 SD per 10 percent larger spending, while the remaining gains are all less than 0.03 SD per 10 percent larger spending.⁸

The results were, unsurprisingly, dramatically altered by the COVID-19 pandemic. The Main NAEP had testing in spring 2019 (included in Table 2) and spring 2022. In math and reading for both grade four and grade eight, average scores fell dramatically with the largest declines being recorded for math performance (Table 3). Grade eight (grade four) gains from 1990 through 2022 were down to 0.33 SD (0.72 SD). For reading, virtually all gains since 1992 were erased by the pandemic; the 1992–2022 gain was 0.01 SD for grade eight and 0.02 for grade four. It is of course difficult to know how to interpret the scores after the pandemic. Clearly, the substantial added funds over the pandemic period were insufficient to overcome the learning disadvantages of the pandemic period.

TABLE 3 PANDEMIC EFFECT ON NAEP SCORES

Exam	Start year	Δ score (SDs), 2019	Δ score (SDs), 2022
<i>Reading</i>			
Grade 4	1992	0.1050	0.0213
Grade 8	1992	0.0867	0.0120
<i>Math</i>			
Grade 4	1990	0.8639	0.7202
Grade 8	1990	0.5399	0.3252

Source: Handel and Hanushek (2023b)

The achievement gains in Table 2 are unconditional changes in student performance. In interpreting this performance data, it is important to note that, because achievement is a function not only of schools but also of parents, peers, and neighborhoods, the data do not indicate the causal impact of schools or spending, but they do provide an important backdrop to finance discussions.

One related pattern that does consider some of the nonschool factors is the historical evolution of achievement gaps by socioeconomic status (SES). Concerns have been raised that the widening of the US income distribution led to expanding SES achievement gaps (Reardon 2011). That concern is unfounded because test information that is linked over time shows a slow shrinking of gaps for birth cohorts born between 1961 and 2001 (Hanushek et al. 2022).

COURT INVOLVEMENT

While the federal courts were involved in school funding issues for a while after the school desegregation ruling in *Brown v. Board of Education*, the US Supreme Court in 1973 declared school finance outside the federal role (*Rodriguez v. San Antonio*), effectively moving all litigation to state courts.

Litigation in the state courts is filed under the state's equal protection clause or the state's education clause as covered by individual state constitutions (see Hanushek and Lindseth 2009). The equity cases under the equal protection clause argue that state efforts to ameliorate either cost of education differences (e.g., for English language learners) or differences in property tax bases are insufficient. The adequacy cases under state education clauses argue that the current level of funding is insufficient to meet the constitutional obligations of the state.

The judicial branch has been asked to assess the level and pattern of school spending in 205 separate court cases adjudicated across forty-eight of the fifty states.⁹ There is no distinct geographical pattern to where these court cases have been found. The prevalence of cases is almost evenly split between below-average and above-average spending states, but the success of defendants in maintaining the existing finance structures is relatively greater in low-spending states. Perhaps surprisingly, decisions in cases focused on adequacy tend to be more successful in states that are already at above-average achievement levels as measured by NAEP.

Interestingly, while the court cases are focused on school spending, there is no overall relationship between spending growth and either decisions that favor the plaintiffs or the number of cases in any state. States with mandates from the courts to increase spending average somewhat larger immediate growth (within five years of the decision) than states where there is no such court mandate, but these short-run changes do not lead to differences in long-term growth of spending. Thus, the school finance litigation has occupied the attention of state legislatures across the country, but it has not changed the overall funding outcomes across the states.

THE SPENDING-ACHIEVEMENT DILEMMA

Since the first major study of school resources and student achievement (Coleman et al. 1966), there have been questions about the strength and consistency of any relationship between the two. This very influential study, the Coleman Report, suggested that school resources were not closely related to student outcomes; instead, families and peers had the primary influence. While the study was not well executed by current scientific standards, it evoked a huge response, with many researchers pursuing related questions about the determinants of student achievement.

The early research confirmed the doubts about whether strong impacts on student achievement would follow added spending (Hanushek 2003). But the early research was marked by studies of highly variable quality, and many would not meet current empirical standards. There are a variety of problems faced by this research, but the main problem is that insufficient attention is given to finding the “causal impact” of added funding. In other words, the correlations of resources and achievement could well be affected by other unmeasured factors that bias any empirical analysis.

A more recent body of research has developed that emphasizes careful identification of the causal impact of resources on student outcomes. The ideal approach to investigating the causal impact of resources is a randomized controlled trial where some group of schools is randomly chosen to receive more resources while another group does not. Such a research design is, of course, not really feasible with schools (or in many other circumstances). As a

result, a variety of other approaches that are designed to mimic randomized controlled trials have been developed. These approaches have two common elements: the existence of a change in resources that is not correlated with other factors that affect student outcomes and the availability of a control group that can indicate what would happen in the absence of the added resources.

Finding circumstances that meet the requirements for these quasi-experimental approaches is not easy. Observations of most actual school operations do not meet these stringent requirements. In fact, the relevant scientific conditions are relatively unusual. But over the past two decades a number of such circumstances have been uncovered by researchers, lending the possibility that evidence on the causal impact of added resources can be more thoroughly investigated.

The studies falling into this category come from a variety of circumstances, ranging from added funding that results from court decisions in finance cases to the impact of budget decreases following the 2008 recession. Because these studies reflect such a wide range of circumstances, it is difficult to provide a direct comparison of the various estimates, but there are now two reviews of the work over the past two decades (Jackson and Mackevicius 2021; Handel and Hanushek 2023b).

Two general conclusions come from the recent studies:¹⁰

1. With high probability, adding resources to schools has a positive effect on student outcomes.
2. The estimated impact of resources is highly variable and depends on the context and constraints on the spending.

The first conclusion largely underscores the contentious political nature of the research in this area. Nobody believes that adding resources to schools is likely to harm students and learning, but because parts of the research enter directly into legislative and judicial decisions about funding, there has been some effort to make this the focus of attention. By phrasing the issue as “does money matter?” the intent is to set the low hurdle of “no harm.” Of course, rational public decision-making would not fund all public programs that don’t harm the recipients.

The second conclusion of the research is much more relevant. The estimates of spending impacts range from too small to reject the possibility of no impact to very large effects on both student achievement and attainment of more schools. The small estimates would not justify added public expenditure because the costs would exceed the social benefits. The large results, on the other hand, would justify considerable commitment of added public funds.

TABLE 4 DISTRIBUTION OF STANDARDIZED SCHOOL SPENDING ESTIMATES

Outcome	Median	Min	Max	N	N pos.	N significant
Test scores	0.070	-0.244	0.543	16	14	9
Attainment	0.057	0.011	0.850	18	18	14

Source: Handel and Hanushek (2023a)

Notes: For test score estimates, results represent the effect of a 10 percent increase in spending on the change in test scores (in individual standard deviation units). For all attainment outcomes, results represent the percent change in the outcome variable for a 1 percent increase in spending. For example, an estimate of 0.05 for graduation indicates that a 10 percent increase in spending led to a 5 percent increase in graduation rates. Estimates are significant if $p < 0.05$.

Table 4 provides a summary of the results from the separate studies of student outcomes that meet modern empirical standards for estimating the causal impact of funding. All estimates represent the expected improvement in outcomes for a 10 percent increase in funding. The preferred estimates relate to achievement test scores. While most are positive and nine of sixteen are statistically significant, they vary widely. Part of the variation just represents normal sampling errors that are present in all studies, but most of it represents true differences in the underlying impact of funding. The estimates for test scores range from a reduction in achievement of -0.24 SD (not statistically significant) to $+0.54$ SD (statistically significant). This large range leaves substantial uncertainty in what can be expected from added funding. Clearly, averaging across these estimates to get a predicted impact would be misleading: in addition to having a small number of estimates in the sample, we could not be confident that they are typical of the full set of funding decisions that have not been measured.

While all of the results for school attainment (high school graduation, not dropping out, and continuing to college) are positive, they also cover a very wide range. They, too, have the same challenges for interpretation.

The major difficulty is that it has not been possible to describe when funds are particularly effective or ineffective (Handel and Hanushek 2023a). The estimated impacts of resources, as noted, come from very different circumstances. They do not reflect differences in the underlying methodology, in whether funds are targeted at a particular group such as disadvantaged students, whether they come from court directives, or whether they reflect differences across states in policies. To date, little headway has been made in describing the features of the particular contexts or the particular use of funds that yields significant learning gains.

In many ways, it is not surprising that the underlying methodology does not provide clear information about the underlying structure of effectiveness. The appeal of randomized controlled trials and quasi-experimental designs is that it is possible to provide causal impact information without knowing or being able to specify the full range of factors that enter into

determining the outcomes. But this does not mean that the specific impact estimates are unaffected by the circumstances or even the design of the specific use of resources. The combination of the use of resources and the context within which they are applied is in *how* funds are used. The current research underscores the importance of how funds are used if student achievement is to be improved.

AN IDEAL FUNDING POLICY

Education policy has two broad goals: reach high levels of achievement and do this in an equitable manner. The way that we fund schools should clearly relate to meeting these goals. The overall level of funding is a political decision, not a scientific decision.¹¹ Legislatures decide on funding levels on the basis of both their judgments about reaching the desired learning standards of the state and their views on the trade-offs with other public expenditures and with private expenditures (as related to tax rates).¹² But because the outcomes of the funding depend on how the funds are used, the education policy surrounding any funding cannot be ignored.

A fundamental problem is that we do not have a set of simple policies that can be put in place and that have a high probability of successful impact on student achievement. We know some things that have an impact, but it is often not clear how they can be put in place at scale.

For example, there is extensive information about the importance of effective teachers (e.g., Hanushek 2011; Chetty, Friedman, and Rockoff 2014; Bacher-Hicks and Koedel 2023). Knowing how important teachers are is different from having a clear set of policies that can be legislated and put into place. There are examples of the application of teacher policies that work in some locations, such as Washington, DC (Dee and Wyckoff 2015, 2017) and Dallas (Hanushek et al. 2023). It is nonetheless difficult to legislate adoption of these complex plans that have been honed to the circumstances of the individual areas.

There are institutional structures that tend to promote better achievement—and that are likely to work in part through promoting better teachers. For example, recent evidence points to good overall performance results from allowing the greater flexibility and parental choice that come with charter schools (see CREDO 2023). Yet the details remain difficult to legislate.

In discussing guiding principles for an effective funding system, Hanushek and Lindseth (2009) proposed seven general principles:

1. If the objective is to improve outcomes, the system should focus on outcomes. Accountability for performance should be substituted for restrictions on local decision-making.
2. The system should reward those who contribute to success—that is, those who bring about high achievement.

3. Rewards should be based on each person's contribution to success and not on external factors such as the education inputs of families and neighborhoods.
4. School funding formulas should minimize unproductive "gaming" by avoiding rewards for things that are easily manipulated by school personnel.
5. School funding policies must recognize the underlying heterogeneity of students and their education challenges and ensure that all schools have the means to succeed.
6. School authorities must gather relevant programmatic and performance data and use it to refine and improve performance.
7. New policies or programs should be introduced in a manner that enables direct evaluation of their results.

These principles can, of course, be filled in a variety of ways, but they revolve around setting up incentives so that the decision-makers take actions that lead to better student outcomes. An example of the application of these principles is what Hanushek and Lindseth (2009) call "performance-based funding."

The central elements of such a system, building on what has previously been successful, include a strong accountability system with incentives and direct rewards for successful performance, empowered local decision-making by both schools and parents, and an ongoing information and evaluation system. This would all be built on a rational and equitable base of funding that provides basic support and that recognizes both different abilities of districts to raise revenue and different costs for educating individuals (e.g., for children from poor families and for students with special needs).

Perhaps the key idea, however, is recognizing and rewarding success. Today many public funding programs actually do the opposite: they reward failure. For example, if a school shows poor performance from its students, more funds are provided; if the school shows improvement, funds are reduced. In other words, they provide an incentive for failure, not for success.

Policies based on incentives for outcomes do not call for completely understanding what works and why. They implicitly acknowledge that there might be alternative ways to achieve the same outcomes and that the choices might reflect both differing demands and differing capacities of schools.

HEADWINDS

An incentive-based funding program faces headwinds from a variety of sources. Perhaps the largest is simply the inertia in the system: "That is not how we do it." There is a long history of approaches to funding that avoid policies offering direct positive incentives. This history is

deeply embedded in both state policies and local decision-making—and leads to a majority of personnel in the current system being happy with the overall structure. Moreover, public views remain supportive of the institutional structure of the public schools. As a result, the system itself resists attempts at alteration.

The strongest force of resistance to change is the teachers' unions. They, as a matter of principle, push back against any attempt to make policies based on differential performance (Moe 2011). As part of this, they resist accountability of schools and of personnel in general, and they resist linking resources to good performance.

At the same time, the unions do not stand alone. This is perhaps easiest to see in states that do not permit collective bargaining and that still resist changes in terms of accountability and incentives. It is also seen in the fact that right-to-work states do not systematically perform better.

COVID-19 brought new challenges to schools, and it has been common to blame all concerns and policy challenges on the pandemic. In reality, NAEP scores began falling after 2012 and simply continued their slide during the pandemic. The prior falls in scores have hit minorities and disadvantaged students exceptionally hard. The COVID cohort as a group has been seriously harmed by learning losses that accrued during the pandemic (Hanushek 2023). Just getting schools back to their 2020 levels appears to be a major challenge in a range of schools. But if we just get back to 2020, the COVID cohort will be permanently harmed. Eliminating the learning losses for this generation is a major policy challenge, but as described, it is far from the only challenge facing the schools. COVID underscores the urgency of the situation but does not provide a long-run solution.

In another matter that affects budgets but is not closely related to student outcomes, many schools are facing significant budget overhang from their retirement programs. The impact of the retirement system varies widely, depending on state rules on funding and depending in part on the character of prior contract negotiations. Most of these issues are beyond the scope of this discussion—with one exception. There is now evidence that schools tend to put too much teacher compensation into retirement plans that are valued by the teachers as having lower value than salary dollars (Fitzpatrick 2015).¹³ Thus, the state funding formula must be sensitive to the incentives sent to districts when they negotiate contracts.

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.

The following finding resonated with me:

The overall level of funding is a political decision, not a scientific decision. . . . We know some things that have an impact, but it is often not clear how they can be put in place at scale.

This resonated for me because scientific discussions rarely, if ever, happen at capitol buildings. Instead, it is a political struggle. Consequently, I am drawn to the recommendation that an incentive-based funding program that rewards outcomes should be considered. If an incentive-based funding program could be devised that rewards performance among comparable schools and school districts (i.e., by comparing schools with similar percentages of students in poverty, with similar expenditures per pupil, etc.), such a system would incentivize the behaviors that we want to see. Would it also be possible to include in an incentive-based funding program additional funds if the district or school leadership were to work with a lower-performing district or school to replicate [higher-performing schools'] success? Can we begin incentivizing districts and schools to scale excellence by offering additional funds?

To help improve education decision-making, our state is in the process of producing data dashboards to document spending and results. A new dashboard published by the Education Oversight Committee with the support of the South Carolina Department of Education can be found at <https://dashboardsc.sc.gov>. The Office of Revenue and Fiscal Affairs will soon release a dashboard focused on how districts spend funds. Such public information may help drive an incentive-based funding program.

—Melanie Barton, senior education advisor for the governor of South Carolina

Dr. Hanushek's chapter thoughtfully captures some of the highlights and lessons learned in the school funding debates over the last forty-plus years, and importantly questions how bold we have really been for students in two areas. First, even in states that have driven toward funding approaches in line with the principles of effective funding systems, good implementation seems to significantly lag the promise of bold policy. In many places across the country, some of the key principles that show promise—like additional investment and expanding the impact of the most effective teachers—seem further away than they were ten years ago, both in terms of broad-based political support and strong exemplars in the field. This chapter challenges us to again center on the principles that connect effective funding with improved student performance.

Second, the chapter pushes us to be bolder in recognizing and rewarding success in educational outcomes. As Hanushek writes, “Perhaps the key idea, however, is recognizing and rewarding success. Today many public funding programs actually do the opposite: they reward failure.” The challenge for policymakers and advocates moving forward is how to invest more resources in the students who need more support while simultaneously rewarding and supporting improved student performance through additional funding. Texas and Tennessee are leading in this approach by funding performance bonuses when K-12 students achieve key outcomes. Striking the balance between funding for need and funding for results will require much bolder policymaking and significantly better implementation in the years ahead.

—David Mansouri, president and CEO of the Tennessee State Collaborative on Reforming Education (SCORE)

REFERENCES

- Bacher-Hicks, Andrew, and Cory Koedel. 2023. “Estimation and Interpretation of Teacher Value Added in Research Applications.” In *Handbook of the Economics of Education, Vol. 6*, edited by Eric A. Hanushek and Stephen Machin, 93–134. San Diego: Elsevier.
- Chetty, Raj, John N. Friedman, and Jonah Rockoff. 2014. “Measuring the Impacts of Teachers I: Teacher Value-Added and Student Outcomes in Adulthood,” *American Economic Review* 104, no. 9 (September): 2633–79.
- Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. 1966. *Equality of Educational Opportunity*. Washington, DC: US Government Printing Office.
- CREDO (Center for Research on Education Outcomes). 2023. *As a Matter of Fact: The National Charter School Study III 2023*, CREDO, Stanford University, June.
- Dee, Thomas S., and James Wyckoff. 2015. “Incentives, Selection, and Teacher Performance: Evidence from Impact.” *Journal of Policy Analysis and Management* 34, no. 2 (Spring): 267–97.
- . 2017. “A Lasting Impact: High-Stakes Teacher Evaluations Drive Student Success in Washington, DC.” *Education Next* 17, no. 4 (Fall): 58–66.
- Fitzpatrick, Maria Donovan. 2015. “How Much Are Public School Teachers Willing to Pay for Their Retirement Benefits?” *American Economic Journal: Economic Policy* 7, no. 4 (November): 165–88.
- Handel, Danielle V., and Eric A Hanushek. 2023a. “Contexts of Convenience: Generalizing from Published Evaluations of School Finance Policies.” National Bureau of Economic Research, NBER Working Paper no. 31653, September.
- . 2023b. “US School Finance: Resources and Outcomes.” In *Handbook of the Economics of Education, Vol. 7*, edited by Eric A Hanushek, Stephen Machin, and Ludger Woessmann. Amsterdam: North Holland.
- Hanushek, Eric A. 2003. “The Failure of Input-Based Schooling Policies.” *Economic Journal* 113, no. 485 (February): F64–98.
- . 2006. “Science Violated: Spending Projections and the ‘Costing Out’ of an Adequate Education.” In *Courting Failure: How School Finance Lawsuits Exploit Judges’ Good Intentions and Harm Our Children*, edited by Eric A. Hanushek, 257–311. Stanford, CA: Education Next Books.

- . 2011. “The Economic Value of Higher Teacher Quality.” *Economics of Education Review* 30, no. 3 (June): 466–79.
- . 2023. “The Economic Cost of the Pandemic: State by State.” Hoover Institution, Hoover Education Success Initiative, January 4. <https://www.hoover.org/research/economic-cost-pandemic>.
- Hanushek, Eric A., and Matthew Joyce-Wirtz. 2023. “Incidence and Outcomes of School Finance Litigation: 1968–2021.” National Bureau of Economic Research, NBER Working Paper no. 31271, May.
- Hanushek, Eric A., Jacob Light, Paul E. Peterson, Laura M. Talpey, and Ludger Woessmann. 2022. “Long-Run Trends in the US SES–Achievement Gap.” *Education Finance and Policy* 17, no. 4 (Autumn): 608–40.
- Hanushek, Eric A., and Alfred A. Lindseth. 2009. *Schoolhouses, Courthouses, and Statehouses: Solving the Funding–Achievement Puzzle in America’s Public Schools*. Princeton, NJ: Princeton University Press.
- Hanushek, Eric A., Jin Luo, Andrew Morgan, Minh Nguyen, Ben Ost, Steven G. Rivkin, and Ayman Shakeel. 2023. “The Effects of Comprehensive Educator Evaluation and Pay Reform on Achievement.” National Bureau of Economic Research, NBER Working Paper no. 31073, March.
- Jackson, C. Kirabo, and Claire Mackevicius. 2021. “The Distribution of School Spending Impacts.” National Bureau of Economic Research, NBER Working Paper no. 28517, March.
- Moe, Terry M. 2011. *Special Interest: Teachers Unions and America’s Public Schools*. Washington, DC: Brookings Institution Press.
- Reardon, Sean F. 2011. “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*, edited by Richard J. Murnane and Greg J. Duncan, 91–116. New York: Russell Sage Foundation.
- US Department of Education. 2022. *Digest of Education Statistics 2022*. Washington, DC: National Center for Education Statistics.

NOTES

1. This section relies heavily on the discussion in Handel and Hanushek (2023b).
2. Statistics on students, districts, and revenues in this section come from various years of US Department of Education data (2022).
3. There are more dimensions of choice, but they do not interact significantly with overall financing. Most importantly, while districts with assigned attendance zones for neighborhood schools predominate, many districts have magnet schools that draw students from the entire district to attend schools with a specialized focus or have open enrollment across all schools in the district. Such choices in general do not affect the total funding for the district, whereas the choices in Figure 1 will affect funding for traditional districts.
4. Note, however, that there have been significant changes in support of private schooling and sometimes homeschooling. The growth in education savings accounts is starting to have a significant impact in some states (see John D. Singleton, chapter 10).
5. Main NAEP has much larger samples of students in order to provide state-by-state performance data. It has also tested grade twelve reading and math and various other subjects such as history, civics, and geography on a less regular basis and using significantly smaller samples of students. These additional tests do not provide consistent time series data.
6. Assessment scores can be reported in different ways. The NAEP assessments are most commonly reported as “scale scores,” which is an arbitrarily set numerical ranking of student

performance. Thus, for example, the announcement of the NAEP scores in math for thirteen-year-olds in 2023 identified a nine-point decline (<https://www.nationsreportcard.gov/highlights/ltt/2023>) compared to students in 2020 before the COVID school closings. Because such scale score changes have little meaning in and of themselves, it is common to translate them into changes in standard deviations, which indicates the dispersion of scores away from the mean. The NAEP decline of nine points is equivalent to a decline of 0.22 SD between 2020 and 2023. This decline implies that the average thirteen-year-old in 2023 would have scored at the 41st percentile (instead of the 50th percentile) of students in 2020.

7. LTT NAEP is age based, while Main NAEP is grade based.

8. A 0.07 SD increase in scores would move the average student from the 50th percentile to the 53rd percentile of the achievement distribution; a 0.03 increase in scores would move the average student from the 50th percentile to the 51st percentile of the achievement distribution.

9. Details of court cases can be found in Hanushek and Joyce-Wirtz (2023). The tabulations consider all school finance court cases between 1968 and 2021.

10. See Handel and Hanushek (2023a) for an analysis of existing quasi-experimental studies.

11. One strand of analysis does “costing-out studies,” which are designed to identify how much it would cost to achieve some specified level of student performance. If, however, we cannot readily identify the relationship between spending and performance, we cannot reliably do these costing-out exercises; see Hanushek (2006).

12. The courts do have a role in addressing the constitutional obligations of each state. State constitutional statements about education vary widely, but most are aspirational and vague. As a result, state court decisions show considerable variation in judgments and in interpretation of the state obligations; see Hanushek and Lindseth (2009).

13. Putting large amounts of compensation into the negotiated contracts may make sense for each district, depending on the funding rules in the state.



9. From the One Best System to Student-Centered Systems

Lessons from a Half Century of K-12 Governance Reform

Michael T. Hartney

Executive Summary

American K-12 education operates at a significant disadvantage. It is burdened by a century-old, one-size-fits-all governance model that prioritizes adult rather than student interests. Owing to interest-group capture, the traditional model of local democratic control—an elected school board, an appointed superintendent, and a central office bureaucracy—is often unresponsive to families and unaccountable to the public for results. What can be done? Since the publication of *A Nation at Risk*, reformers have variously turned to site-based management, state takeovers, and mayoral control to try to weaken the local district and board monopoly. While each of these approaches has improved student outcomes in some systems, none has been a silver bullet. So, rather than seeking to find a single “one best” system, state and local policymakers should focus on identifying a bifurcated strategy to move governance in a direction more focused on student outcomes.

First, for chronically low-performing systems, policymakers can disrupt the “district as monopoly” education provider by pursuing a portfolio management model (PMM) strategy that takes districts out of the business of *running* schools and instead has them provide performance-based oversight in a diverse ecosystem of regulated, but still autonomous, schools of choice. While charter, magnet, and traditional district-run public schools would all be free to pursue their own strategies, they would only be permitted to continue operating in the ecosystem if they meet agreed-upon performance objectives.

Finally, all districts can and should adopt a series of commonsense governance reforms that more tightly link political accountability to student-centered outcomes: (1) establishing on-cycle and nonstaggered school board elections; (2) providing more transparency about

student outcomes timed to coincide with election cycles; and (3) creating mechanisms to change district leadership when students perpetually fail to improve.

- America’s one-size-fits-all school governance system is outdated and ineffective.
- School districts should provide oversight for schools using a variety of strategies to reach agreed-upon educational objectives.
- Electoral success should be linked to student-centered outcomes.



*Improving education in America is hard because it doesn’t have an education system. It has 13,491 of them.*¹

—Dylan Wiliam

*[School] reformers have been busy trying to take politics out of schools rather than considering how politics—of which governance is a part—can be managed, constrained, and transformed to serve public purposes.*²

—Paul Hill and Ashley Jochim

BACKGROUND AND MOTIVATION

Despite its bold rhetoric and urgent call for action, *A Nation at Risk (ANAR)* notably said nothing about reforming “education governance”—the institutions and actors empowered to decide which education policies will (and will not) be put into practice. Nonetheless, shortly after the landmark report ignited a wave of reforms across the states, it became clear to many observers that the nation’s governance system—known colloquially, if not derisively, as the *one best system*—makes it exceedingly difficult to enact reforms that improve student learning at scale.

For example, in their pathbreaking book *Politics, Markets, and America’s Schools*, John Chubb and Terry Moe presaged their indictment of public education at the end of the 1980s by noting: “[The one best system] is so thoroughly taken for granted that it virtually defines what Americans mean by democratic governance of the public schools. At its heart are the school district and its institutions of democratic control: the school board, the superintendent, and the district office.”³ Thirty years later, America remains wedded to this same system, one in which the school district is a sacred cow that often serves the interests of adults more than students. Even the most committed and visionary reformer will make little headway when

constrained by a political system that makes it easier for reform opponents to defeat bold ideas and uphold the status quo.

The simple truth is that the actors who occupy and benefit from our current political institutions have a vested interest in perpetuating the existence of those crusty institutions *irrespective of their performance behind the wheel*. “It is tempting to think that the public schools must be different somehow,” Moe explains. “Their purpose, after all, is to educate children. So it might seem that everyone would want what is best for kids and would agree to change the system . . . [to] make sure it is performing effectively. But this is a Pollyannaish view that has little to do with reality.”⁴

Irrespective of their virtues in other contexts, federalism and localism in K-12 education have evolved to produce a governance system that, due to special-interest capture, is neither responsive to consumers (families and students) nor accountable for producing results. As Chester Finn and Michael Petrilli argue, this one best system offers the “worst of both worlds.” “On one hand, district-level power constrains individual schools; its standardizing, bureaucratic, and political force ties the hands of principals, keeping them from doing what is best for their pupils with regard to budget, staffing, and curriculum. On the other, local control [as practiced in the United States] is not strong enough to clear the obstacles that state and federal governments place before reform-minded board members and superintendents in the relatively few situations where these can even be observed.”⁵

Why is the United States saddled with this patchwork quilt system of school governance? With some simplification, it all boils down to a historical accident followed by a combination of what political scientists call policy diffusion and path dependence (a fancy term for institutional stickiness). Most notably, the key developments that brought and then locked the current system into place had everything to do with adult concerns and very little (if anything) to do with designing a coherent education system to best serve kids. Political scientist Vladimir Kogan outlines the “bottom-up” origins of the first key development—US education’s commitment to governance that is local and diffuse rather than centralized and coherent:

In much of the developed world, schools are typically overseen by centralized national agencies. [The US] model is largely a historical artifact, dating back to the first public-education law adopted in the Massachusetts Bay Colony in the mid-1600s. As evident from the law’s title, the Old Deluder Satan Act [1647], it was the moral concerns of adults, rather than a desire to address the holistic educational needs of children, that mainly drove the public-school effort. . . . The Massachusetts law, which charged local government with the responsibility for funding and operating local schools so kids would become literate enough to read the Bible, was copied across the country in one of the earliest examples of what political scientists now call policy diffusion.⁶

Later, in the early twentieth century (1890-1930), the moral concerns that Kogan highlights here were superseded by more modern, secular ones: leaning on public schools to assimilate immigrants and prepare workers for a second wave of industrialization. Governance experts Paul Hill

and Ashley Jochim deftly summarize the most important changes that accompanied this latter development, the ones that ultimately gave us the one best system that we have today:

Progressive Era reformers sought to rationalize and centralize control of the system. . . . They hoped to create more capable schools—better than the fragmented one-room schoolhouses that dotted the rural landscape and less political than the patronage-driven system that dominated urban centers. Thus emerged the local education agency (LEA). The core of an LEA was an elected school board with power to make most [education] decisions and a bureaucracy largely staffed by professional educators. The LEA was insulated from normal local politics by off-cycle nonpartisan elections. . . . [This] rationalized system . . . gave way to a larger and politically fragmented system in the second half of the 20th century. Laws to encourage and broaden the scope of collective bargaining among public sector employees . . . greatly strengthened teachers' unions.⁷

One final development warrants a brief mention: the district consolidation movement. As Christopher Berry and Martin West document, between 1930 and 1970, the nation's tiny one-room schoolhouses were steadily supplanted by the age-graded schools we know today.⁸ This shift, Kogan explains, "necessitated consolidation into larger school systems, moving the locus of political control from boards overseeing individual schools to districtwide bodies [LEAs]."⁹ Ultimately, the nation eliminated one hundred thousand districts, and consolidated LEAs became larger bureaucracies. What did all this mean for students? Berry and West found that "although larger districts were associated with modestly [better student outcomes], any gains from the consolidation of districts . . . were far outweighed by the harmful effects of larger schools."¹⁰

The key point in all of this is that the forging of education governance in the United States was, as Kogan emphatically states, "not intentionally designed with student academic outcomes in mind and has become less local (and perhaps less democratic) over time."¹¹ In other words, largely through historical happenstance, today we are saddled with the worst of both worlds: a system that is neither especially responsive to community (and especially parental) concerns nor efficient at ensuring that system leaders prioritize student learning outcomes.

The aim of this chapter is straightforward: to assess what the education community has learned since *ANAR* about the challenges to good governance and the most promising solutions for reform. The chapter proceeds in four parts. I first summarize the major political obstacles that have kept a lid on education reform in the United States. After laying out these challenges, I discuss some of the governance reforms that have been tried and what the scholarly evidence says about how those efforts have fared. The third section of the chapter condenses the research into some lessons for policymakers who are considering different governance changes. Since America's students cannot afford to wait for politicians to construct the perfect governance *system* from scratch (an impossible task), the chapter concludes with two types of recommendations for how state and local policymakers can move toward more *student-centered governance systems*: (1) an ambitious alt-governance framework well suited to troubled districts that need immediate and dramatic turnaround, followed

by (2) a more modest set of reforms that are likely to do no harm and some reasonable amount of good in most any district. The guiding ethos in both sets of recommendations is the belief that enough lessons have been learned about governance in the intervening years since *ANAR* to identify a set of best practices for adopting political structures that incentivize the adults in districts and buildings to put student outcomes at the center of policymaking and day-to-day decision-making.

Before proceeding, the reader should be aware of two scope conditions. First, because of their relative fiscal contribution (large) and their central role in implementing policy on the ground, governance issues related to state and (especially) the local school district (rather than the federal government) are the primary concern addressed in the chapter. Second, when discussing problems and solutions, the chapter starts with the point that improving student academic outcomes is the *central* purpose of public education and that other values and “community interests” are of secondary importance. Focusing on how governance can enhance (or impede) reforms intended to bolster student learning outcomes is consistent with the spirit of the goals of *ANAR* (student achievement) and the public’s primary concern with their schools.¹² With these two caveats out of the way, let us turn to discuss the many challenges of America’s traditional model of school governance, better known as the “one best system.”

GOVERNANCE CHALLENGES

The excellence movement that arose out of *ANAR* had two primary objectives: to raise student achievement and to close performance gaps between poor and advantaged students. As is well illustrated by the other chapters in this series, while the federal report helped drive education reforms in several different areas (often with mixed results), all these efforts faced a common hurdle: overcoming political resistance and governance challenges.

While all reforms faced these challenges, two proposals garnered outsized political resistance: (1) school choice and (2) consequential accountability. This is hardly surprising. As Terry Moe explains, “The two great education reform movements of the modern era, the movements for accountability and for school choice, are attempts to transform the traditional structure of the American education system—and the changes they pursue are threatening to the [teachers’] unions’ vested interests.”¹³ Since *ANAR*, the choice and accountability movements’ most significant political victories have been (1) the rapid expansion of charter schooling (1990–present) and (2) the *consequential* test-based federal accountability regime that endured during the Bush and Obama presidencies (2002–2015).¹⁴

A complete assessment of the impact of these policies on student learning is beyond the scope of this chapter. However, research has shown that both choice¹⁵ and accountability¹⁶ reforms can *improve* student achievement and promote education opportunity for underserved kids but that success has often been uneven and difficult to sustain, especially at a statewide (let alone national) scale. For example, the demise of consequential test-based

accountability and the difficulty of increasing the number of high-quality school choice options (e.g., charter schools) can both be traced to major shortcomings in the policies and practices of our traditional system of K-12 governance and politics. Three persistent challenges stand out.

ADULTS ARE NOT INCENTIVIZED TO PRIORITIZE STUDENT OUTCOMES

First, the current governance system does little to nothing to ensure that education professionals are sufficiently incentivized to prioritize student learning above all else. In 2009, for example, just four in ten superintendents surveyed by the National School Boards Association (NSBA) said that student learning was an “extremely important” factor in how they were evaluated by their school board employers.¹⁷ These results mirror a more recent analysis of North Carolina superintendent contracts that showed fewer than 5 percent of these agreements contain provisions to hold leaders “accountable for student achievement and attainment [outcomes].”¹⁸

The failure of too many school boards to prioritize and focus on student outcomes is a widespread problem with tangible consequences. For example, one analysis of the NSBA data uncovered a strong relationship between a school district’s academic performance and the extent to which board members prioritized student achievement outcomes in their board work.¹⁹ Alarming, though, while two-thirds of school boards agree that “the current state of student achievement is unacceptable,” nine out of ten boards said that “defining success only in terms of student achievement is narrow and short-sighted . . . and one-third are nervous about placing ‘unreasonable expectations for student achievement in our schools.’”²⁰ School districts send the wrong message (and the wrong incentives) to the education professionals they employ (e.g., teachers, superintendents) when they make student outcomes a secondary concern. Indeed, elected board governance may not work at all if boards aren’t held accountable by voters for learning outcomes or they don’t expect to be held accountable at the ballot box.²¹

COORDINATING MULTILEVEL GOVERNANCE IS A CHALLENGE

Everyone seems to acknowledge that K-12 governance has too many cooks in the kitchen such that “if everybody is in charge then no one is.”²² This “tangled web” of school governance challenges the public to hold any single entity or public official accountable and encourages political buck-passing.²³ Unfortunately, this problem is inherent in our federal political system. Political scientist Patrick McGuinn refers to it as the 50/15,000/100,000 problem, noting: “We have fifty different state education systems which collectively contain approximately 15,000 school districts and almost 100,000 schools. While the US now has clear national goals in education, it lacks a national system of education within which to pursue these goals, and the federal government can only indirectly attempt to drive reform through the grant-in-aid system.”²⁴

Uncle Sam tried to step up to the plate in 2002 with the federal No Child Left Behind (NCLB) law. By requiring that student performance outcomes be made public, the law was intended to put pressure—including electoral pressure—on school boards to either improve or face

consequences. Unfortunately, the devil was in the details, and federal accountability mandates failed for two primary reasons. First, the law prioritized student academic proficiency over student learning gains (growth), leading many schools where students were improving to be classified as failing. Second, as political scientist Paul Manna has documented, NCLB erred by taking the sound logic of public administration (management) theory and turning it on its head. For example, rather than have the principal (the federal government) set rigorous standards and free up the agents (states and local districts) to innovate and meet these standards in creative ways, the law let states set their own standards while Washington dictated weak and specific consequences for failure.²⁵

Perhaps the problem is not so much too many cooks in the kitchen, but rather that the kitchen lacks thoughtful coordination, and we have not placed each cook at the station where they have a “comparative advantage.”²⁶ For example, NCLB was born out of a real problem whereby localities gave insufficient attention to (and often hid) poor academic outcomes and achievement gaps, but the federal foray into accountability also served to remind us that localities are functionally needed to implement reform from afar.²⁷ Yet, as previously noted, those localities are easily captured by vested interests, and they themselves have incentives to focus on maintaining their institutional existence rather than holding themselves to account. For example, under both NCLB and Race to the Top (RttT), states and districts “took the easy way out,” rarely opting to impose the toughest forms of restructuring on themselves.²⁸

VESTED INTERESTS DOMINATE EDUCATION POLITICS

The third major obstacle to effective governance is the fact that too many adults—be they union leaders, school employees, administrators, colleges of education, or vendors—either benefit from existing K-12 policies and procedures or are reluctant to consider any reforms that may bring about changes that leave them materially worse off. Such opposition ensues even if proposed reforms could be shown to benefit student learning.²⁹ Because vested interests pursue concentrated occupational benefits whose costs are widely distributed, these actors tend to be more politically organized and influential than groups like parents, whose own connection to their public schools is transitory in nature. What’s more, the widespread use of nonpartisan off-cycle school board elections often ensures low voter turnout and a lack of robust competition among competing interests. This anemic electoral environment enables teachers’ unions to win seven out of every ten school board elections when they make an endorsement. The consequence: rather than management (school boards) representing parents and taxpayers by serving as a “check” on labor, the relationship becomes reversed, with management owing its very election and political survival to the employees it is supposed to hold accountable.³⁰ This well-documented dynamic has been shown to lead directly to pro-union school boards that (1) agree to more restrictive collective bargaining agreements (CBAs),³¹ (2) authorize fewer charter schools,³² and (3) spend more on salaries with little to no improvement (and often worse outcomes) in student achievement gains.³³

Although they arguably face greater political competition in federal and state politics, teachers’ unions are still rated the top education lobby in most statehouses, limiting experimentation with choice and accountability, especially on issues related to teacher accountability and pay

reform.³⁴ Finally, teachers' unions are not alone in opposing new approaches to public education outside of the traditional district delivery model. School board members (regardless of party) are far less enthusiastic about school choice and charter schooling than are parents and the public. Yet many states still have charter school laws that either make boards the sole authorizer or limit growth through caps that unions and board associations lobby for in state law. All in all, the politics of education reform remain constrained by governing structures (formal and informal) that empower the producers of education (e.g., teachers' unions, district central offices) at the expense of the consumers of it (parents and students).³⁵

ATTEMPTS TO REFORM THE ONE BEST SYSTEM

Looking back on the history of education in the United States, one can't help but notice the governance pendulum swinging back and forth between decentralization and centralization. The hyper-localism that originated in the mid-1600s held sway until the turn of the twentieth century before yielding to the Progressives' centralized and professionalized LEA. A few decades later, that bureaucratic one best system became a focal point of contention between teachers' unions and minority communities in New York City who wanted more of a say in their kids' schools—what they called “community control.” While the unions, led by then United Federation of Teachers (UFT) leader Albert Shanker, mostly won that battle and the primacy of the central office endured, by the 1980s advocates of a new strategy they called “site-based management” (SBM) were pinning their hopes on giving schools, rather than districts, more autonomy. When student outcomes again failed to improve in any meaningful way, especially in large urban districts, reformers once again saw potential in recentralizing, pursuing alternatives to school board control through mayoral control of the district³⁶ or through state takeovers.³⁷ At the federal level, after promising for decades to “end federal meddling in our schools,” in the 2000s a Republican president embraced more centralized accountability with NCLB, ushering in a decade of bipartisan support for a test-based accountability regime overseen by Washington.³⁸ After political and practical considerations rendered NCLB unworkable, a new breed of school reformers focused on building “parallel” school systems, abandoned trying to bring political reform to the one best system itself, and turned their attention to expanding local autonomy linked to greater school choice (charter schooling).³⁹ In some cases, such efforts have even included trying to partner with or reconstitute districts under a “portfolio” management model (PMM) that combines district accountability/oversight with local school autonomy/choice.⁴⁰ Have any of these governance reforms worked, and if so, where and under what conditions?

SITE-BASED MANAGEMENT

The earliest efforts to rethink K-12 governance after *ANAR* were a series of “site-” or “school-based” management reforms that spread across several states (e.g., Kentucky) and cities (e.g., Chicago). It is difficult to provide a coherent definition of SBM because the specific changes implemented across states and districts that all claimed to be using “SBM principles” varied significantly. However, some common SBM themes that emerged at various

implementation sites included decision-making councils at the school level rather than the district level, formal representation for stakeholders like parents and educators, and direct involvement in hiring building leaders and instructional staff.⁴¹

SBM's "theory of action" is that taking power away from central-office bureaucrats and giving more autonomy to school leaders (with input from educators and families) promotes innovative and customized solutions that result in more effective teaching and learning in buildings and classrooms.⁴² According to one estimate, as many as 30 percent of all US school districts tried some variation of SBM by 1990. However, little systematic evidence emerged to show that the SBM model—at least as it was put into practice—widely improved student learning outcomes across implementation sites at scale.⁴³ To be clear, this is *not* because the idea of having local councils or providing greater autonomy to building leaders is wrongheaded. To the contrary, a recent study from Chicago Public Schools (CPS) found that "schools with high-quality principals and student populations requiring atypical policy decisions [benefit] from more autonomy."⁴⁴ However, that analysis showed that leader quality is often the linchpin to making governance reforms work in practice. As the author of that CPS study concluded, "[school] autonomy should be granted to effective and motivated school leaders [but it may] lead to worse outcomes in settings with agency problems or low principal capacity. . . ."⁴⁵ In other words, successful governance reforms cannot rely solely on building better institutions. Better people (human capital) is a prerequisite to reaping the rewards of well-designed institutions.

Finally, retrospective evaluations of SBM reform frequently mention another challenge that inhibited success: the lack of political will in following through on authentically devolving power and autonomy to building leaders. In practice, many state and district leaders talked a big game about handing over decision-making authority through SBM but were subsequently unwilling to yield on big-ticket items (e.g., budgeting, hiring) when push came to shove or vested interests resisted.⁴⁶ As Paul Hill and Ashley Jochim explain:

School boards and state governments may promise to give schools a great deal of freedom, but over time they take it away. . . . This first became evident with SBM. In the early 1990s, many districts encouraged schools to use time and money in novel ways. . . . Superintendents encouraged principals and teachers to think big, *but no rules were changed*. Schools were encouraged to think of new ways to organize teaching, *but they were still bound by the collective bargaining agreement*. That meant school leaders had little control over who was assigned to teach in the school and the kinds of work they could do. Schools were encouraged to use time and materials differently, *but they did not control their budgets or make purchasing decisions*. And so on. In any clash between school autonomy and actual practice, school leaders soon learned that for every freedom they were promised [under SBM], a rule existed that effectively took it away.⁴⁷

ALT-GOVERNANCE (MAYORAL CONTROL, STATE TAKEOVERS)

Because they are keenly aware of the linkage between education and economic growth in their states and cities, political executives like governors and mayors were often in the vanguard of the excellence movement right from the outset of ANAR. Frustrated with the outright

failure of their cities' largest school systems to improve academically, in the 1990s several mayors sought more authority in especially long-troubled districts (e.g., Boston, Cleveland, Chicago, Detroit, New York). The two primary approaches to robust executive involvement became state takeovers and mayoral control/involvement. While these alternative or "alt-governance" arrangements often involve different mechanisms, they share the common feature of removing or demoting elected school boards, either replacing them with a mayor-appointed board or relegating the board itself to have mere "consultative" status in lieu of policymaking authority. Importantly, in such cases, the district superintendent is chosen by and serves at the pleasure of the mayor—or in the case of takeover, the state education agency (SEA).⁴⁸

Mayoral control's "theory of action" arises from the belief that political executives are more likely to focus on their political legacies (what's best for their city) than parochial-minded legislators (e.g., school board members) who are more prone to single-issue interest-group capture. "Mayors," Terry Moe explains, "are constantly in the public eye; they have larger, more diverse constituencies than school board members do; they have far more resources for wielding power; and they may decide to make their mark by reforming the local schools."⁴⁹ Additionally, one benefit to vesting education authority in a mayor or governor is that it can streamline political accountability under a single actor, making it easier for the public to know whom to hold accountable. Indeed, some research has shown a linkage between greater state-level centralization and student performance: gubernatorial authority to appoint state boards/chiefs has been connected to better outcomes on the National Assessment of Educational Progress (NAEP) and reduced achievement gaps.⁵⁰

Admittedly, efforts to evaluate the impact of mayoral control or state-led takeovers are hampered by small sample sizes and obvious selection biases: districts that turn to mayors for help or those that are taken over by SEAs are difficult to compare to districts that do not have these governance reforms imposed on them. Nonetheless, it is worth noting that the most comprehensive empirical assessment of mayoral control tends to show positive effects on both academic outcomes and fiscal efficiency.⁵¹ Yet it is equally important to keep in mind that not all mayoral involvement is similar in nature. Mayoral involvement in education in cities like Cleveland and Boston operated very differently than it did in New York City and Washington, DC. In the latter two cases, the political executives of those cities were given *complete* autonomy to choose the district's superintendent, and there was no policymaking school board with which the superintendent had to deal politically. Moreover, in the case of Washington, DC—arguably the most successful mayoral turnaround story—the mayor won additional governance changes that empowered the superintendent in hiring and evaluation, removing these policies from the collective bargaining process. Therefore, while research shows that mayoral control in Washington led to reforms that improved student achievement outcomes in the nation's capital,⁵² it does not necessarily follow that more minor forms of mayoral involvement (e.g., appointing a few of a city school board's members) will replicate this unique success story. Indeed, one factor stands out in helping to explain why mayoral control in Washington led actors to prioritize student, rather than adult, interests: centralized political accountability. One anecdote from that city is especially telling. Years after departing his post as president of the Washington, DC, teachers' union, George Parker explained, in

retrospect, why mayoral control forced his hand in accepting a student outcomes-focused teacher evaluation system:

One of the most important things is that we went from board governance to mayoral control. . . . Previously I was able to use politics to block a lot of reforms. But once mayoral control came into place, and there was only one person who had all the control, I no longer could prevent a lot of the reforms, so I had to decide: do I take a good look at these reforms and how do these reforms impact students, or do I try to continue to fight?

In my previous contract [negotiations] when the Superintendent put things on the table that I didn't like all I had to do was go to several of the board members that we supported financially and just say, 'We helped get you elected'. . . . And I come back to [the] negotiating table the next day and it's off the table. When we had mayoral control there was only one person. And I tried it with Mayor Fenty. I remember I went down to his office, but he made it clear that he promised Michelle [Rhee] that he was going to support what it was she was going to do. So, for the first time, to be very honest, I had to take a different position for negotiations because I had no one to go to [to] block reform.⁵³

In a similar vein, advocates of state takeover can point to impressive turnarounds like New Orleans, where the bold post-Hurricane Katrina choice and accountability reforms overseen by that state's "Recovery School District" (RSD) led to dramatic improvements in student outcomes in both achievement (test score gains) and attainment.⁵⁴ To be sure, New Orleans does not represent the typical state takeover. As Terry Moe explains, the all-charter system that emerged in the aftermath of the storm was an extreme outlier that was made possible by the sudden elimination of vested interest opposition (United Teachers of New Orleans and the Orleans Parish School Board).⁵⁵ In fact, the most comprehensive empirical study of state takeovers to date found little systematic evidence that abolishing local control (elected boards) leads to higher student achievement at scale.⁵⁶ Moreover, critics can and do point to a clear downside of state takeover: disempowering communities from having a direct hand in running their local public schools, with communities of color being disproportionately targeted for takeover.⁵⁷

On the other hand, the average effect of state takeover may not be the right quantity of interest to focus on given the theory of action for granting states temporary control. As with mayoral control, state takeover advocates rightly note that democratic accountability can become so broken in some school districts that boards can no longer be trusted to do right by their kids and that dramatic leadership change is needed. Of course, not all state takeovers are created equal; for example, some are driven by fiscal concerns and others are provoked by chronic student achievement failure. *What seems to matter most is what policymakers (state leaders) do with their newfound authority when takeover occurs.*⁵⁸ For example, research shows that when states can use takeovers to close a district's lowest-performing schools and replace them with higher-performing schools, student outcomes can and do improve substantially.⁵⁹ But the key to an SEA succeeding in this endeavor is ensuring that students will, in fact, move to a higher-performing school. If students are instead relegated to another

low-performing school (or even a middling school), then the instability associated with moving schools can be a net negative for student learning.⁶⁰ It is not altogether surprising, then, that state takeovers have been a mixed bag. Takeovers in Camden (NJ), Newark (NJ), and especially New Orleans—where the close and replace strategy was pursued—stand out as successful. In contrast, both Michigan’s and Tennessee’s efforts to replicate Louisiana’s success in New Orleans fell short.⁶¹

PORTFOLIO FRAMEWORK OR PORTFOLIO MANAGEMENT MODEL

Frustrated by the lack of progress in turning around chronically low-performing schools in the late 1990s, political scientist Paul Hill began to advocate for a new governance framework for large city school districts: the portfolio management model.⁶² In one sense, PMM was partly an effort to fix a core failure of SBM—the unwillingness of states and districts to hand over the car keys of autonomy on key issues like budgeting and hiring to school leaders. But PMM proposed even more.

PMM reimagines the district’s role as the monopoly education provider (e.g., “district schools”) and instead sees its role as a chief incubation officer that simply oversees “schools.” In other words, PMM envisions getting districts (e.g., school boards, central offices) out of the business of *running* school buildings and into the business of gently overseeing an ecosystem of autonomous schools of choice. But PMM is not an unfettered school choice program. To the contrary, the framework melds autonomy and choice with a centralized accountability system for all schools (irrespective of type) and (often) a single districtwide application process. While charter schools, magnets, and traditional district-run schools are all free to innovate at the school building level under the PMM framework, *all schools, irrespective of type, are only permitted to continue operating if they meet agreed-upon performance objectives*. In part, the allure of the PMM approach is that it helps soften the unhelpful charter versus traditional public school debate because the district and charter sectors are incentivized to collaborate with all schools in the portfolio, as every school is seen as an equal member of the same citywide ecosystem.

Where has it been tried and how well has it worked? Standouts include New Orleans, Denver, Indianapolis, Washington, DC, and New York City.⁶³ Notably, several of these cities pursued alt-governance models first or along the way, which helped provide (at least temporary) political cover for this choice ecosystem to blossom and gain constituents (families) whose favorable experience in this new system could create a new constituency that would protect the model from being undone by vested interest opposition. However, alt-governance clearly is not a prerequisite to embracing PMM, and there is no single definition of the approach in practice, perhaps other than sector agnosticism (charters and district-run schools are equal in the eyes of the system). In fact, in some cases, because traditional district-run schools have seen firsthand some of the advantages of site-based autonomy in personnel and school calendar/time use, for example, PMM has led to state legislation that spawned charter-like district schools, called “innovation schools,” in Indianapolis and Denver. On the other hand, progress has been uneven in many of the other systems that have incorporated

PMM principles. In 2022, Hill and Jochim reported that “of the 52 districts that participated in CRPE’s portfolio network and nominally adopted the strategy at some time or another, few sustained it for more than a few years.”⁶⁴ Moreover, the charter-district détente that PMM imagines has been far less successful in systems with strong teachers’ unions, such as Los Angeles.⁶⁵

One aspect of the theory of action behind PMM is that offering more options whets the appetites of and expectations among families for the district to provide them with a variety of learning models from which to choose. One of the most powerful levers of policy reform is the ability to create new constituencies who have a vested interest of their own in new school models and delivery systems. Creating value for education consumers (parents) and *potential* consumers will give more voters reason to defend the entire fleet of options in a district’s portfolio, and future board members who wish to go back to “the way things were” (with the district as sole provider) may find themselves facing political resistance that rivals the power of locking in a formal governance change in law or regulation. This matches the well-known (successful) mobilization effort among charter school parents to prevent New York City’s then incoming mayor, Bill de Blasio, from diminishing the charter sector that they had a personal stake in continuing to use. In that way, PMM helps reshape the politics of education more generally.

LESSONS AND RECURRENT TENSIONS IN GOVERNANCE REFORM DEBATES

What broader lessons can policymakers, reform advocates, and educators take away from past and present efforts to use governance changes to spur school improvement? Relatedly, what are the key tensions in our governance reform debates that are likely to persist moving forward?

1. DEMOCRATIC PROCEDURES ARE LESS IMPORTANT THAN DEMOCRATIC OUTCOMES

“Fundamentally, democracy is really about representing the interests of adults,” Vladimir Kogan explains. “Whether school board elections are democratic tells us absolutely nothing about whether public schools are doing a good job delivering on their core mission [of educating kids].”⁶⁶ In other words, when policymakers sit down to evaluate K-12 governance models, they should recognize the difference between democratic procedures (important) and the substantive outcomes that public education is trying to achieve: creating an educated populace that is equipped to participate in self-governance (most important). Consider, for example, the tension between the right for students to go to school and learn without interruption and the right of school employees to pursue their occupational self-interests through a labor action. This is not a hypothetical. Teachers’ unions often claim that the right to strike fundamentally promotes democracy for workers (their members), yet we know that keeping

children out of school for prolonged periods of time is not in their best interest.⁶⁷ How should policymakers wrestle with these tensions, ones where democratic procedures collide with democratic outcomes? Consider the following thought experiment (again) from Kogan:

In many communities drinking water is delivered by public agencies. Yet very few people ask if these agencies are democratic. They ask whether they deliver clean and safe water. I think few would be okay with these agencies delivering cholera contaminated water just because they were satisfied with voter turnout and other metrics of democratic process or procedure. In many parts of the US, we also have publicly run hospitals. Again, when we're evaluating their performance, I think most people care about how all these hospitals are serving patients, not about whether their board meetings follow Robert's Rules and allow opportunity for community engagement.⁶⁸

As agencies of government (subject to the demands of interest groups and voters), public schools will always be in the political arena. And to be sure, many adults will have a vested interest in upholding school board governance and in maintaining the traditional district/LEA as the sole provider of public education. These actors have obvious incentives to oppose alt-governance arrangements or portfolio management approaches. Policymakers should expect nothing less. However, at the end of the day, policymakers will need to prioritize, while remembering, most of all, that public education systems exist to serve students, not adults.

2. THERE'S NO "FOOLPROOFING" A GOVERNANCE SYSTEM IN THE ABSENCE OF POLITICAL WILL AND BOLD, CAGE-BUSTING LEADERSHIP

Well-defined governance arrangements with clear lines of accountability are typically necessary to deliver improved outcomes for kids, but they are almost always insufficient to the task at hand. Well-designed governance systems are only as good as the leaders who make use of them. As the author of a recent book on the delivery of government services in our digital age put it, "culture eats policy's lunch."⁶⁹ In the case of education reform moving the needle for kids, this means that governance reform can create new possibilities and provide political cover, but it takes bold leaders to step up to the plate and make use of those new institutional levers. For all their faults (noted below), the architects of the turnaround in Washington, DC—then chancellor Michelle Rhee and then mayor Adrian Fenty—were each willing to put it all on the line and make tough decisions to change the culture of the city's school system (and its future trajectory) even when those decisions cost them their jobs. In a similar vein, recall the key finding about the importance of leadership from economist Kirabo Jackson's study of school autonomy in Chicago that was discussed earlier in this chapter. Jackson found that providing more school-level autonomy to principals improved student learning outcomes in schools with high-quality leaders. In places where leaders had a poor or middling track record, providing greater autonomy predictably did not lead to better decision-making and did not improve student outcomes; it led to worse performance. In sum, strong district and school leadership both matter immensely.

3. LOCK IN GOVERNANCE AND POLITICAL REFORMS TO INCENTIVIZE STUDENT-CENTERED DECISION-MAKING WHENEVER POSSIBLE, BUT REMEMBER THAT ETERNAL VIGILANCE WILL REMAIN ESSENTIAL

As we've seen with the history of both the SBM and PMM governance reform models, politics always has a way of undoing progress, and a reform-minded majority today is no assurance of one tomorrow. When in power, reformers should try to lock in governance reforms that will maximize the chances that future district leaders will remain student centered in their decision-making. For example, in New Orleans, state lawmakers ensured that even after RSD transferred authority back to the local Orleans Parish School Board, the superintendent would retain authority to hold schools accountable without meddling from individual board members. This was crucial, because the entire PMM framework functions only when school renewals are based on transparent and objective student performance criteria, not political criteria such as whether a school is in a board member's electoral district. Similarly, as we saw in Washington, DC, the fact that some key decisions (around teacher evaluation) were taken out of collective bargaining enabled the system leader to make more efficient student-centered decisions when it came to managing human capital. This would not have been possible without changes in the governance protocols centralizing authority in the mayor's office. In Indianapolis, empowering the mayor to authorize charters has helped ensure that the PMM framework can remain in place even if there is board turnover, as has happened in Denver in recent years, putting reforms that helped improve district performance in jeopardy.⁷⁰

4. IN EDUCATION REFORM, A MANTRA OF "MOVE FAST AND BREAK THINGS" OFTEN BACKFIRES

Bedside manner matters in education reform. On the one hand, Americans appear comfortable with their state, rather than local government, addressing chronically failing schools.⁷¹ However, when it comes to formal takeover proposals, issues related to race and the loss of political power become salient in city school systems that were often important sites where racial minorities gained a foothold in politics or found a pathway to the middle class in a teaching career. For example, a survey commissioned by journalist Richard Whitmire found that while many Black Washingtonians believed Michelle Rhee's tenure improved their schools, they also believed her reform methods (e.g., school closures, firings) were overly draconian and unnecessary.⁷² Irrespective of whether the critics are right or wrong on the merits, reformers will come up on the short end of the stick if they refuse to consider the timing, temperament, and input of local actors in an authentic manner. Rhee's own tenure as chancellor was cut short because voters soured on her and Fenty's "move fast and break things" ethos. In contrast, by being more intentionally "collaborative and accessible," Rhee's successor managed to maintain the very same reforms that put the city's children first while keeping her post for three times as long. This isn't a criticism of Rhee per se, but a warning to other reformers who have been turned out of power swiftly because community perception and a lack of engagement did them in (e.g., in Memphis and Detroit).⁷³

To avoid alienating potential allies in the local community, reformers should consider the timing and sequence of their actions. School closures are invariably controversial. When necessary, they should be done using a consistent and transparent set of metrics so that critics cannot claim bias in sites chosen. Additionally, some reformers have been able to put closures off until goodwill has been established in the community, and, especially in the context of takeovers/alt-governance, local actors believe that reform efforts are well intended. This won't please everyone, and opposition will surely remain, but acting capriciously and without any attention to bedside manner is both counterproductive and an unforced, self-inflicted error. In places like New Orleans, Memphis, and Detroit, where takeovers led to complaints about outsiders imposing closures without community input, it is essential for reformers to ensure demographic representation on charter boards and other bodies, for example, so that alt-governance is not interpreted as an effort to disempower local communities.⁷⁴

CONCLUSION AND RECOMMENDATIONS

Given the immense size and scale of public education in the United States, it would be foolish and impractical to conclude this retrospective by recommending that a single governance model be applied everywhere. Instead, the broader lessons that have been outlined here recommend two paths forward on governance reform, with careful attention to context.

In the first case, large school districts with poor academic outcomes that have remained unchanged under the constraints of the traditional "district as monopoly" education provider should give serious consideration to an alt-governance model that would allow for a portfolio framework to blossom. While formal governance changes are not a prerequisite to incorporating the portfolio framework, the author of that reform approach notes that in the absence of "a galvanizing event" or "the entrance of new [often nontraditional] leadership," the "adoption of [the portfolio] strategy [is] often precipitated by a major shift in education governance via state takeover or mayoral control." The reason is simple: "these events [help] to restructure local education politics such that traditional actors . . . [are] sidelined, creating a window of opportunity for new reform ideas to take root."⁷⁵

Since these districts can and will rarely initiate alt-governance on their own (Washington, DC, being a rare exception), leaders who wish to pursue a portfolio framework may do well to begin their effort by working with their counterparts in state government. To avoid the negative perceptions that invariably arise from "outsiders" ignoring local context and concerns,⁷⁶ advocates could benefit by framing their effort to leverage state support as an exercise in "freeing" local schools to enjoy more autonomy or "innovation" opportunities even if they remain under traditional district governance. Alternative governance arrangements need not mean the formal elimination of an elected school board en route to a portfolio framework. As Indianapolis has shown, having an executive (mayor) with charter-authorizing power opens new possibilities. Likewise, Denver Public Schools also remained under elected board

control, but innovation schools there nevertheless provided autonomy and choice consistent with the portfolio framework.

The second path forward is probably more appropriate for the nation's (smaller) suburban and rural school districts that maintain the traditional elected board-appointed superintendent structure. Although these districts (which are more numerous but enroll far fewer students) may not need to abandon traditional governance structures, states should nonetheless require (or at least encourage) them to adopt a series of more modest reforms aimed at promoting a political structure that creates stronger incentives for aligning democratic accountability with improved student academic achievement outcomes.

First, state governments should move to on-cycle school board elections. A political system that allows one special interest group to dominate low-turnout, low-information elections isn't a model of robust democracy.⁷⁷ A large research literature shows that off-cycle elections unfairly advantage unions over other stakeholders and decrease the representation of parents, the poor, and racial minorities in school board elections.⁷⁸ Most importantly, shifting to on-cycle elections increases the likelihood that voters will reward/punish incumbent school board members based on student achievement growth in their district during their tenure.⁷⁹ In sum, this is a small but important policy change that comes with few downsides and a big upside.

Relatedly, states might consider (or at least investigate) the benefits of using non-staggered school board elections. Currently, with staggered board elections, the ability for the public to make a wholesale change in district leadership is deferred across election cycles. If voters are constitutionally empowered to "throw the bums out" of Congress every two years, perhaps they should have that same opportunity in local school politics. This reform would, in theory, also simplify participation in school politics, encourage slate running, and make it easier for the public to identify whom to hold accountable at a given point in time (since all incumbents would run at the same time, there would be a *de facto* referendum on their performance).

Second, as A. J. Crabill has argued, state governments should require school board training or coaching that focuses specifically on student outcomes. Ideally, states could find ways to make this more than a compliance exercise.⁸⁰ In fact, Crabill makes a good case that states could add to this the incentive for board candidates to get certified before running for office. One benefit might be dissuading candidates who do not want to do the serious work and who are running for reasons other than raising district achievement.

Third, states must ensure that their accountability systems provide useful and easy-to-understand information about the performance of each district's public schools. Those metrics should include and emphasize information on student growth, not simply proficiency. Letter grades, though imperfect, often make it easier on the public. Importantly, SEAs need to be prepared (and required under state law) to release report card data earlier on and

preferably in the month prior to when school board elections are held, to maximize the likelihood that voters will prioritize student learning outcomes during board elections.⁸¹

States should consider electoral reforms that provide information about student performance on the ballot, identifying any incumbents seeking reelection so that voters know how their board members have fared in raising achievement when they decide whether to rehire them for the job.⁸² As a gentler form of “takeover,” states could first have a policy whereby an automatic board recall election is held when a district’s academic improvement stagnates for a period under the same leadership. Relatedly, similar legislation could call for a superintendent’s replacement in the event of severe achievement failure or stagnation.

FINAL THOUGHTS

A total governance failure is typically observed only in an ad hoc fashion. Examples might include a district embezzlement scheme or a school cheating scandal. This leads to the mistaken belief that K-12 governance problems are rare and isolated to specific districts or leaders. Yet in 2020, the COVID-19 pandemic laid bare the broader dysfunction beneath the surface of America’s traditional system of K-12 board-based governance. While more centralized education systems in other parts of the world reopened far more quickly,⁸³ in our highly decentralized system partisanship and the lack of political will to negotiate reopening agreements with teachers’ unions played no small role in keeping half of all students out of school for a full year. In fact, numerous studies revealed that in the absence of thoughtful state political leadership, too many local school boards made decisions to keep schools closed more because of adult politics than in response to thoughtful reflection about neutral public health criteria, including the cost-benefit calculation regarding what was best for students.⁸⁴

As the second epigraph of this chapter noted, the root of the K-12 governance problem, Paul Hill and Ashley Jochim explain, is that ever since the turn of the twentieth century, “[school] reformers have been busy trying to take politics out of schools rather than considering how politics—of which governance is a part—can be managed, constrained, and transformed to serve *public purposes*.”⁸⁵ This failure of imagination is a key reason that our public schools are encumbered by bureaucratic structures and work routines that too readily prioritize the interests of adults rather than the students they serve. Ironically, then, one hundred years after progressive reformers dismantled the nation’s large and unwieldy urban school boards, America’s fourth-largest school district, CPS, is returning to this relic of the past. Despite making real strides under mayoral control,⁸⁶ at the behest of the city’s powerful teachers’ union, CPS will soon be governed by a large (twenty-one members!) elected board beginning in 2024.⁸⁷ Meanwhile, the SEA in Texas has decided to pursue takeover of the nation’s third-largest district, Houston Independent School District (ISD). The Texas Education Agency recently tapped former Dallas ISD superintendent Mike Miles to bring to Houston the muscular human capital reform strategy previously pursued in Dallas. Miles has announced that he will use his authority to introduce pay incentives that induce top teachers to work in struggling

schools, an approach that some research shows can make a positive impact on student learning.⁸⁸ Despite the obvious similarities they share in size and demographic challenges, Chicago and Houston suddenly appear to be two ships passing in the night. They remind us once more that the decentralized nature of K-12 politics and governance too often influences a child's chances of receiving a high-quality education and obtaining a shot at upward mobility in this patchwork quilt we call public education in the United States.

HESI PRACTITIONER COUNCIL RESPONSE

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.

Michael Hartney's paper makes several important points. First, the multiple levels of governance responsibility are disjointed, overlapping, and confusing—and mostly undefined. Second, school boards generally lack incentives to prioritize student achievement. Third, school board elections tend to be heavily influenced by vested interests, more specifically teachers' unions. The net result is that K-12 governance tends to be ineffectual, especially when it comes to improving educational outcomes.

Among Hartney's most important recommendations is to respond to this sorry state of affairs by using mayoral control or state receiverships not just to stabilize dysfunctional districts and begin the process of school improvement, but to put in place sustainable governance and management systems and structures that can survive a return of control to an elected school board. This approach aligns incentives so that both school personnel and parents have a vested interest in defending school-level autonomy through better student outcomes.

As much as I agree that participation in local elections is to be encouraged as part of good citizenship, I fear that more voter participation in school board elections may have unintended consequences. The problem today is that low turnout privileges the organized professional interests.

I would recommend instead that we have a more robust conversation and hopefully consensus about what the role of a local school board really is. Local school boards should provide for the equitable allocation of resources, offer a robust set of educational options for parents, and ensure that district administration is both supported and accountable.

—James Peyser, former secretary of education for Massachusetts

NOTES

1. Dylan Wiliam, *Creating the Schools Our Children Need* (West Palm Beach, FL: Learning Sciences International, 2018), 6.
2. Paul T. Hill and Ashley E. Jochim, *A Democratic Constitution for Public Education* (Chicago: University of Chicago Press, 2014), 2.
3. John E. Chubb and Terry M. Moe, *Politics, Markets, and America's Schools* (Washington, DC: Brookings Institution Press, 1990), 5.
4. Terry M. Moe, "Vested Interests and Political Institutions," *Political Science Quarterly* 130, no. 2 (Summer 2015): 277–318.
5. Michael Petrilli and Chester E. Finn, "The Failures of US Education Governance Today," in *Education Governance for the Twenty-First Century: Overcoming the Structural Barriers to School Reform*, ed. Paul Manna and Patrick McGuinn (Washington, DC: Brookings Institution Press, 2013), 32.
6. Vladimir Kogan, "Locally Elected School Boards Are Failing: Pandemic Stress-Tested School Governance, Revealing Many Flaws," *Education Next* 22, no. 3 (Summer 2022): 8–13.
7. Hill and Jochim, *Democratic Constitution*.
8. Christopher R. Berry and Martin R. West, "Growing Pains: The School Consolidation Movement and Student Outcomes," *Journal of Law, Economics, & Organization* 26, no. 1 (April 2010): 1–29.
9. Kogan, "Locally Elected School Boards," 10.
10. Berry and West, "Growing Pains," 1.
11. Kogan, "Locally Elected School Boards."
12. For example, as Kogan notes, "According to the latest *Education Next* public opinion survey, two-thirds of Americans say that schools should prioritize academic achievement over student 'social and emotional wellbeing'"; Kogan, "Locally Elected School Boards." Also see Vladimir Kogan, "The Choice in Education Governance Debates: Complacency or Reform?," *Education Next* 23, no. 1 (Fall 2023): 68–74.
13. Moe, "Vested Interests," 305.
14. The federal government's heavy-handed approach to school accountability softened considerably in 2015 when the Every Student Succeeds Act (ESSA) replaced NCLB. Although ESSA still conditions federal funding on state testing, the prescriptive and punitive consequences that NCLB imposed were eliminated in favor of granting states flexibility to respond. For an overview of the evolution from NCLB to ESSA, see Thomas S. Dee, "Learning from the Past: School Accountability before ESSA," Hoover Institution, Hoover Education Success Initiative, June 15, 2020.
15. See, for example, Patrick J. Wolf, Jay P. Greene, James D. Paul, and Matthew Ladner, "Education Freedom and Student Achievement: Is More School Choice Associated with Higher State-Level Performance on the NAEP?," *Journal of School Choice*, published online March 3, 2023; Center for Research on Education Outcomes (CREDO), "As a Matter of Fact: National Charter School Study III," Stanford University, June 2023.
16. See, for example, 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; 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.
17. Frederick M. Hess and Olivia Meeks, "School Boards Circa 2010: Governance in the Accountability Era," Thomas B. Fordham Institute, 2010.
18. Robert Maranto, Julie Trivitt, Malachi Nichols, and Angela Watson, "No Contractual Obligation to Improve Education: School Boards and Their Superintendents," *Politics & Policy* 45, no. 6 (December 2017): 1003–23.

19. Arnold F. Shober and Michael T. Hartney, "Does School Board Leadership Matter?," Thomas B. Fordham Institute, 2014.
20. Hess and Meeks, "School Boards Circa 2010."
21. See, for example, Patrick Flavin and Michael T. Hartney, "Racial Inequality in Democratic Accountability: Evidence from Retrospective Voting in Local Elections," *American Journal of Political Science* 61, no. 3 (2017): 684–97.
22. See, for example, Petrilli and Finn, "Failures."
23. See Noel Epstein, *Who's in Charge Here? The Tangled Web of School Governance and Policy* (Washington, DC: Brookings Institution Press, 2004).
24. Patrick McGuinn, "Divided Democrats: The Two Narratives of School Reform," *EdWeek* (blog), June 18, 2010, <https://www.edweek.org/education/opinion-divided-democrats-the-two-narratives-of-school-reform/2010/06>.
25. Paul Manna, *Collision Course: Federal Education Policy Meets State and Local Realities* (Washington, DC: CQ Press, 2011).
26. See, for example, Frederick M. Hess and Andrew P. Kelly, eds., *Carrots, Sticks, and the Bully Pulpit: Lessons from a Half-Century of Federal Efforts to Improve America's Schools* (Cambridge, MA: Harvard Education Press, 2011).
27. Manna, *Collision Course*.
28. See, for example, Paul E. Peterson, "The End of the Bush-Obama Regulatory Approach to School Reform: Choice and Competition Remain the Country's Best Hope," *Education Next* 16, no. 3 (2016): 22–33; Sara Mead, "'Easy Way Out': Restructured Usually Means Little Has Changed," *Education Next* 7, no. 1 (2007): 52–56.
29. See, for example, Michael T. Hartney, *How Policies Make Interest Groups: Governments, Unions, and American Education* (Chicago: University of Chicago Press, 2022); Terry M. Moe, *Special Interest: Teachers Unions and America's Public Schools* (Washington, DC: Brookings Institution Press, 2011).
30. See, for example, Terry M. Moe, "Political Control and the Power of the Agent," *Journal of Law, Economics, and Organization* 22, no. 1 (2006): 1–29.
31. See Katharine O. Strunk and Jason A. Grissom, "Do Strong Unions Shape District Policies? Collective Bargaining, Teacher Contract Restrictiveness, and the Political Power of Teachers' Unions," *Educational Evaluation and Policy Analysis* 32, no. 3 (2010): 389–406.
32. See Ying Shi and John D. Singleton, "School Boards and Education Production: Evidence from Randomized Ballot Order," *American Economic Journal: Economic Policy*, forthcoming.
33. See, for example, Joshua M. Cowen and Katharine O. Strunk, "The Impact of Teachers' Unions on Educational Outcomes: What We Know and What We Need to Learn," *Economics of Education Review* 48 (2015): 208–23; Joshua M. Cowen and Katharine O. Strunk, "How Do Teachers' Unions Affect Education Outcomes? Reviewing the Latest Research on the Rent-Seeking Debate," in *Walkout: Teacher Militancy, Activism, and School Reform*, ed. Diana D'Amico Pawlewicz (Charlotte, NC: Information Age Publishing, 2022); Michael T. Hartney, *How Policies Make Interest Groups: Teacher Political Activism and American Education* (Chicago: University of Chicago Press, 2022); Terry M. Moe, "Collective Bargaining and the Performance of the Public Schools," *American Journal of Political Science* 53, no. 1 (2009): 156–74.
34. See, for example, Leslie K. Finger, "Vested Interests and the Diffusion of Education Reform Across the States," *Policy Studies Journal* 46, no. 2 (2018): 378–401; Leslie K. Finger, "Interest Group Influence and the Two Faces of Power," *American Politics Research* 47, no. 4 (2019): 852–86; Michael T. Hartney and Patrick Flavin, "From the Schoolhouse to the Statehouse: Teacher Union Political Activism and State Education Reform Policy," *State Politics and Policy Quarterly* 11, no. 3 (2011): 251–68.
35. See, for example, Paul E. Peterson, Michael Henderson, and Martin R. West, *Teachers versus the Public: What Americans Think about Schools and How to Fix Them* (Washington, DC:

- Brookings Institution Press, 2014); Martin R. West, Michael Henderson, and Paul E. Peterson, "The Education Iron Triangle," *The Forum* 10, no. 1 (2012).
36. Henig notes that four major cities (Boston, Chicago, Cleveland, and Detroit) turned to mayoral control in the 1990s. See Jeffrey R. Henig, "Mayors, Governors, and Presidents: The New Education Executives and the End of Educational Exceptionalism," *Peabody Journal of Education* 84, no. 3 (2009): 283–99.
37. See, for example, Beth E. Schueler and Joshua F. Bleiberg, "Evaluating Education Governance: Does State Takeover of School Districts Affect Student Achievement?," *Journal of Policy Analysis and Management* 41, no. 1 (2022): 162–92.
38. See, for example, Michael T. Hartney and Christina Wolbrecht, "Ideas about Interests: Explaining the Changing Partisan Politics of Education," *Perspectives on Politics* 12, no. 3 (2014): 1–28.
39. Peterson, "Bush-Obama Regulatory Approach."
40. See, for example, Paul Thomas Hill, *Strife and Progress: Portfolio Strategies for Managing Urban Schools* (Washington, DC: Brookings Institution Press, 2013); David Osborne, *Reinventing America's Schools: Creating a 21st Century Education System* (New York: Bloomsbury Publishing, 2017); Katrina E. Bulkley, Julie A. Marsh, Katharine O. Strunk, Douglas N. Harris, and Ayesha K. Hashim, *Challenging the One Best System: The Portfolio Management Model and Urban School Governance* (Cambridge, MA: Harvard Education Press, 2020).
41. Jane L. David, "The Who, What, and Why of Site-Based Management," *Educational Leadership* 53 (1995): 4–9.
42. Scott C. Bauer and Ira E. Bogotch, "Analysis of the Relationships among Site Council Resources, Council Practices, and Outcomes," *Journal of School Leadership* 11, no. 2 (2001): 98–119.
43. Rodney T. Ogawa and Paula A. White, "School-Based Management: An Overview," in *School-Based Management Organizing for High Performance*, ed. Priscilla Wohlstetter and Susan A. Mohrman (San Francisco: Jossey-Bass, 1994), 53–80.
44. Kirabo C. Jackson, "When Does School Autonomy Improve Student Outcomes?," Annenberg Institute at Brown University, EdWorkingPaper No. 23-808 (2023), <https://edworkingpapers.com/ai23-808>.
45. Jackson, "School Autonomy," 27.
46. Eleanor R. Odden and Priscilla Wohlstetter, "Making School-Based Management Work," *Educational Leadership* 52, no. 5 (1995): 32–36; Hill and Jochim, *Democratic Constitution*.
47. Hill and Jochim, *Democratic Constitution*.
48. In some rare cases, takeover and mayor involvement have been collaborative, as was the case in Newark, New Jersey, when then Republican governor Chris Christie and then Democratic mayor Cory Booker ran a turnaround operation in tandem. See Dale Russakoff, *The Prize: Who's in Charge of America's Schools?* (Boston: Houghton Mifflin Harcourt, 2015).
49. Moe, "Vested Interests."
50. See, for example, Paul Manna, "State Education Governance and Policy: Dynamic Challenges, Diverse Approaches, and New Frontiers," *Peabody Journal of Education* 87, no. 5 (2012): 627–43; Paul Manna, "Centralized Governance and Student Outcomes: Excellence, Equity, and Academic Achievement in the US States," *Policy Studies Journal* 41, no. 4 (2013): 682–705.
51. Kenneth K. Wong, Francis X. Shen, Dorothea Anagnostopoulos, and Stacey Rutledge, *The Education Mayor: Improving America's Schools* (Washington, DC: Georgetown University Press, 2007); Kenneth K. Wong and Francis X. Shen, "Mayoral Governance and Student Achievement: How Mayor-Led Districts Are Improving School and Student Performance," Center for American Progress, 2013; Kenneth K. Wong and Francis X. Shen, "Big City Mayors

and School Governance Reform: The Case of School District Takeover," *Peabody Journal of Education* 78, no. 1 (2003): 5-32.

52. See, for example, Thomas S. Dee and James Wyckoff, "Incentives, Selection, and Teacher Performance: Evidence from IMPACT," *Journal of Policy Analysis and Management* 34, no. 2 (2015): 267-97; Kristin Blagg and Matthew M. Chingos, "Does Gentrification Explain Rising Student Scores in Washington, DC?," *Urban Wire*, May 23, 2016; Thomas Dee and James Wyckoff, "A Lasting Impact: High-Stakes Teacher Evaluations Drive Student Success in Washington, DC," *Education Next* 17, no. 4 (2017): 58-67; Dallas Dotter, Duncan D. Chaplin, and Maria Bartlett, *Impacts of School Reforms in Washington, DC, on Student Achievement* (Princeton, NJ: Mathematica Incorporated, 2021).

53. George Parker and Michelle Rhee, "The Washington DC Teachers Union Contract: Negotiating for Change," October 6, 2022, <https://www.youtube.com/watch?v=w3zcmhdx6JA>.

54. See Douglas N. Harris and Matthew F. Larsen, "What Effect Did the New Orleans School Reforms Have on Student Achievement, High School Graduation, and College Outcomes?," Education Research Alliance Policy Brief, 2018; Douglas N. Harris, *Charter School City: What the End of Traditional Public Schools in New Orleans Means for American Education* (Chicago: University of Chicago Press, 2020).

55. Terry M. Moe, *The Politics of Institutional Reform: Katrina, Education, and the Second Face of Power* (Cambridge, UK: Cambridge University Press, 2019).

56. Schueler and Bleiberg, "Evaluating Education Governance."

57. See, for example, Domingo Morel, *Takeover: Race, Education, and American Democracy* (Oxford, UK: Oxford University Press, 2018).

58. Ashley Jochim, "Measures of Last Resort: Assessing Strategies for State-Initiated Turnarounds. Linking State and Local School Improvement," Center on Reinventing Public Education, 2016.

59. See, for example, Mark J. Chin, Thomas J. Kane, Whitney Kozakowski, Beth E. Schueler, and Douglas O. Staiger, "School District Reform in Newark: Within—and between—School Changes in Achievement Growth," *ILR Review* 72, no. 2 (2019): 323-54.

60. Molly F. Gordon, Marisa de la Torre, Jennifer R. Cowhy, Paul T. Moore, Lauren Sartain, and David Knight, "School Closings in Chicago: Staff and Student Experiences and Academic Outcomes. Research Report," University of Chicago Consortium on School Research, 2018.

61. See, for example, Mary L. Mason and Sarah Reckhow, "Rootless Reforms? State Takeovers and School Governance in Detroit and Memphis," *Peabody Journal of Education* 92, no. 1 (2017): 64-75; Ron Zimmer, Gary T. Henry, and Adam Kho, "The Effects of School Turnaround in Tennessee's Achievement School District and Innovation Zones," *Educational Evaluation and Policy Analysis* 39, no. 4 (2017): 670-96.

62. See, for example, Paul Hill, Lawrence C. Pierce, and James W. Guthrie, *Reinventing Public Education: How Contracting Can Transform America's Schools* (Chicago: University of Chicago Press, 1997); Paul T. Hill, "Put Learning First: A Portfolio Approach to Public Schools. PPI Policy Report," Progressive Policy Institute, 2006.

63. See, for example, Parker Baxter, Anna Nicotera, Erik Fuller, Jakob Panzer, Todd Ely, and Paul Teske, "The System-Level Effects of Denver's Portfolio District Strategy: Technical Report," Center for Education Policy Analysis, University of Colorado Denver School of Public Affairs, 2022; Osborne, *Reinventing America's Schools: Creating a 21st Century Education System*; Robin Lake, "The Hoosier Way: Good Choices for All in Indianapolis," *Education Next* 20, no. 2 (Spring 2020): 26-39.

64. Paul T. Hill and Ashley Jochim, "Education Reform That Sticks: Ten Years of 'Portfolio Strategy' Efforts Reveal How to Navigate Politics and Change K-12 Public Education," Center on Reinventing Public Education, 2022.

65. See, for example, Bulkley et al., *Challenging the One Best System*.

66. Vladimir Kogan, "Are School Boards Democratic? Session 4: Should School Boards Run Schools?," virtual conference series on school governance, Program on Education Policy and Governance (PEPG) at Harvard University, May 25, 2023, <https://www.youtube.com/watch?v=yObO7EtQBdM>.
67. See, for example, David Jaume and Alexander Willén, "The Long-Run Effects of Teacher Strikes: Evidence from Argentina," *Journal of Labor Economics* 37, no. 4 (2019): 1097-139.
68. Kogan, "Are School Boards Democratic?"
69. Jennifer Pahlka, "Culture Eats Policy," Niskanen Center (blog), June 21, 2023, <https://www.niskanencenter.org/culture-eats-policy>.
70. Parker Baxter and Alan Gottlieb, "Dismantling Denver: The City Was a National Model for Education Reform. Then Union-Backed Candidates Took over the School Board," *Education Next* 22, no. 2 (2022): 26-36.
71. Beth E. Schueler and Martin R. West, "Federalism, Race, and the Politics of Turnaround: US Public Opinion on Improving Low-Performing Schools and Districts," *Educational Researcher* 51, no. 2 (2022): 122-33.
72. Richard Whitmire, *The Bee Eater: Michelle Rhee Takes On the Nation's Worst School District* (Hoboken, NJ: John Wiley & Sons, 2011).
73. See, for example, Russakoff, *The Prize*.
74. See, for example, J. Celeste Lay and Anna Bauman, "Private Governance of Public Schools: Representation, Priorities, and Compliance in New Orleans Charter School Boards," *Urban Affairs Review* 55, no. 4 (2019): 1006-34.
75. Hill and Jochim, "Education Reform That Sticks."
76. J. Celeste Lay, *Public Schools, Private Governance: Education Reform and Democracy in New Orleans* (Philadelphia: Temple University Press, 2022).
77. Michael T. Hartney, *Revitalizing Local Democracy: The Case for On-Cycle Local Elections* (New York: Manhattan Institute for Policy Research, October 2021).
78. Vladimir Kogan, Stéphane Lavertu, and Zachary Peskowitz, "Election Timing, Electorate Composition, and Policy Outcomes: Evidence from School Districts," *American Journal of Political Science* 62, no. 3 (2018): 637-51; Zoltan L. Hajnal, Vladimir Kogan, and G. Agustin Markarian, "Who Votes: City Election Timing and Voter Composition," *American Political Science Review* (2021): 1-10.
79. Julia A. Payson, "Test Scores and School Boards: Why Election Timing Matters," Brookings Institution: Brown Center Chalkboard (blog), March 22, 2017, <https://www.brookings.edu/blog/brown-center-chalkboard/2017/03/22/test-scores-and-school-boards-why-election-timing-matters>.
80. A. J. Crabill, *Great on Their Behalf: Why School Boards Fail, How Yours Can Become Effective* (Austin, TX: Lioncrest, 2023).
81. See, for example, Chad Aldeman, "State Test Results Must Be Released More Quickly to Benefit Kids," NWEA: The Education Blog (blog), May 12, 2023, <https://www.nwea.org/blog/2023/state-test-results-must-be-released-more-quickly-to-benefit-kids>; Kogan, "The Choice in Education Governance Debates: Complacency or Reform?"
82. Crabill suggests "current student outcomes data being mandated to appear on school board ballots next to the incumbent's name." Crabill, *Great on Their Behalf*, 57.
83. See UNESCO, "Education: From School Closure to Recovery" (accessed September 21, 2023), <https://www.unesco.org/en/covid-19/education-response>.
84. Crabill, *Great on Their Behalf*, 57.
85. Hill and Jochim, *Democratic Constitution*, 2.

86. Sean F. Reardon and Rebecca Hinze-Pifer, "Test Score Growth among Chicago Public School Students, 2009-2014," Stanford Center for Education Policy Analysis, 2017.
87. Samantha Smylie and Kalyn Belsha, "Chicago Will Have the Largest Elected School Board of Any Major US City," *Chalkbeat*, July 30, 2021, <https://chicago.chalkbeat.org/2021/7/30/22602068/illinois-governor-approves-elected-chicago-school-board>.
88. See Eric A. Hanushek, Jin Luo, Andrew J. Morgan, Minh Nguyen, Ben Ost, Steven G. Rivkin, and Ayman Shakeel, "The Effects of Comprehensive Educator Evaluation and Pay Reform on Achievement," National Bureau of Economic Research, Working Paper no. 31073, March 2023.



10. School Choice Policies in the United States

Retrospect and Prospect

John D. Singleton

Executive Summary

The adoption of school choice policies over the last several decades was a major K-12 education policy development in the United States. These policies take several different forms, including public school choice (e.g., “open enrollment”), charter schools, and voucher programs that defray or reduce the cost of private school tuition. This chapter surveys the present landscape of school choice policies in the United States. While vouchers have attracted renewed policy interest recently, more than ten times as many students attended a charter school in fall 2019 as used a voucher to attend a private school. The chapter then assesses what global experience so far suggests about the potential of school choice policies for reforming public education in the United States.

It is emphasized that this potential rests in large part on the degree to which the policies improve education in traditional public schools. The survey of the empirical evidence to date supports the view that school choice policies usually—but not necessarily—have positive spillover effects and that they do so by causing public schools to raise their quality in the face of competition. At the same time, there are important gaps in knowledge concerning impacts of school choice on students’ long-run success and how the design of choice policies contributes to those impacts. For a post-COVID future that is likely to feature greater school choice, several policy takeaways are emphasized. These include leveraging incentives that stand to intensify competition, such as providing public funding for transportation to non-public schools; producing and supplying families with information about schools and their impacts; and targeting vouchers to students underserved by public education.

- School choice policies are a relatively new development in the United States, with charter schools in particular affecting large numbers of pupils.

- The most important element of school choice is the extent to which it improves education in traditional public schools.
- Shaping school choice policies is especially important as educators adjust to a post-COVID world.



The past forty years witnessed several significant initiatives in K-12 education policy in the United States. These include school finance reforms to ensure equity and adequacy of funding as well as the adoption of school accountability policies (exemplified by the No Child Left Behind Act) linking measures of student success with sanctions and rewards for teachers and schools. A major development was likewise school choice: public policies that provide families with increased access to schooling options other than a neighborhood-assigned public school.

Although school choice policies in the United States take several forms in practice—this chapter surveys findings from public school choice (e.g., “open enrollment”), charter school, and private school voucher programs—the school choice movement is unified around a common logic and set of policy goals. Its intellectual heritage typically is traced to economist Milton Friedman’s argument that, while government rightly should fund primary and secondary education, a robust system where substantially more private providers also produce and deliver that education and where the funding a school (public or private) receives is directly linked to its enrollment could be more efficient and more equitable (Friedman 1955). As with the debates over school finance and school accountability, a central issue is the low productivity of public schools in the United States in comparison with other nations. School choice policies distinctly diagnose this problem as a symptom of monopoly power. Poorly performing public schools are the logical product of weak incentives to supply a high-quality education in this view. Introducing choice—and therefore competition—to K-12 education follows as a remedy.

This broadly economic rationale connects with two other motivations for the introduction and expansion of school choice policies. The first can be described as a pragmatic appeal that has historically proven important: given the entrenched power of the public education bureaucracy in the United States (e.g., the political influence of public teachers’ unions), reform from without—as opposed to from within—has often proved more feasible. In other words, rather than attempting to reform public education from the “top down,” school choice policies advance reform in a more decentralized, “bottom-up” fashion. The other motivation recognizes that the students who are most underserved by public education in the United States are disproportionately disadvantaged. In particular, it is families of low socioeconomic status who otherwise effectively lack school choice; they often do not have the resources to pay for private school tuition or to move to affluent suburban neighborhoods with good public schools. From this perspective, school choice policies stand to level the playing field for

low-income and minority students both by improving access to high-quality schools and by spurring improvement of public schools.

This chapter surveys the present landscape of school choice policies in the United States and assesses what global experience so far suggests about their potential for reforming public education. This experience includes both several decades of state- and district-level choice policies in the United States as well as the experience of countries where choice has been introduced (e.g., Sweden and India) or has long been a feature of the national education system (e.g., Chile). A major point of emphasis is that the potential of school choice policies as a large-scale education reform rests in large part on the degree to which they improve the quality of education in traditional public schools. Such learning spillovers are likely to depend on how school choice programs are designed and may or may not be positive.

The chapter then summarizes key elements of the major school choice policies present across the United States, with a focus on the incentives that funding and accountability provisions may or may not create. These policies include public school choice options, such as magnet schools and open enrollment policies, as well as charter schools, which are operated independently but are publicly funded and tuition-free. While vouchers (and voucher-like programs) that defray or reduce the cost of tuition at private schools have attracted renewed policy interest recently, a much greater share of current public funding for school choice flows to charter schools: more than ten times as many students attended a charter school in fall 2019 as used a voucher to attend a private school. Charter schools educated around 6 percent of all students. Survey numbers also suggest that around as many students “chose” their public school (as opposed to attending an assigned one).

This leads into a critical appraisal of the empirical literature to date. The review focuses primarily on the evidence regarding the effects of school choice programs on public school students. It is argued that a consensus, built on the findings from both domestic and international programs, supports the view that school choice policies usually—but not necessarily—have positive spillover effects and that they do so by causing public schools to raise their quality in the face of competition. The chapter then highlights several important gaps in current knowledge. These include the need to understand the sources and significance of heterogeneity in impacts across different settings and programs and for greater evidence of effects on long-run student success (such as education attainment). There is also a need for work that considers the implications of the growing reality in the United States where several different school choice policies can be present at the same time.

The last sections of the chapter reflect on takeaways for school choice policy moving forward and on challenges of evidence-based policymaking in a post-COVID 19 future that is likely to feature greater school choice. One conclusion is that policymakers should consider and leverage incentives that stand to intensify competition. Specific applications include providing public funding for transportation to nonpublic schools; producing and supplying families with information about schools and their impacts; and targeting vouchers to those students who are underserved by public education.

ORIENTING THE FOCUS TOWARD EDUCATION QUALITY IN THE AGGREGATE

How should school choice policies be evaluated as successful or not? The perspective adopted in this chapter is that their impact on the quality of education, understood as the level of skills that students acquire in schools, should be the foremost focus. An important point is that this focus accommodates a broad conception of skills that encompass cognitive and other (noncognitive) aptitudes; it is not reducible to academic achievement. While measurement raises crucial questions returned to later, skill production in schools as a central focus has two advantages. The first is that it serves to summarize in terms of student outcomes the multitude of individual elements that constitute a schooling environment—for example, the quality of the teachers, the culture and leadership, the curriculum, and so on. The second advantage is from the perspective of policy: it provides a common basis for comparing the effectiveness of school choice policies with other major K-12 education reforms.

One way the question is often framed of whether school choice policies are successful or not is from the perspective of parents and caregivers: “Will sending my student to a charter school be better than sending them to their assigned public schools? Will using a voucher program to send my student to a private school be better than sending them to their assigned public schools?” This framing connects with a mature empirical literature and body of evidence that asks whether charter and private school alternatives provide higher education quality than traditional public schools. A fundamental empirical challenge for research is self-selection—the fact that the students who choose to use a voucher or switch to a charter school are not randomly drawn from the population—and scholarly work has advanced several approaches to address this. Important findings in the literature include that certain charter schools (many aligned with “No Excuses” practices) generate dramatic improvements in students’ academic achievement and whether they go on to college (e.g., Angrist et al. 2016; Dobbie and Fryer 2020); that recent national data sources show learning gains of charter school students exceed those of traditional public school students (Shakeel and Peterson 2021; CREDO 2023); and that voucher recipients in statewide programs in several US states perform lower than comparable traditional public school students on state standardized tests (e.g., Mills and Wolf 2017; Abdulkadiroğlu et al. 2018).

From a policy perspective, however, some of the attention to whether school choice alternatives are higher quality is misplaced. This is because the promise of school choice policies as large-scale education reforms rests in a fundamental way on how those policies impact students who choose to attend traditional public schools. In theory, school choice policies can increase education quality in the aggregate by creating incentives for public schools to raise their productivity. Analogized to a “tide that lifts all boats” (Hoxby 2003), this competitive response prospectively redounds to the benefit of students who remain in public schools. The key point, emphasized in the next section’s survey of the school choice landscape in the United States, is that those students who participate in choice programs (i.e., use a voucher program to attend a private school or choose to attend a charter school) are a minority of all students. As a consequence, even small spillover effects on the comparatively larger share of

students who remain in public school will matter more in the aggregate. A numerical example helps illustrate this point: if 10 percent of students used a voucher to attend a private school and their learning was consequently advanced by six weeks each, education quality on average would be negligibly improved if there was zero impact on the 90 percent of public school students. While the voucher program in this example might pass a cost-benefit analysis, the point is that it would be difficult to justify the attention of policymakers concerned with developing and implementing policies that address the systemic failures of K-12 public education in the United States.

A growing literature, drawing on evidence from programs in the United States as well as international findings, tests for competitive effects of school choice policies. School choice policies can have negative spillovers, too, however, and a central empirical question concerns effects on resources and student composition in public schools and separating those effects from changes in public school productivity. A major empirical challenge for this branch of the literature is that the public schools that are exposed to competition are likely not random. The review of the empirical literature below summarizes the major findings about spillovers of school choice policies.

THE SCHOOL CHOICE POLICY LANDSCAPE IN THE UNITED STATES

There are various policies that create or expand access to school options across the United States. This chapter provides a brief overview of three broad categories: (1) policies that provide choice among public schools (e.g., open enrollment and magnet programs); (2) charter school laws that allow for the creation and operation of publicly funded and tuition-free schools by independent organizations; and (3) voucher and voucher-like programs that defray the cost of tuition at private schools for qualified students.¹

PUBLIC SCHOOL CHOICE

Public school choice policies offer families choice among public schools. One example is magnet schools, which are typically district run, feature specialized curricula, and often have selective admissions. Students must apply to be considered for admission. Examples include “exam” schools such as Stuyvesant in New York City, Thomas Jefferson High School for Science and Technology in Fairfax, Virginia, and many schools of the arts around the country. Another example of public school choice is open enrollment policies intended to provide a process whereby students can transfer between public schools. Most states allow school districts to voluntarily accept students from other districts, and twenty-three have policies in place that make such interdistrict transfers mandatory (Education Commission of the States 2017). Transfers within a school district are mandatory, subject to limitations, in nineteen states.

Choice mechanisms are a kind of open enrollment policy that facilitates within-district public school choice. Such mechanisms allow students to submit ranked lists of school preferences,

which are then aggregated (and can be combined with preferences identified by schools and specialized programs and other criteria) to assign students to schools via an algorithm. An example is Denver’s SchoolChoice process, first put in place in 2012, where families (most with students in either kindergarten, sixth, or ninth grade) rank up to twelve schools, and matching priorities include neighborhood zones and siblings. School districts often pair choice mechanisms with guides for families to understanding the ranking and allocation process and that detail aspects and attributes of the various schools and programs available. While public school choice mechanisms expand choice options beyond charter or private schools—and introducing choice into the system may be a response to the availability of charter or voucher options—there is greater ambiguity about the theoretical system-level implications of magnet, specialized programs, or open enrollment policies on education quality. This is because choice *between* public schools does not obviously create meaningful competitive incentives: whichever public school a student chooses to attend, they (and the revenue attached to them) ultimately stay in the public system.

It is difficult to provide an exact number as to how many students in the United States “chose” their public school via a policy option (as opposed to being assigned it based on residence). Pre-2020 survey numbers from the National Center for Education Statistics (NCES) imply a figure of around three million students (about 6 percent of all public and private enrollment) (Hanson and Pugliese 2020). A number of large public school districts have at least limited school choice mechanisms in place, including New York City, Los Angeles, Indianapolis, and Charlotte-Mecklenburg (in North Carolina). These programs can take different forms. In some cases, magnet program and charter school options are integrated into the same choice mechanism as for traditional public schools.

CHARTER SCHOOLS

Charter schools, in contrast, are essentially privately operated public schools. Charter schools are supported by tax revenues, are regulated by government entities, do not charge tuition, and cannot have selective admission criteria. They also participate in school accountability programs (i.e., standardized testing) and can be closed by public school authorities for poor performance. At the same time, they are created, operated, and managed by independent organizations and, like private schools, have considerable autonomy when it comes to decisions over curricula, human resources, and where to locate. Charter school operators are generally nonprofit organizations (well-known examples include the Knowledge is Power Program, Rocketship, and Success Academy), though for-profit management companies operate in several states.

How (and at what level) charter schools are funded has long been a major point of debate. This is in part because most per-pupil funding for charter school students moves out of the public school system to the charter school the student enrolls in. Thus, the fiscal implications of school choice can be immediate for public schools in the case of charter school competition. From an economic perspective, this means that the presence of charter schools in principle creates a meaningful incentive for districts to retain students (though public school leaders may not act on those incentives). In several states, however, these incentives are

mutated by “hold harmless” provisions that at least temporarily offset the full financial impact of enrollment lost to charter schools. Public school leadership, at both district and state levels, has at times been hostile to charter school growth in the face of fiscal impacts, and charter schools are a frequent target of public school teachers’ unions. At the same time, the amount of funding is also typically disparate—charter schools generally receive less revenue per pupil than is spent on a student in a traditional public school—and this disparity has itself been (and continues to be) a focus of policy attention.

About 3.4 million students in the United States attended a charter school during 2019–20. This equates to more than 6 percent of all school enrollment and reflects steady growth; charter schools enrolled about 4 percent of students in 2010–11. The average charter school student is much less likely to be White and more likely to be economically disadvantaged than the average public school student (NCES 2022).² Although all but four states have charter school laws in place, charter school penetration varies considerably across places. In several states, more than one in ten non-private school attendees attended a charter school in 2019, including Washington, DC (45 percent), Arizona (18 percent), Colorado (14 percent), Louisiana (12 percent), and Florida (11 percent) (NCES 2022). Note that these numbers mask major within-state heterogeneity: for example, about 25 percent of Miami-Dade non-private school students attend a charter school. In New Orleans, virtually all students who do not attend private schools attend charter schools.

VOUCHERS

The last variety of school choice programs highlighted in brief here are those that defray the cost of tuition at private schools for eligible students. These programs are generically referred to as “voucher” policies but can take several different forms across the United States. These forms include prototypical government-funded programs that cut families a check; tax credit scholarships—which deputize nonprofits to receive tax-advantaged contributions and provide scholarships to students; and education savings accounts. Education savings accounts instead allow families to allocate a given sum of government funds across education expenses, including private school tuition.

Voucher programs have historically been means tested, whereby eligibility is restricted to students in families whose income does not exceed some threshold. A number of programs are targeted specifically to students with disabilities, and a few restrict or expand eligibility to students assigned to persistently low-performing public schools. Similar to the revenue disparity between traditional public schools and charter schools, effective voucher amounts are typically well below per-pupil expenditures in public schools and often do not cover the average sticker price at private schools. For example, Ohio’s means-tested statewide vouchers are worth \$5,500 at the elementary and middle school levels for 2022–23. Other policy parameters concern accountability provisions of voucher policies. At issue are typically three aspects of the programs: (1) whether participating private schools may apply their own admission criteria or not (with seats in oversubscribed private schools allocated by lottery); (2) whether private schools can require supplemental tuition from voucher recipients; and (3) whether voucher recipients must take the same achievement exams as in public schools.

While the country’s first voucher program (which began in Milwaukee in 1990 and served nearly thirty thousand students in 2022–23) requires lotteries and allows tuition supplements only at high schools, almost all major voucher programs instituted in the United States in the past twenty years relax these provisions.

In fall 2019, about 4.7 million K–12 students were enrolled in a private school nationally (NCES 2022). This level, which is a little less than 10 percent of all students, has been relatively constant over recent years. However, the estimated share of private school enrollees who are part of a voucher or scholarship program has grown. More than five hundred thousand students (about 1 percent of all public and private school students) attended a private school via a school choice policy in 2019—more than double the number ten years prior (EdChoice 2023). As of 2023, thirty US states (plus Washington, DC) had some kind of voucher program in place. The country’s largest voucher program, Florida’s Tax Credit Scholarship, began in 2001 and served more than one hundred thousand students as recently as 2020–21. While private school students as a whole tend to be more advantaged (and less likely to belong to an underrepresented minority) than public school students, those students who use vouchers—reflecting in part the income-based eligibility criteria for such programs—are generally less advantaged and relatively more minority (Wolf 2020).

WHAT DO WE KNOW?

What do we know about the externalities of school choice policies on the learning of students who nonetheless attend a traditional public school? This section considers the evidence on this question. The topic of spillovers juxtaposes countervailing forces. On the one hand, incentives to retain enrollments may lead public schools to respond to school choice policies by raising quality, as intended. On the other hand, expanded school choice may negatively impact students who remain in public schools through a resource or peer effect channel. If, for example, more advantaged or higher-achieving students leave public schools (and exposure to such students benefits their peers), the learning of students left behind will suffer. There is also an important concern that low information about school quality (or preferences for other attributes of schools) could incentivize responses by public schools that are unrelated to impacts on learning.

The body of rigorous evidence to date supports the conclusion that spillover effects on student learning are usually—but not necessarily—positive on net. The pieces of evidence supporting this are twofold. First, the net effect on students remaining in an assigned traditional district school is generally zero or positive; this implies, and more direct evidence affirms, that negative externalities from peer sorting are limited.³ Second, the patterns for net improvements in student outcomes are consistent with increased competition causing public schools to raise their productivity.

A recent review of the findings from private school choice programs in the United States highlights that most studies find positive evidence of competitive effects and that no study finds evidence of negative effects on students attending traditional public schools (Wolf 2020). Figlio and Karbownik (2016), to highlight one example, compare two sets of Ohio public school students: (1) those in traditional public schools who marginally exceeded a score threshold (based on standardized test performance) below which all students would become eligible to receive a voucher; and (2) those in public schools who marginally failed the score threshold. The latter group, who were exposed to private school competition but should otherwise be on average the same as the first group, showed greater test score growth. The body of evidence on spillovers of charter schools on public school students is more ambiguous. Older findings drawn from a variety of states suggest that test score impacts are limited.⁴ One important study, by Imberman (2011), reports negative impacts from charter schools on the achievement of students in traditional public schools. The more recent evidence is generally more positive. Studies that find improvements in public school students' test scores include Cordes (2018) for New York City, Gilraine et al. (2021) for North Carolina, and Ridley and Terrier (2022) for Massachusetts.⁵ The New York City study compares student test scores between traditional public schools exposed to charter school competition at different times and at different distances. These studies of US programs are accompanied by results showing increases in public school student outcomes following the introduction of vouchers in 1992 in Sweden (Sandström and Bergström 2005) and, despite increases in stratification that likely reduced average student academic potential in public schools, no changes associated with entry of private schools in India (Bagde et al. 2022).

Two kinds of evidence additionally support the claim that increased competition can cause public schools to raise their quality. The first kind is from settings where competition increases but in which other responses cannot arise. Figlio and Hart (2014) is a foremost example: students could apply for Florida's means-tested voucher one year in advance of the program launch. The authors present evidence of increases in test scores at competitively exposed public schools in even this "pure competition" year where students could not yet re-sort. Implementing the same idea, Gilraine et al. (2021) provide evidence for competition in the context of charter schools in North Carolina. The second kind of evidence is drawn from models of school choice that generate predictions for how much competitive pressure different public schools experience. Gilraine et al. (2023), for example, show (1) that demand for charter schools offering a "nontraditional" curriculum is not very sensitive to the quality of public schools; and (2) that, as would be expected, the quality of public schools does not increase following nearby entry by a "nontraditional" charter school. They further show that the test score value added of public schools does increase following entry of math and reading skills-focused charter schools. Campos and Kearns (2023) take a similar approach to argue that competition among public schools is the mechanism supporting improvement in student outcomes, including high school completion, in Los Angeles's high school Zones of Choice. Card et al. (2010) for Ontario, California, and Neilson (2021) for Chile, respectively, are two examples that bring international evidence of this kind to bear.

WHAT DO WE NOT YET KNOW THAT WE NEED TO KNOW?

While the evidence summarized in the previous section points to emerging consensus on the spillover effects of school choice policies, this section highlights several gaps in current knowledge. These gaps include (1) the paucity of evidence on long-run impacts, such as on wages; (2) the little attention to implications of how choice programs are designed; and (3) the need to consider markets (which may feature several overlapping choice options and programs) as the appropriate units of analysis, not individual programs and policies.

First, a major limitation of the existing evidence on the aggregate effects of school choice policies is the limited amount of work that speaks to impacts on students' "long-run" outcomes. These outcomes include educational attainment (such as college entrance or graduation), marriage, employment, work history, and labor market earnings. This is important because long-run indicators of success are reliably better measures of human capital acquisition than test scores, which are more widely used due to their greater availability in US datasets. Long-run markers of success, for example, depend on many kinds of skills, while test scores may only reliably measure cognitive skills acquisition. A large literature recognizes that near-term impacts on math and reading scores or measures derived from them may not capture durable (and multidimensional) skill gains. An exception in the literature is studies that examine effects on postsecondary outcomes of charter school attendance. But the lack of long-term outcome data is especially relevant to the evidence on statewide voucher programs, where negative effects on test scores for voucher recipients may conflate lower education quality with private schools' nonalignment with the public school curriculum. Further, little to no existing work estimates competitive impacts from school choice policies on measures of student success in the long run.⁶

A second gap in the existing evidence concerns how impacts on education quality relate to how choice programs are designed and implemented. This is because how policy elements combine has implications for the incentives facing public school leaders and thus for the potential of choice to generate competition. A simple example in illustration is whether (and how much of) public funds actually follow students who switch from public schools to a school choice option like a private school via voucher or a charter school. However, other elements of choice programs are also relevant. These include whether funding and policy support a competitive threat to public schools. For example, are negative impacts of statewide voucher programs on participants partly due to high-quality private schools being insufficiently incentivized to participate? Should barriers to charter school entry be kept low, or is it better that authorizers screen applications to open new schools or expand existing schools based on proven success and limit where new charter schools can open? Answers will require more work that tackles how schools—public, private, and charters alike—make decisions. Though they have clear implications for policy, little existing evidence in the literature speaks to these questions.

This relates to a third limitation of the existing stock of knowledge, which concerns its applicability in the current policymaking environment: increasingly, K-12 education "markets" feature multiple school choice programs. It is no longer unusual for a voucher program that defrays private school tuition, several charter schools, and some kind of choice among public schools to all be options available to families in a district. One issue this creates is

interpreting findings from individual choice programs: a case in point is that recent data from the Washington, DC, voucher program shows that 42 percent of the control group students attended charter schools (Dynarski et al. 2017). The bigger question raised, however, concerns how combinations of school choice programs interact at scale. How should limited resources be allocated across schools and programs? Is the marginal dollar better spent on expanding charter schooling, or on expanding vouchers, or on implementing better public school choice mechanisms, or on increasing public school quality generally? This requires a shift in focus from school choice options and policies in isolation to “education markets” as they exist and evolve.

WHAT SHOULD BE DONE?

This section collects several takeaways from the global experience with school choice programs to date. A first recommendation is that policymakers should recognize the central importance—and leverage the power—of incentives. The simple point is that, for choice policies to create meaningful competition, traditional public schools and districts must feel threatened with losing students and funding. One practical application of this recommendation is to transportation: a compelling argument that education funding should be directed toward providing transportation of students to charter schools or, in the case of voucher programs, to private schools is that doing so will make it easier for students to potentially leave traditional public schools; it will increase competition. A second application is providing parents with better information about schools (public, private, and charter). A large body of evidence indicates that parents may not know about available choice options and do not have accurate information about school quality. The implication is that supplying information to parents about the effectiveness of different schools at advancing student learning will strengthen incentives facing traditional public schools to increase their quality.

Given limited resources, policy decisions would additionally benefit from considering where returns are likely to be highest. This recommendation channels an animating motive for school choice programs: expanding choice for low-income families and students will yield greater improvements in education quality overall than will expanding choice for high-income families, who already have the means to choose private schools or public schools in affluent neighborhoods. But combined with the logic of incentives, this recommendation produces additional insight: because high-income families experience little real change in their ability to exercise choice, the schools under their consideration also experience limited changes in incentives. This stands in marked contrast for low-performing schools that disproportionately serve low-income students: continuing to miss the mark, when effective choice programs are in place, will risk losing enrollments. This observation has several applications. One is that policymakers could provide financial incentives to charter schools to locate in neighborhoods and areas where traditional public schools underserve students. Several pieces of evidence show that financial considerations are important for where charter schools open (and whether they survive). Long-standing eligibility criteria for voucher programs that are based on the performance of the student’s local public school or family income carry a similar logic. This

implies that making vouchers and voucher-like programs universal, while perhaps appealing on fairness grounds, also may reduce their overall effectiveness. An alternative would be to scale the generosity of the voucher with family income. Policy decisions about charter school authorization (and funding) could also benefit from considering which providers or operators generate greater competitive externalities. It stands to reason that high-quality or “proven” providers likely have an advantage in this regard, and evidence, mentioned earlier, suggests that core skills-focused charter schools compete more closely with traditional public schools.

A final recommendation is just that policymakers value the potential of data and evidence to inform policy development and implementation. For example, there is no technical obstacle to creating measures of school (and teacher) quality that recognize multidimensional, durable skill development and reflect long-run success. Such measures could be built on rich evidence from test scores across many subjects, from GPAs, from attendance and discipline records, from college entry and completion data, and even from earnings in the labor market. These measures could then be used to inform parents so that they can make better decisions about schools (publicly available information about the quality of private schools is especially scarce), to target policies in effective ways, and to allocate resources. Rather, the obstacles are tragically often bureaucratic.

LOOKING FORWARD

The COVID-19 global pandemic and the resulting school closures and shift in education delivery models will likely prove to be the largest single “shock” to school choice in the United States to date. This shock has a demand side and a supply side. On the demand side, widespread dissatisfaction with the responses of public school leaders to the pandemic—in particular, delays in reopening schools—drove many families to seek choice alternatives. Private and charter schools were generally quicker to return to in-person instruction. This dissatisfaction is reflected in data on enrollments, which shows declines in public school enrollment (particularly among kindergartners) that were especially pronounced in districts that were slow to open buildings to in-person instruction (Dee et al. 2021). Public school enrollment declined nationally by 4 percent between fall 2019 and fall 2020, whereas charter school enrollments over the same time period increased by 7 percent (NCES 2023).⁷ Rates of students attending private schools and being homeschooled also increased (Musaddiq et al. 2022).

The durability of this increase in demand for choice—particularly as new cohorts of students come of age—will combine with policy developments on the supply side that likewise portend a future with more school choice. In the wake of the COVID-19 pandemic, policy momentum behind expanding private school voucher programs has rapidly gained steam across numerous states. An example is 2023 legislation that extended the eligibility for Florida’s Tax Credit Scholarship, mentioned earlier, to all Florida students. Previously, eligibility was means tested (though vouchers remain prioritized to lower-income students under the new legislation) and was restricted to students either attending a public school or entering kindergarten or first grade. Florida is joined by Arizona, Iowa, North Carolina, South Carolina, Utah, and

West Virginia in creating new or expanding existing private school voucher programs post-pandemic. Like the Florida program, several of these will be universal in student eligibility, marking a significant departure from earlier voucher programs in the United States.

The fresh momentum for vouchers highlights a fundamental challenge of evidence-based policymaking: research findings typically lag policy. This is brought into relief by the fact that the recent turn to vouchers arrives amid important new findings pointing to successes from charter school choice.⁸ Moving forward, the post-pandemic policy developments raise at least two questions for policymakers and researchers alike. The first is, for effective school choice policies, how important are accountability provisions—as opposed to accountability to the “market” alone? As detailed earlier, charter schools are generally subject to accountability provisions that are typically absent for private schools in voucher programs. The second question is how to produce rigorous evidence about new voucher (and voucher-like) programs as they grow. Many existing data systems do not contain reliable information about private schools (or private school teachers), may not record or track recipients of vouchers (especially those who do not at some point have contact with the public school system), and—in states where administration of state tests is not required—will not readily contain information suitable for evaluation. Producing rigorous research findings in the post-pandemic world will thus require renewed cooperation to develop and make available high-quality data as well as creativity to identify and draw from other information sources.

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](https://www.hoover.org/hesi).

The conclusion that choice between public schools does not create meaningful competitive incentives may be true for intradistrict choice programs, but Arkansas public schools are highly incentivized through the interdistrict choice law. Because of this school choice law and the school funding formula being based on average daily membership in which the funding “follows the student,” many districts successfully advertise their performance levels, programs, and extracurricular options to attract students from other districts.

The recommendation that policymakers value the potential of data and evidence strikes a chord, as the future success of choice policies depends on the consideration of data and evidence beyond state assessments. The absence of an assessment and accountability requirement for nonpublic schools will forever be a hurdle for policymakers looking to expand choice options. This challenge could be mitigated by requiring other measures, such as an assessment of higher-order skills that are indicators of future success.

—Johnny Key, former commissioner of the Arkansas Department of Education

In the last two years, ten states have passed universal or nearly universal educational choice programs, the vast majority of which are education savings accounts (ESAs). If vouchers are like mobile phones, education savings accounts are like the smartphones of education choice. Families can use the accounts for multiple educational needs, including tuition, fees, therapy, tutoring, digital devices, instructional materials, uniforms, and even college savings plans. ESAs put families in the driver's seat, not just for choosing educational options that best meet their child's needs but also for customizing those options and evaluating alternatives based upon how much they want to spend. As policymakers continue to pursue these policies to empower families, they should consider several items: program design that does not suppress private school, microschool, or other innovative education service supply; transportation supports to ensure real choice that empowers families to get their child to the school they desire; and sufficient weighted financial support for low-income students and those with special needs.

—Patricia Levesque, CEO, Foundation for Excellence in Education

REFERENCES

- Abdulkadiroğlu, Atila, Parag A. Pathak, and Christopher R. Walters. 2018. "Free to Choose: Can School Choice Reduce Student Achievement?" *American Economic Journal: Applied Economics* 10, no. 1: 175-206.
- Angrist, Joshua D., Sarah R. Cohodes, Susan M. Dynarski, Parag A. Pathak, and Christopher R. Walters. 2016. "Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice." *Journal of Labor Economics* 34, no. 2:275-318.
- Bagde, Surendrakumar, Dennis Epple, and Lowell Taylor. 2022. "The Emergence of Private High Schools in India: The Impact of Public-Private Competition on Public School Students." *Journal of Public Economics* 215: 104749.
- Bettinger, Eric P. 2005. "The Effect of Charter Schools on Charter Students and Public Schools." *Economics of Education Review* 24, no. 2:133-47.
- Bifulco, Robert, and Helen F. Ladd. 2006. "The Impacts of Charter Schools on Student Achievement: Evidence from North Carolina." *Education Finance and Policy* 1, no. 1:50-90.
- Booker, Kevin, Scott M. Gilpatric, Timothy Gronberg, and Dennis Jansen. 2008. "The Effect of Charter Schools on Traditional Public School Students in Texas: Are Children Who Stay Behind Left Behind?" *Journal of Urban Economics* 64, no. 1:123-45.
- Bruhn, Jesse, Scott Imberman, and Marcus Winters. 2022. "Regulatory Arbitrage in Teacher Hiring and Retention: Evidence from Massachusetts Charter Schools." *Journal of Public Economics* 215: 104750.
- Campos, Christopher, and Caitlin Kearns. 2023. "The Impact of Public School Choice: Evidence from Los Angeles' Zones of Choice." *Quarterly Journal of Economics*, forthcoming.
- Card, David, Martin D. Dooley, and A. Abigail Payne. 2010. "School Competition and Efficiency with Publicly Funded Catholic Schools." *American Economic Journal: Applied Economics* 2, no. 4:150-76.
- Cohodes, Sarah R., and Katharine S. Parham. 2021. "Charter Schools' Effectiveness, Mechanisms, and Competitive Influence." Working Paper No. w28477. National Bureau of Economic Research.
- Cohodes, Sarah R., Elizabeth M. Setren, and Christopher R. Walters. 2021. "Can Successful Schools Replicate? Scaling Up Boston's Charter School Sector." *American Economic Journal: Economic Policy* 13, no. 1:138-67.

- Cordes, Sarah A. 2018. "In Pursuit of the Common Good: The Spillover Effects of Charter Schools on Public School Students in New York City." *Education Finance and Policy* 13, no. 4:484–512.
- CREDO (Center for Research on Education Outcomes). 2023. *As a Matter of Fact: The National Charter School Study III*, Center for Research on Education Outcomes.
- Dee, Thomas S., Elizabeth Huffaker, Cheryl Phillips, and Eric Sagara. 2021. "The Revealed Preferences for School Reopening: Evidence from Public-School Disenrollment." *American Educational Research Journal* 60, no. 5: 916–40.
- Dobbie, Will, and Roland G. Fryer. 2020. "Charter Schools and Labor Market Outcomes." *Journal of Labor Economics* 38, no. 4:915–57.
- Dynarski, Mark, Ning Rui, Ann Webber, and Babette Gutmann. 2017. "Evaluation of the DC Opportunity Scholarship Program: Impacts after One Year." Working Paper No. NCEE 2017-4022. National Center for Education Evaluation and Regional Assistance.
- EdChoice, 2023. *The ABCs of School Choice* (2023 Edition). 2023. Indianapolis, IN: EdChoice.
- Education Commission of the States. 2017. "Does the State Have Open Enrollment Programs?" October. <https://ecs.my.salesforce-sites.com/mbdata/MBQuestNB2n?rep=OE1701>.
- Figlio, David, and Cassandra M. D. Hart. 2014. "Competitive Effects of Means-Tested School Vouchers." *American Economic Journal: Applied Economics* 6, no. 1:133–56.
- Figlio, David N., Cassandra Hart, and Krzysztof Karbownik. 2022. "Effects of Maturing Private School Choice Programs on Public School Students." *American Economic Journal: Economic Policy*, forthcoming.
- Figlio, David, and Krzysztof Karbownik. 2016. "Evaluation of Ohio's EdChoice Scholarship Program: Selection, Competition, and Performance Effects." Thomas B. Fordham Institute.
- Friedman, Milton. 1955. "The Role of Government in Education." In *Economics and the Public Interest*, ed. Robert A. Solo, 123–44. New Brunswick, NJ: Rutgers University Press.
- Fryer, Roland G. Jr. 2014. "Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments." *Quarterly Journal of Economics* 129, no. 3:1355–407.
- Gilraine, Michael, Uros Petronijevic, and John D. Singleton. 2021. "Horizontal Differentiation and the Policy Effect of Charter Schools." *American Economic Journal: Economic Policy* 13, no. 3:239–76.
- . 2023. "School Choice, Competition, and Aggregate School Quality." Working Paper No. w31328. National Bureau of Economic Research.
- Hanson, Rachel, and Chris Pugliese. 2020. Parent and Family Involvement in Education: 2019 (NCES 2020-076). US Department of Education. Washington, DC: National Center for Education Statistics, <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2020076>.
- Hoxby, Caroline M. 2003. "School Choice and School Productivity. Could School Choice Be a Tide That Lifts All Boats?" In *Economics of School Choice*, 287–342. Chicago: University of Chicago Press.
- Imberman, Scott A. 2011. "The Effect of Charter Schools on Achievement and Behavior of Public School Students." *Journal of Public Economics* 95, no. 7–8:850–63.
- Mills, Jonathan N., and Patrick J. Wolf. 2017. "Vouchers in the Bayou: The Effects of the Louisiana Scholarship Program on Student Achievement after 2 Years." *Educational Evaluation and Policy Analysis* 39, no. 3:464–84.
- Muralidharan, Karthik, and Venkatesh Sundararaman. 2015. "The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India." *Quarterly Journal of Economics* 130, no. 3:1011–66.
- Musaddiq, Tareena, Kevin Stange, Andrew Bacher-Hicks, and Joshua Goodman. 2022. "The Pandemic's Effect on Demand for Public Schools, Homeschooling, and Private Schools." *Journal of Public Economics* 212:104710.

- NCES (National Center for Education Statistics). 2022. Private School Enrollment. Condition of Education. US Department of Education, Institute of Education Sciences. Last updated May 2022. <https://nces.ed.gov/programs/coe/indicator/cgc>.
- . 2023. Public Charter School Enrollment. Condition of Education. US Department of Education, Institute of Education Sciences. Last updated May 2023. <https://nces.ed.gov/programs/coe/indicator/cgb>.
- Neilson, Christopher. 2021. “Targeted Vouchers, Competition among Schools, and the Academic Achievement of Poor Students.” Working Paper.
- Ridley, Matthew, and Camille Terrier. 2022. “Fiscal and Education Spillovers from Charter School Expansion.” *Journal of Human Resources*, forthcoming.
- Sandström, F. Mikael, and Fredrik Bergström. 2005. “School Vouchers in Practice: Competition Will Not Hurt You.” *Journal of Public Economics* 89, no. 2-3:351-80.
- Sass, Tim R. 2006. “Charter Schools and Student Achievement in Florida.” *Education Finance and Policy* 1, no. 1:91-122.
- Sass, Tim R., Ron W. Zimmer, Brian P. Gill, and T. Kevin Booker. 2016. “Charter High Schools’ Effects on Long-Term Attainment and Earnings.” *Journal of Policy Analysis and Management* 35, no. 3:683-706.
- Shakeel, M. Danish, and Paul E. Peterson. 2021. “Charter Schools Show Steeper Upward Trend in Student Achievement Than District Schools.” *Education Next* 21, no. 1: 40-47.
- Slungaard Mumma, Kirsten. 2022. “The Effect of Charter School Openings on Traditional Public Schools in Massachusetts and North Carolina.” *American Economic Journal: Economic Policy* 14, no. 2:445-74.
- Wolf, Patrick J. 2020. “Private School Choice Programs: What Do We Know about Their Effects on Outcomes for Disadvantaged Students?” Background paper for the Hoover Education Success Initiative. https://www.hoover.org/sites/default/files/research/docs/wolf_webready.pdf.
- Zimmer, Ron, and Richard Buddin. 2009. “Is Charter School Competition in California Improving the Performance of Traditional Public Schools?” *Public Administration Review* 69, no. 5:831-45.

NOTES

1. The focus exclusively is on brick-and-mortar school options. Some states have created publicly run virtual schools, whereas in other states public virtual schools are typically a kind of charter school. Vouchers in several states can be used at private voucher schools.
2. For 2021-22, about 29 percent of charter school students were White, whereas nearly 47 percent of public school students were so classified. This is partly a reflection of the fact that charter schools generally do not open in low-density and more rural areas.
3. Recent compelling evidence concerning peer spillovers from school choice is drawn from India: Muralidharan and Sundararaman (2015) experimentally estimate the effect of private school voucher recipients exiting government schools on the students left behind. The paper does not find evidence that the test scores of government school students are lowered by voucher students leaving.
4. Estimated effects in the literature range from zero (e.g., Bettinger 2005; Bifulco and Ladd 2006; Zimmer and Buddin 2009) to positive (e.g., Sass 2006; Booker et al. 2008).
5. In-progress work from two research teams also finds positive effects in recent data from Florida. Slungaard Mumma (2022) finds no effect on test scores of public school students on average, using data from North Carolina and Massachusetts for a limited sample of charter school entries.
6. One exception, noted earlier (Campos and Kearns 2023), attributes to competition the effects of the public school choice program on education attainment in Los Angeles.

7. This compares with a 4 percent increase in charter school enrollment between fall 2018 and fall 2019.

8. These include the recent evidence on competitive impacts, summarized earlier, but also new work on the quality of charter schools. Cohodes et al. (2021), for example, present evidence that new charter schools of “proven providers” in Boston are as highly effective as the parent schools they are replicants of. The latest findings of CREDO’s national study indicate—for the first time—higher average annual learning gains for charter school students than for matched traditional public school students (CREDO 2023).



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.¹

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.² 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” frame³—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.⁴ 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.⁵ 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.⁶

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.⁷

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.”⁸ 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.⁹ 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.¹⁰ 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*.¹¹ 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).”¹²

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.¹³ 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.¹⁴

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.¹⁵ 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.¹⁶ 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.¹⁷

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."¹⁸

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.¹⁹ (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.²⁰) 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.²¹

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.^{22,23} 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.²⁴

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.²⁵ 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.²⁶

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.²⁷ To be sure, another set of voucher studies finds positive long-term impacts on measures such as high school graduation and college enrollment.²⁸ 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.²⁹ 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.³⁰

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.³¹

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.³²

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.

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; 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.
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.
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.
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; 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.

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.

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; 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.
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 Bachofer, “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/Crede-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 Abdulkadirglu, 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.
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; 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>.



12. Lessons from Newark

The Lineage of Modern School Reform and Where We Go Next

Cami Anderson

Executive Summary

Since the release of *A Nation at Risk* in 1983, the school reform movement has illuminated significant insights and promising practices for improving schools for children in poverty and students of color. The work of trying to radically increase student outcomes also produced glaring missteps and tough lessons. Few efforts demonstrate the complexity of attempting to provide a bold citywide plan to ensure educational excellence for all children better than the experiences in Newark, New Jersey. Much has been written about the political drama during my tenure as superintendent from 2011 to 2014. However, very little has been written about the actual playbook, results, and concrete implications for educational policymakers and leaders.

Starting with the announcement of Mark Zuckerberg’s unprecedented philanthropic gift on *The Oprah Winfrey Show*, in this chapter I walk through the early efforts to stabilize district schools, and I build up to unveiling the ambitious One Newark plan. This plan established a nation-leading universal enrollment system and fostered an environment in which both district and charter schools saw unprecedented academic growth. Along the way, I share explicit details of our efforts to build a “system of great schools” (not a great school system) for every Newark student. I provide details about how our bold approach built on promising practices and emerging research from around the country.

I share data on the real, material gains we achieved. I also reflect candidly on the enormous challenges we faced and own up to mistakes that may have prevented deeper and more permanent change. The lessons learned from that period in Newark provide a blueprint for systems leaders who seek to foster environments where a variety of school models thrive. These recommendations include (1) enabling successful schools that possess what I call the “four ingredients”; (2) ensuring quality and equity across the system; (3) breaking bureaucracy and cutting needless red tape to clear the runway for innovation; and (4) creating cross-system

and community-based solutions to address the failures at the intersections of social services and schools.

- Public education in Newark, New Jersey, took the spotlight in 2010 when Mark Zuckerberg donated \$100 million to help struggling schools.
- Even with bold reform efforts, however, ensuring educational excellence for all children is work left undone.
- One Newark’s successes and failures show where other school districts can build on Newark’s approach.

• • •

When *A Nation at Risk* was published in 1983, I was twelve years old. I was growing up in Los Angeles as one of three birth children in a diverse family of fourteen, and I had witnessed firsthand the unacceptable circumstances my adopted brothers and sisters experienced in a broken child welfare system. It was clear to me that our education system was similarly broken, and I could plainly tell that succeeding in school was much easier for those of us who held privilege.

As the child of two dedicated civil servants who are avowed Democrats, I never would have guessed that a report coming out of the Reagan administration would help spur a transformative movement that would tap into my activist, social justice upbringing and provide a path for me to play a small part in moving the needle on education equity.

We’ve come a long way since 1983. Over these many years, my friends and colleagues in the school reform movement and I have led hard-fought progress—as well as committed some glaring missteps. These years of experience, coupled with a deep and growing body of research, have yielded invaluable lessons for how we, as a nation, can come ever closer to delivering on the promise of education excellence for all kids.

Educational leaders have used these decades to explore an entire continuum from tweaking and refining existing systems to innovating and disrupting the status quo. The solutions are out there. The talent is out there. Even the money is out there. So what is holding us back?

In this chapter I will share reflections from our attempt to answer this question in Newark, where I served as superintendent from 2011 to 2015. Building on the lessons and experience from nearly twenty years of reform, we launched a number of ambitious but commonsense reforms, culminating in the “One Newark” plan. Our goal was to engage the entire community and its leaders to address the educational needs of every child in every neighborhood in Newark. It was a bold and controversial effort that aimed to bring the best of what educators and policymakers had learned about driving change through top-down systems reform, bottom-up community demand, innovative labor approaches, and new school models.

Much has been written about the personalities and political drama in Newark during that time, but almost nothing has been written about the actual playbook and results. More importantly, I believe the reform efforts there still serve as useful examples to illustrate what is possible, what we should correct, and where we should go next.

Examination of this period in Newark will raise critical questions about what policymakers and community leaders at all levels should do to foster a system holistic and flexible enough that it would address the needs of all children, especially those our current systems of education and social service have historically and consistently failed the most. I will outline four steps policymakers can take to catalyze change and center the students and families who face the most challenges.

THE ROAD TO REFORM IN NEWARK

I was appointed superintendent of Newark Public Schools (NPS) in 2011 by then governor Chris Christie and the state's education commissioner at the time, Chris Cerf. While most school districts have a local board charged with hiring a superintendent, NPS had lost that authority back in 1995, when the state took control of the district.

New Jersey's constitution has uniquely strong language compelling the state to intervene in the event of chronic failure. As far back as the 1967 rebellion in Newark, there had been growing political support for the state to step in and address the city's abysmal student achievement, despite vehement objections from local elected officials.

I was the fourth in a line of state-appointed district leaders to arrive with a mandate for reform and improvement to Newark. But my arrival came with an even greater sense of urgency. About nine months before my first day, multibillionaire Mark Zuckerberg famously announced on *The Oprah Winfrey Show*, seated next to Governor Christie and Cory Booker, then mayor of Newark, that he was making an unprecedented one-time gift of \$100 million to the city's school reform efforts.¹ Just prior to this, the governor had also announced his decision not to renew my predecessor's contract, sending a clear signal that change had not moved quickly or decisively enough in Newark.²

It was clear that Newark school improvement was a signature political issue for Christie, and the pressure to raise achievement and improve efficiency was intense. On a trip to Washington, DC, earlier that year, he proclaimed to national media that "in the city of Newark, we are spending \$24,000 per pupil, and public money for an absolutely disgraceful public education system—one that should embarrass our entire state."³

Indeed, as I arrived in Newark, 39 percent of students who entered the system failed to graduate, and only 40 percent of third-graders could read and write at grade level.⁴ Enrollment was plummeting.⁵ The district's nearly forty thousand⁶ students and one hundred schools made it the largest in the state, with the majority of students living below the poverty level. The

schools themselves were a physical manifestation of the deterioration and decay. I still shudder at the memory of rats floating in basements⁷ amid boilers dating back to the 1950s and of Lafayette Elementary School (built when Abraham Lincoln was president), which had actual mushrooms growing in the cafeteria after Hurricane Irene.⁸

Local politicians and families had grown impatient. For the five years prior to my arrival as superintendent, many elected leaders had become early adopters of a growing national charter school movement that aimed to free individual schools and networks of schools from government red tape and allow them autonomy to innovate and build excellent schools. These supporters included Booker (then a young councilman), school board member Shavar Jeffries (who now heads the charter school behemoth KIPP Foundation), and state senator Teresa Ruiz, among other notable local leaders. Charters weren't the only new option—other school models, such as magnet high schools (often with entrance requirements) and partner-run small high schools, had gained momentum too.

Some of these schools had notable evidence of improving achievement for Newark students, and it was understandable that they were gaining strong support from local leaders, influential funders, and certainly the families of the nearly 5,500 students who were enrolled in them. Indeed, Zuckerberg's gift earmarked a significant amount of funds to growing the charter and small-school footprint. A plan between Christie and Booker that was leaked to the press around this time (causing uproar in the community) revealed that a central focus was to build more charter schools and open new small schools as quickly as possible.⁹

It was clear that the most impactful efforts at improving schools in Newark were actually working *around* the very system they were trying to improve. And in New Jersey, these new schools were funded on a per-pupil basis; in other words, the money followed the child out of the traditional system and into the public charter system. Logically, this made sense. But in practice, this proliferation of competitors to district-run schools was creating unintended consequences that few wanted to discuss.

During the interview process to become superintendent, though, I was very eager to discuss it. What was the long-term plan to ensure that students who remained in traditional district schools benefited from the cash infusion as much as those who were lucky enough to win school enrollment lotteries? What about the school closures and layoffs that would be inevitable when the footprint of traditional schools shrank? Would charter and new-school growth help bring excitement and excellence to our poorest neighborhoods, or would it give some kids better schools but make the conditions in their communities worse? Wouldn't the district end up serving more students who required the highest levels of support? No one seemed to have answers for these tough questions.

I had questions for the anticharter folks too. How could you blame families for exercising choice given the abysmal conditions of many traditional schools? Why were thousands of families on waiting lists? What policies—labor and otherwise—were holding traditional schools back from succeeding? Why hadn't those policies been advanced? Were community

leaders ready to push the bold reforms necessary for traditional schools to compete with charters? And what should we do about the chronically failing, profoundly underenrolled schools? The answers to these questions were complicated and generally had only to do with adult politics, not what's best for kids.

My background and the questions I asked made a lot of people very uncomfortable. Some saw me as “antireform” because I wasn’t “charter-friendly” enough, especially with the eyes of the nation now on Newark. At the same time, others felt I was in cahoots with the “privatizers” trying to kill public education by supporting new school models. In hindsight, it is easy to see that my story was just an example of a growing and deeply polarized national debate on whether and how we can radically improve the quality of education for low-income families and students of color at scale. Meanwhile, from day one, our team wasn’t focused on winning favors or promoting specific ideologies. We were focused on great schools for all kids in every neighborhood, by any means necessary.

BUILDING A “SYSTEM OF GREAT SCHOOLS”

Given the perilous state of the city’s schools, the unrealistic expectations around quick achievement gains, and the pressure from ideologues on all sides, many speculated that the superintendent role wasn’t doable. But I was inspired by the scale and the ferocious commitment of many leaders in the community.

I interviewed with a large committee and countless stakeholders, who spent hours debating diverse theories about what to do to improve schools. I was also moved by the unusual personal and political alliance formed by Christie, a popular Republican governor in a Democratic state, and Booker, a popular mayor and rising Democratic star, to do something bold. Something was in the air, and it felt like transformation was possible for children and families who had been failed by public systems for generations. I was convinced, perhaps naively, that if we could harness the debates and emotion around what to do, we could lift up a whole city. I remember thinking, *There are one hundred schools—we can do this!*

I assembled a diverse team of exemplary senior leaders—some known and trusted from within Newark, some with a track record of results elsewhere, and some with a lot of promise who were ready to step up. When I say “we” in describing the work in Newark, it is not simply rhetorical. I cannot take sole credit for anything we accomplished in those four years; it was this stunning team that acted as both the architects and the engineers behind the policies and practices we implemented, tirelessly, on behalf of Newark families. I wish I could list them all here by name.

It was time for us to craft our own plan for Newark’s schools. We started with the theory that the unit of change was the school itself and embraced the idea that what we were building was what my former boss, then New York City Schools chancellor Joel Klein, called “a system of great schools,” not a “great school system.”¹⁰ This was a subtle but profound distinction,

because it meant we were seeking to ensure that there were one hundred excellent schools serving every child in every neighborhood—regardless of governance structure.

In the long term, we knew that Newark required a citywide master plan that would account for every school building and every child. But in the immediate term, we had obvious and dire issues to address in the district’s own schools.

First, we needed to set a unifying goal for the district: every child would be college ready. That’s right, *college*, not just *career*—because we believed that choice of higher education should be up to the student, not simply determined by the inadequacy of their preparation, and because Newark families were demanding this.

While college readiness is an obvious educational goal for affluent families and communities, in Newark it was far from obvious that this was attainable or even desirable for most students. But to families, it was obvious. In poll after poll, focus group after focus group, they told us very clearly: they wanted their children to graduate college ready.¹¹ Moreover, they believed that “career ready” was a euphemism for low expectations. Families felt that academic excellence was a passport out of poverty.

Most parents were with us from day one. The challenge was the well-meaning funders and other influencers who wanted to muddy the waters and talk about everything except whether students could read, write, and do math at grade level.

To make a case for action, we shared baseline achievement data and started to create a sense of urgency around this critical goal. Looking unflinchingly at this data wasn’t easy, and it made many educators and leaders uncomfortable to acknowledge so blatantly how poorly our schools were preparing our students to succeed in life. When we started sharing actual data about proficiency rates and the number of young people earning diplomas indicative of their mastery of hard content, we started to encounter real pushback, both within and outside the school system. This was a theme I became increasingly familiar with: often what families say they want can be quite different from what those who speak for them are willing to stand for.

ENSURING “FOUR-INGREDIENT” SCHOOLS

With our North Star established, we rolled up our sleeves to improve the district, school by school. By this point in US education reform, it had been nearly thirty years since the release of *A Nation at Risk*, and there was a large and growing body of research and evidence about high-performing schools in high-poverty neighborhoods. Combined with our team’s years of on-the-ground school transformation experience, we zeroed in on four basic ingredients that every high-quality school possessed: people, content, culture, and conditions.

To extend the cooking analogy, we knew each ingredient could have a different flavor profile in a different environment. The exact mix of ingredients varies based on context as well. Getting all four ingredients right in a school building is a huge challenge on a good day and

nearly impossible the rest of the time. And yet the ingredients themselves are relatively uncontroversial—even commonsense—and should be the central focus of systems leaders and policymakers.

Our aim: ensure that every NPS school was a four-ingredient school so that we could make steady progress toward college readiness for all. Our philosophy: focus on what works regardless of ideology, which often led to “third-way” solutions—combining the best of seemingly disparate views or forging a new path to transcend old, binary thinking. Our mantra: implementation matters.

People

It’s critical to have the right people in the right seats, from the leadership team to the teachers to mental health professionals to custodial staff. No matter their role, every adult in the building must be equipped with the right mindsets and skill sets to uphold the mission and goals of the school.

Education reform birthed a broad array of nonprofits and policies focused specifically on teacher quality, notably Teach for America (TFA) and TNTP (formerly the New Teacher Project), the latter with its landmark Widget Effect report.¹² We know intuitively the power that a great teacher has, and a growing body of research reinforced this belief, showing us that teachers are the most significant in-school factor determining a child’s level of achievement.¹³ Further, the most significant factor in getting great teachers in every classroom is the quality of the principal. Meanwhile, author Amanda Ripley¹⁴ showed us that in Korea, master teachers are treated like rock stars, which is hardly the case here in the United States.¹⁵

Partly because of my experience working on the talent side and the emerging research about the strong correlation between school leaders and student achievement, we focused on leadership from day one in Newark.¹⁶ I’ve never been to a great school with a mediocre principal, and I have never been to a failing school with a terrific principal (except perhaps at the very beginning of a turnaround). Within two years, we had replaced nearly one-quarter of our principals through aggressive recruiting and selection, giving preference to Newarkers and leaders who not only knew instruction but thought of themselves as community organizers and change agents.¹⁷

We took a page from New York City’s playbook, enacting a “mutual consent” policy that allowed principals to select teachers aligned to their school’s goals, as opposed to having them force-placed according to seniority.¹⁸ We adopted a new teacher evaluation system that required more evidence-based classroom observations and feedback, taking a page from Chancellor Michelle Rhee and her team’s work in the District of Columbia.¹⁹ We trained and empowered principals to hold teachers accountable when they failed to uphold expectations—and we had their back when teachers needed to be removed. We created “career ladders” for teachers to become leaders without having to leave the classroom, taking a page out of the Baltimore playbook.²⁰

As was a consistent theme with our approach in Newark, pursuing and advancing third-way ideas had us making people on all sides of the issues uncomfortable. Many states at this time were starting to use quantitative test score data in teacher evaluations, and New Jersey was eager to follow suit. However, my team and I felt that the science for such “value-added” approaches didn’t hold up when it came to determining the effectiveness of individual teachers.²¹ We took a lot of flak from hardline education reformers, who had become fixated on using test scores as a shortcut to accountability and who worried that our questioning the use of test scores in teacher evaluations would water down reform efforts more broadly since Newark was such a high-profile example. But not only did we feel that using the value-added approach in teacher evaluations would be unfair to teachers, we also knew that including such a poison pill in our new evaluation plan would create a backlash that could sabotage the entire effort.

To help noncharter schools accelerate the “people” ingredient, we negotiated what was widely considered an ambitious contract with Newark teachers.²² We were able to find agreement with both the local and national teachers’ unions on contentious issues such as freezing pay for teachers who were ineffective; on providing bonuses for high performers working in hard-to-staff subjects; on expediting firing for the small number of teachers caught doing egregious things; and on finding pathways for individual schools to innovate outside the four corners of the contract. We also asked the state to grant us a waiver from traditional tenure laws so that we could consider quality alongside tenure when making decisions about whom to retain as we set about the necessary downsizing.

Despite agreeing to key labor reforms after more than two hundred hours at the bargaining table, some in the Newark Teachers Union²³ and their national affiliate, the American Federation of Teachers, vociferously advocated against them within weeks of the contract being ratified by an overwhelming majority of teachers. Both groups had a long track record of preserving some of the sacred cows of labor negotiations with teachers: seniority-based placement, infallibility of teachers with tenure (regardless of what they do), and resistance to any form of accountability—no matter how nuanced. Meanwhile, we found many of these ideas to be popular among everyday teachers, who told us the quality of the teacher in the classroom next door is a factor in whether or not they want to stay at a school. In fact, research shows teachers want access to high-quality curricula, comprehensive assessment systems, and the ability to collaborate with colleagues.²⁴ As the granddaughter of a union organizer and as a strong believer in collective bargaining, I was pushing largely because I believed then—and still believe now—that teachers’ unions need to evolve to become part of the solution or they will become obsolete.

We also had to completely restructure and reimagine the central office to be in service to schools and families. This required breaking senior leaders into new teams and inviting them to clearly articulate how they would enable the four school-level ingredients. It also meant crafting clear plans with goals aligned with good management and coaching—not simply doing what had always been done. Many rose to the occasion and embraced the opportunity for clarity and coaching. Some didn’t. This was another necessary and politically fraught task. Newark Public Schools was one of the biggest employers in the city, so many staff on the

central team were connected to a local politician, vendor, or influencer who had their back if their job was in jeopardy, regardless of whether or not the job was necessary or they were doing a good job.

Content

A high-quality school needs high-quality and culturally competent curricula. It also needs frameworks, protocols, and data that drive great instruction and continuous improvement. As computer programmers like to say, “Garbage in, garbage out.” This “content” ingredient is all about replacing that “garbage” content with engaging and carefully crafted content.

I started in Newark about a year after the Common Core State Standards had become a force nationally and the same month that New Jersey adopted a version of them. It was good timing, because I have long believed in fewer, clearer, more rigorous standards, as opposed to the laundry list I was handed as a young teacher in California. Common Core gave us an unambiguous and evidence-based target. It also served as a catalyst to scrutinize our curricula with a more rigorous lens.

The research here is undeniable; high-quality instructional materials are critical to ensuring that students are truly internalizing difficult content.²⁵ Historically, though, we had all underinvested in this area in the early reforms after *A Nation at Risk*. A lot of us made the mistake of keeping a hyperfocus on teachers, which indulged an assumption that if we had a perfect person in every classroom, they could invent brilliant content from scratch. Fortunately, the introduction of common standards forced the issue and led to game-changing work by leaders like John King and his team in New York to develop what would become the EngageNY curriculum.²⁶ Many education advisors and publishers have caught on, and there are far more “good enough” options out there, but there are still not enough.

We also were informed by systemic approaches to “managed instruction” like efforts in Charlotte-Mecklenburg Schools, the Success for All initiative in Los Angeles, and Newark’s own NorthStar Academies (a charter network). These reforms, often literally telling teachers what to teach and how to teach minute to minute, yielded some impressive results. But they also predated important reforms around teacher quality. Our team aimed to blend the best of both by making “good enough” choices about curricula, creating spaces for teachers to use existing scope and sequences and lesson plans but also building the capacity of teachers to use data and knowledge of their classrooms to adjust.

High-quality instructional materials are an ingredient that is hard to get right when you are working only at the school or small network level. Scale is your friend. These decisions are better made at a system level, where content experts can dedicate the necessary time to addressing academic needs and cultural contexts, as well as coherence and alignment between the plethora of different curricula and assessments. It is also the area that, at the time in Newark, brought the most consensus. We did “teach-ins” for administrators, educators, influencers, and families who all really seemed to get and support the mandate for good, rigorous content that was consistent across the city.

As I write this, the country has been embroiled in a resurgence of culture wars around what we teach in our classrooms. It's unfortunate, considering that this content ingredient was actually a rare point of local and national consensus not that long ago. The current political climate adds an unwelcome layer of complexity for systems leaders to battle.

Culture

We know from research that schools with intentionally curated environments characterized by high standards alongside high support produce better student outcomes. Students learn healthy habits, and the school community has well-established values and expectations. Norms and protocols prevent incidents, and when incidents happen, adults minimize shame and exclusion to keep students learning. In these schools, there is joy and choice.

From day one in Newark, we focused on the seminal research work and promising practices that had emerged, connecting how kids feel, how adults feel, and student outcomes. Years after comparing student achievement results to staff, student, and family survey responses, researchers Tony Bryk and Barbara Schneider found that the schools with high levels of trust were far more likely to get beat-the-odds results than their counterparts.²⁷ Economists like Ron Ferguson and social policy experts like Christopher Jenks found a direct correlation between adult expectations, student surveys, and student outcomes.²⁸ Though controversial, it is also no surprise to me that recent research has shown student surveys can predict student achievement²⁹ as well as teacher evaluations.³⁰ When young people say things like, "My teacher doesn't stop until I get it" or, "My teacher believes I can understand anything," or, "When my teacher asks me how I am doing, I believe s/he wants to know," we know those students will do better. Jeannie Oakes's *Keeping Track* showed very explicitly how adult biases and expectations can have a negative impact on student achievement.³¹

Relatedly, an area where I have seen some of the greatest challenges for adults in establishing and preserving culture is in response to conflict and disruptive incidents. How we handle student discipline, struggle, and conflict is where adult biases show up the most. Black students are four times as likely to be suspended than their White peers, often for the same behaviors.³² Black students make up one-third of school-based arrests, though they are only 16 percent of the student population.³³ Moreover, nationwide, Black boys are almost twice as likely to experience corporal punishment as their White peers, and Black girls are about three times as likely. Adults rate Black girls as "less innocent" than their White peers, with damning implications.³⁴ Students with disabilities are far more likely to receive exclusionary discipline for subjective things like "disrespect" and "insubordination."³⁵ This is a problem not only from an equity and justice lens but also from a student achievement standpoint. Often students who need the most support and time on task are being excluded the most. Students can't learn when they feel shame and helplessness. So it is no surprise to me that data shows that the relationship between the discipline gap and achievement is more than correlative—it is also causal.³⁶

The research squared with our team's lived experience in Newark. Traditional schools like Chancellor Avenue Elementary School and Sussex Avenue Elementary School, charter schools

like KIPP Spark Academy and North Star Academy’s Alexander Street Elementary School, and partner schools like Bard Early College got this ingredient right, and their results showed it. Adults, kids, families, and community partners rallied together around a common vision and values—and shared expectations and norms at the school reinforced them. Their results showed how critical it is to build a collective culture of high expectations and high support.

For these reasons, we hired administrators who showed skill in building culture and partnering with families. We created an entire central-office team focused on student well-being and discipline and charged them with building the capacity of schools to create environments where all kids would thrive and to respond to incidents skillfully. We reinvented the role of school resource officer and hosted weeklong restorative practice institutes that brought together student leaders, administrators, families, teachers, and police officers.³⁷

We made progress, but admittedly, the playbook on culture is harder to run for many reasons. Too often, discussions about what student culture should feel like are preachy, ideological, or theoretical—devoid of practical, research-based, promising practices. Building culture is far from a paint-by-numbers task. Cultures that are simply cheap forms of imitation, are inauthentic, or are misaligned to the needs of a particular community don’t create the conditions for achievement. Frequently, we think about individual tactics for establishing and preserving culture, such as specific expectations or restorative circles, but not about how they all fit together, and this leads to cultures that feel disjointed and incoherent. Effective cultures don’t feel the same in every school, but they do share key components. This is nuanced and hard to teach to administrators. At a systems level approaches to this ingredient often devolve into compliance lists and checklists. Further, the culture work requires us to surface and address adult biases about what kids can accomplish and what is considered “dangerous” behavior, and this can cause real discomfort and resistance.

Conditions

This ingredient is all about strong operations and infrastructure. The building is clean, well equipped, and well run. The trains run on time. You have the facilities, management structures, and funds to support learning, and you have the funding, supplies, and technology to support all of the other ingredients of a high-quality school.

While setting the right culture creates a social and emotional environment where both students and adults can thrive, it is important to simultaneously address the physical environment and the day-to-day operations. It may not be as compelling or sexy as the other ingredients, but none of the other ingredients of a strong school or system can succeed if we don’t address the conditions in which our children learn and our teachers teach. In Newark, we had a lot of work to do on this ingredient.

When I started in Newark, Malcolm X Shabazz High School had a river running through its fourth floor on rainy days. Many schools didn’t have air conditioning, in a city where average temperatures reach above a humid ninety degrees for months. Some schools weren’t even wired for internet access, and only a few had laptops to check out to students for the day.

Local leaders openly talked about a “rolling start” at the beginning of the school year, which referred to the fact that it took weeks to sort out the basics: enrollment, special education schedules and services, buses, and even books. Honestly, I had never heard of a system where instruction didn’t start on day one.

And while I’ve never bought into the idea that market forces are a panacea for improving school quality (and research suggests I’m right³⁸), you can bet my team took note when our colleagues at KIPP would make an overnight run to the hardware store for window A/C units to survive a Newark heat wave, while we were forced to navigate a maze of vendor regulations and nepotistic relationships just to open a window. Trust me, there’s an entire chapter to be written on reforming procurement and purchase ordering alone.

Some of these intolerable conditions were due to bad public policy and some were because of poor management. My team and I would say we could tell if a school was getting results by how visitors were greeted at the door (if at all) and how quickly families could get the answer to whatever they were asking. When organizations are well run, their primary constituents—in our case, students, families, and community members—are at the forefront of everyone’s mind. This service orientation is shared by noninstructional staff members, from the custodial staff to the crossing guard to the budget team. All are invested in the mission, goals, and shared values. We created school operations managers (similar to how some charter networks like KIPP have created business operations managers) to attend to the operational needs of the school. At the time, this got me in trouble with the administrators’ union (because I was seen as encroaching on district administrator roles and jobs). Even today our approach to operations is considered innovative, which just shows how little we prioritize the conditions in our schools.

Getting the conditions for achievement at every school in Newark was an expensive and backbreaking task, and progress was excruciatingly slow. Bureaucracy at every level—local, state, and federal—made our vision of goal-based budgeting, twenty-first-century facilities, nutritional food, and high-quality customer service feel nearly impossible on some days, even with exemplary people working on it. When academics and policymakers talk about disrupting systemic inequity, rethinking our entire system of education, or rising to the grave challenges initially posed by *A Nation at Risk*, they’d do well to spend a lot more time talking about basic operations. We need an entire movement to break the bureaucracy that is crippling school infrastructure.

THE ONE NEWARK PLAN

While establishing a focus on college readiness and building four-ingredient schools was our primary focus right out of the gate, we knew we had to make progress on a citywide plan that addressed the schools beyond our purview. Looking at the full picture in Newark, you saw that everyone—local early childhood providers, the district, third-party school operators, private schools, individually run charters, and large charter networks—was doing their own

thing, and the unintended consequences of this lack of coordination were becoming more evident and unsustainable every day.

From our earliest school visits, we could see that the poorest neighborhood schools were emptying out and becoming concentrated with the highest-need students and the lowest-quality staff. Historic buildings were crumbling while new facilities were being built, sometimes down the street or downtown. Supply and demand were misaligned: for example, the number of magnet high school seats requiring a certain level of academic attainment far exceeded the number of eighth-graders prepared to meet them. The diversity and variety of school models wasn't materializing; with all the new schools, we weren't actually providing a lot of choice, just more flavors of "no excuses" ice cream at the elementary level and a bunch of run-of-the-mill high schools.

Meanwhile, every year, including my first, our district had to cut about \$50 million. While there was certainly a lot of bloated bureaucracy to streamline, more than 80 percent of that money was wrapped up in people. Newark Public Schools employs many Newarkers³⁹ in a city with double the national poverty rate.

With every set of data that we unearthed and every school that I visited, the pit in my stomach grew as I saw the magnitude of the challenges. Keeping open too many schools without enough students meant that thousands of Newark's students were attending schools that were crumbling, not just physically but educationally. We were spending more money per pupil than almost every other district in the United States, with terrible results. And there was no denying that the rapid growth of the charter movement further complicated the problem.

Past district leaders—and many city leaders—viewed charters as "the enemy." Families able to navigate the charter lotteries were fleeing the traditional system, and thousands more were on waiting lists. Who could blame them, given that many charters were radically outperforming traditional schools? As money followed students out of the traditional schools and into the charters, the available resources to revive these schools or attract talent were being stretched increasingly thin—and the trend was likely to continue. Charter leaders planned to grow the sector from serving 5 percent of students to serving more than 40 percent, which would mean \$250 million of funds leaving the district with no quick or painless way to shrink infrastructure or union-guaranteed jobs. Seeing that dozens of district schools were dying a slow death, with some of the city's neediest students trapped in them, I knew something had to be done—and soon.

As a city, we had to ask ourselves: "Is it even possible for every child in Newark to have access to a school that meets their needs? Even those children facing the longest odds?" Note that this is a fundamentally different question than "Can we get *some* kids in this community access to great schools?" That framing suggests we are not responsible for all of the children in our community, only those whom our school model can accommodate. It is also a fundamentally different question than "Can we build more great schools?" That question ignores the community context within which the school exists, and it fails to address very real

and sometimes serious consequences when we focus on building some great schools and letting others fail. We were seeing that play out in Newark already.

Our team had no choice but to stare down these questions, which led us to some unconventional and controversial answers. The first thing we had to do: try to rise above political arguments rooted in ideology and self-interest about what type of school models should exist. There were about a hundred schools in Newark. We knew we would get to excellence more quickly if we had a variety of governance structures: traditional, charter, magnet, partner run, and hybrid. But we also knew we couldn't simply let a thousand flowers bloom and allow others to die, especially when those vulnerable schools were serving our students with the highest need. We also knew that the community deserved excellence citywide.

Just as always, our team was guided by national promising practices and research. Only this time, we were limited by a dearth of examples in which systems leaders thought about the entire community. So we pored over our own data: student enrollment trends across governance models, overall city population trends, facilities assessments, and (of course) student outcomes. We fanned out and hosted more than a hundred community-based meetings with faith-based leaders, nonprofit executives, families in struggling schools, families in high-performing schools, charter advocates, charter operators, private schools, local funders, elected officials, union leaders, and early childhood providers. We began to socialize the idea that we needed one citywide plan across governance structures, as well as the harsh reality that the district's footprint had to shrink. We wanted to find a way to preserve the best of the new-schools movement while also addressing some of the unacceptable consequences of its uncoordinated growth.

This process—over the course of about a year—led to a comprehensive plan we called One Newark. The plan was more than just a collection of policies and tactics. It was carefully architected with a clear and accessible framework to communicate honestly and transparently with an extremely broad array of stakeholders and get buy-in across the city. We knew that the best plan in the world would mean nothing if we didn't tell a coherent story that motivated change.

The plan opened with three core values to drive our collective decision-making: equity, excellence, and efficiency:

- **Excellence:** We must ensure that every child in every neighborhood has access to a “four-ingredient” school as quickly as possible and that no kid is in a failing school.
- **Equity:** We must ensure that all students—including those who are facing the longest odds—are on the pathway to college and a twenty-first-century career.
- **Efficiency:** We must ensure that every possible dollar is invested in staff and priorities that make a positive difference for all students.

It was followed with seven focus areas, which we mapped onto the letters in the word “success” to send a clear signal of optimism and affirmation:

- **(S)ystemwide Accountability:** Envision and publish a standardized approach to tracking school and system success and progress across all schools (district, charter, and provider run).
- **(U)niversal Enrollment:** Launch a straightforward, user-friendly enrollment system that empowers our students and families to choose the school (charter, partner, or district) that best meets the student’s needs.
- **(C)itywide Facilities and Technology Revolution:** Create a bold plan to operationalize twenty-first-century learning environments for all students, ensuring no vacant buildings.
- **(C)ommon Core Mastery and PARCC Readiness:** Lead the nation in the number of students living below the poverty level (especially those currently struggling) who make progress toward Common Core mastery and readiness for PARCC, the standardized test aligned to those standards.
- **(E)quity and Access for All Students:** Increase the number of high-quality seats for all students, especially those currently in low-performing schools.
- **(S)hared Vision for Excellent Schools:** Cultivate demand for one hundred excellent schools and the groundswell of support for the changes necessary to get there.
- **(S)ystemic Conditions for Success:** Radically transform the district itself to ensure that it is a high-performing organization for years to come.

While the backbone of the overall picture and the building blocks of the plan were emerging in the spring of 2013, we didn’t feel we had enough operational capacity or community momentum to implement the plan that fall. Instead, we continued to engage, discuss, and refine the plan’s tenets with diverse stakeholders. We launched headlong into implementation in the winter of 2013-14.

We knew from the start that we needed a fair way to compare school quality across model types to meet the goals of our citywide vision of excellence for all. I’d seen firsthand how smart ways of tracking progress could drive good change in New York City under former mayor Mike Bloomberg, and we knew of years of research by the Consortium for Policy Research in Education and others about the correlation between accountability and student outcomes.⁴⁰

We started publishing “family-friendly” snapshots—across both district and charter schools—so that community members could see how their schools were doing in comparison to schools with similar populations. We looked at overall proficiency but also at growth, critical in a city like Newark with low rates across the board. We also compared schools with

similar student populations to one another. There was no question that students at schools like Shabazz High School and South Street Elementary School were in much higher need than students at Science Park High School and Ann Street School. Success was possible, of course, but harder—and we needed a fair accountability system to make decisions and create the right incentives.

We created a simple red, yellow, and green system so that the community could see the landscape clearly. “Red schools” were low-proficiency, low-growth schools. Green were high proficiency and high growth (e.g., we didn’t want selective high schools to recruit “proficient” students and fail to grow them to “highly proficient”). Yellow schools were “on the move” (low proficiency, high growth) or “to watch” (high proficiency, low growth). The color-coding was clear and intuitive, and many in the community started talking about “no red schools.”

We placed an emphasis on transparent data about how schools were doing with students in poverty, students with disabilities, and English learners. We created standard measures—across district and charter schools—to report on student retention. Up to this point, accountability systems implemented under the No Child Left Behind Act of 2001 focused almost exclusively on proficiency, which unfortunately incentivized schools to enroll and retain students who were likely to be successful and to subtly counsel out (e.g., “you are not a good fit for this school”) or explicitly push out (e.g., through suspensions and expulsions) students who were harder to serve.⁴¹ We couldn’t afford to make the same mistake.

People from all sides fought us on this level of transparency—the unions, some charter schools (which weren’t obligated to share their data with us), and some funders who worried we were reducing children to numbers. But many families and policymakers embraced the information. There’s no perfect system, but there was no way to make a citywide plan without a decent measure of school quality.

We performed detailed enrollment analysis and defined the need for a common definition of a “minimum viable school.” From a funding standpoint, schools with fewer than five hundred students are hard to sustain with a staffing model that ensures things like appropriate class size, electives, teacher preparation times, and staff to attend to running operations. Newark had a lot of “red” schools that were also not financially viable, and many of them were in the poorest neighborhoods.

We also looked at demand data—who was applying to charters and from what neighborhoods, who was seeking new small high schools and from what neighborhoods, and which neighborhoods were growing and which were shrinking.

The picture was becoming increasingly clear: the need for a course correction was long overdue. We had traditional schools where 80 percent of families were on charter school waiting lists, but the district’s resistance to collaboration and the charters’ insistence on growing only one grade level each year meant large-scale closures and consolidations were inevitable.

The district had too many elementary schools overall, due to a population decrease, neighborhood shifts, and charter growth. We didn't have enough early learning centers to meet the increased demand. We had too many selective high schools. Most of the new small high schools being incubated downtown were serving families from other wards, while iconic and historic high schools were emptying out. Overall, we had too many old buildings that were crumbling due to underinvestment and age, and some of them simply weren't fixable. At one point, the district was paying more than \$1 million just for scaffolding on vacant buildings that were never going to reopen. The picture was bleak. We had to make some hard decisions.

We decided to be radically transparent about our findings and the implications in a proposed ward-by-ward plan. Some charters should take over existing schools with high demand, keep families who opted in, and keep the buildings and the school name, instead of simply continuing to build new schools one grade at a time. Some elementary schools needed to convert to early learning centers. Some small high schools that were performing well needed to move into our comprehensive high schools, and some underperforming partner-run high schools needed to close. Magnets had to change their enrollment process. And some buildings had to be shut—some condemned, some repurposed, and some sold, potentially to charters.

Another anchor of the One Newark plan was ensuring that every family had equal access to choice. Both psychologically and practically, it didn't make sense for one-third of families to get what they wanted and the rest to get what was left over. For starters, this dynamic was creating an almost civil war-like atmosphere, with charter and noncharter families pitted against each other and magnet and nonmagnet families screaming at each other in meetings.⁴² Also, one goal of establishing high-performing schools in high-poverty neighborhoods is to feed the groundswell of belief that kids can achieve. Newark's choice system was helping create a self-fulfilling prophecy of failure in the noncharter schools.

We had to find a way for the idea of choice to lift all boats, but it wasn't happening—and it can't happen without good public policy and collective action. I've had many school choice advocates dispute this. Some ideologues will have you believe that the mere presence of competition somehow magically raises everyone's game. It certainly didn't happen that way in Newark, nor in the dozens of systems I have worked in since.

This is where universal enrollment came into play. Cities like New Orleans and Denver had implemented systems where families could access a common application instead of having to apply to and navigate multiple lotteries.⁴³ We built on what they had learned (even flying in officials from both systems to participate in community panels) and took it a step further. We envisioned and implemented a system where all schools—charter and traditional—marketed themselves on the same timeline, using citywide approaches, and alongside our common accountability system.⁴⁴ All families could access the system and apply to all schools. An algorithm gave preference to kids in the neighborhood, followed by kids in poverty, then kids with disabilities, and then everyone else at random.

It was a game changer. Now all schools were required to think about how to market themselves and own their quality, or lack thereof. By year two, more than three-quarters of the families of kindergartners and ninth-graders were using the system.⁴⁵ At one point, we opened a family support center to help families exercise choice. We had actually planned for a soft launch, but word got out and more than a thousand families showed up on the first day, and the situation almost devolved into chaos. While our critics crowed about our operational failure—and it was indeed a failure—it also showed how much family demand there was for choice and quality. This is one of the hundreds of examples I’ve had throughout my career that defies the ridiculous stereotype that poor families don’t care about education.

The universal enrollment system may have been hardest on some members of Newark’s political elite who were used to the benefits afforded to them in an unfair, transactional system. I recall one meeting in which a prominent official—previously a supporter of mine—yelled, “You made a liar out of me! I told my cousin I could get her kid into this school!”

We had other “lift all boats” strategies. In partnership with the Newark Trust for Education, we created shared campus grants so that charter and district schools in the same building were incentivized to envision projects that helped their students and staff collaborate on school-wide and community improvement projects.⁴⁶ We asked charter teams to lead professional development for some of our turnaround schools on things like comprehensive approaches to instruction where many of them had more robust practices than we did. We created a collective action team of special educators across district and charter schools to help share and promote promising practices.

The plan meant a lot of changes for a lot of people. Some shuttered buildings were historic, and even though it was clear these buildings needed to have a divestment plan, community elders who remembered their heyday didn’t want to hear it, understandably. Many charter leaders and their supporters dug in their heels on their model of growing slowly and where they wanted to grow according to optimal facilities, regardless of the consequences. The idea of small schools within high schools—which had been successfully implemented in New York at scale—was new to Newark and, therefore, scary for many who had found their pet school to support. Some local and national funders who were excited about ribbon cuttings and smaller projects simply didn’t want to get involved in the far messier project of citywide progress.

Our team knew that the tenets of the plan were bold, unconventional, and controversial and that the politics were going to be tough to navigate. Choice, charters, labor reforms, and teacher excellence polled well. Laying off Newarkers and teachers and “closing” traditional schools or turning them over to highly successful charters were wildly unpopular. But to have the plan succeed citywide, you couldn’t have one without the other.

To add a deeper degree of difficulty, while the plan was emerging and leading up to the official launch, we suffered a series of seismic political blows at the worst possible moment. In September 2013, the Bridgewater scandal broke and increasingly sidelined Governor Christie. My team went from coordinating with his team and political allies in Newark (he had a lot)

on a weekly basis to going months with virtually no communication. Shortly thereafter, then senator Frank Lautenberg tragically passed away. Mayor Booker, who had also been an active and strong supporter of the plan and was working hard to build momentum around it, announced he was running for that US Senate seat. This not only effectively took him off the field from a local political standpoint, but also created a scenario where he needed the support of local officials and union leaders who opposed many parts of the plan. His announcement also spurred the need for an earlier-than-expected mayoral election where the leading candidates spent considerable time spewing hatred about charters and about me personally (although backstage and publicly, they had previously supported both). Shortly thereafter, Commissioner Cerf resigned. To use a sports analogy: the entire offensive line left the field.

The overall approach was comprehensive, and it had to be to ensure that none of our kids were trapped in failing schools, the district didn't go bankrupt, communities weren't living with vacant buildings, and the city was on a path to success. I described the plan to author Dale Russakoff as "three-dimensional chess" in an effort to convey why all the pieces had to happen at one time and couldn't be phased. There were too many interdependent parts to a very complex system, and the stakes couldn't be higher. Unfortunately, in her book about Newark, *The Prize*, which went on to become a bestseller, this quote fed an inaccurate portrayal of me as a top-down, cold technocrat—a narrative that was taking shape across much of the media coverage about our work in Newark. It couldn't have been further from the truth—the emotional pieces of what needed to happen were not lost on me or the team. I lived with my husband and baby son in Newark and had conversations with neighbors in grocery stores and local watering holes on a daily basis. It all felt so heavy, but also necessary.

RESULTS AND LESSONS

During my tenure and the subsequent years under Cerf, our district teams improved outcomes for students in every neighborhood and every age group—from early childhood to high school.⁴⁷

In early childhood, we secured a \$7 million Head Start grant⁴⁸ (only the second district in the country to do so) to add more than one thousand early childhood seats. We brought early childhood standards to life and sounded the alarm to focus on the importance of high-quality early learning. Newark went from having fewer than half of our residents eligible for free early childhood programs (which was most families) to enrolling nearly 90 percent.⁴⁹

In 2015, the Center on Reinventing Public Education named Newark as the number-one district in the country for high-poverty, high-performance elementary schools that beat the odds.⁵⁰ By 2019, more than one-third of Black students attended schools that exceeded the state average, compared with 10 percent in 2011.⁵¹ A study conducted by Harvard University showed across-the-board increases in reading and initially slow but then impressive gains in math.⁵² The number of good schools and schools "on the move" grew every year due to our district-run turnaround approach, charter conversion schools, and some outright closures

and consolidations. Newark was among the top four cities in the country for student outcomes of Black students living in poverty.⁵³

The citywide graduation rate rose 14 points, closing the gap with the state average by 7 percentage points—with almost double the percentage of students graduating having passed the state exit exam.⁵⁴ About 87 percent of Newark graduates returned for a second college term, far exceeding national averages given the high poverty rates.⁵⁵

And we saw signs that the overall community—despite the political rancor we encountered—was starting to believe in the “system of great schools.” For the first time in decades, student enrollment was increasing overall in Newark, as was the population of the city.

Our labor agreement, too, was a long-term success. More than a decade later, most of these terms still exist in the contract today, and an independent study of the agreement found that the “new evaluation system is perceived as valid, accurate, fair, and useful.”⁵⁶ This suggests that its durability is not just due to luck and that the approach could and should be replicated elsewhere.

Similarly, universal enrollment still exists in a modified form to this day and is used by nearly 20 percent of families.⁵⁷ I believe managed school choice was starting to play a role in dissipating the city’s deeply concentrated poverty.

Despite these significant accomplishments, we knew from day one that we would not succeed in Newark unless everyone, from the grandmother at a profoundly struggling school to the dad at a magnet school to the aunt at a charter school, believed things could be different—better for everyone, for Newarkers. We knew we had to build a completely new normal and that some of that work involved helping the entire community see that it didn’t have to accept the failed status quo. We exerted a lot of effort that, in the end, fell short of generating the kind of collective momentum we needed. The reasons are complicated but instructive.

Because we felt responsible for every child in Newark, we engaged all families, charter and district, with equal vigor. This was a good and mission-aligned approach, but it was almost impossible to execute, given the tensions (both perceived and very real) inherent in growing the charter footprint. The conundrum is perfectly exemplified by the mother who called in to ask me a question on-air during a local NPR show. She had just dropped off her kids at North Star Academy Charter School, she said, because she needed them to have access to excellence. At the same time, she was on her way to my office to picket against me on behalf of her nephew, who had lost his job as a school aide due to the smaller footprint of the district.

Our strategy all along was to be up front about failure and embrace accountability. Again, while our radical transparency seemed like a good idea on its face, it turned out that a lot of people don’t want to hear their school is failing—no matter how carefully crafted the message. Also, while some community members were grateful that someone was “finally telling the truth” (an actual quote from a community meeting I led at a failing school), others were

understandably angry. Our team was on the receiving end of the grief and loss that result from telling the patient they have stage IV cancer when someone should have caught it years ago.

We were lucky to have a popular Republican governor in Christie and a Democratic mayor in Booker, who teamed up to create a real mandate for change and put a laser focus on what's best for students. This was a tremendous asset (and seems unthinkable in today's environment) but also a challenge. Some Newarkers resented the involvement of the state (particularly in managing the school system directly) and, by extension, me. And local officials fought even harder to exert influence, sometimes over off-the-mark things, to show they were relevant.

We prioritized students who were at the back of the line. Our universal enrollment system gave preference to students from the poorest neighborhoods and those with disabilities. We revamped the magnet school admissions process to look at multiple factors for student admissions at the central office. These were good decisions for children, families, and equity, but it also put us in the crosshairs of power brokers who were used to getting what they wanted and considered coveted seats theirs to give out. They also had access to the biggest microphones and would use them to mobilize the community against our efforts.

Some charter school operators and their supporters mobilized their constituents in opposition to these citywide efforts as well. They wanted to grow where they wanted to grow, not necessarily in alignment with supply-and-demand patterns or the overall plan. Many (not all) were content to crow about how much better they were than the district without digging into what percentage of "high-need" students they were serving—conveniently avoiding an apples-to-apples comparison that was much more complicated. They enjoyed promises from politicians and funders that were out of alignment with the One Newark collective plan. Many liked running their own lotteries because they had more control over admissions; some would say things like "we don't offer that kind of special education program." Also, many had legitimate concerns about turning over enrollment processes to a district that had been underperforming and had actively sought to extinguish them for many years.

Charters weren't the only group stuck in their own goals and plans—and at least most of their concerns were in service of building quality schools. School-based partners and vendors, local nonprofits, funders, and other leaders all had their individual projects, schools, and pet issues. The incentives to keep doing one's own thing were profound. I was stuck in a daily loop of explicit and often threatening demands to support individual agendas—many of them having nothing to do with what was best for individual neighborhoods and schools, let alone the collective. A local reporter continually nagged me about shoring up my "natural allies." I remember wondering who they were. Breaking up monopolies and pursuing third-way ideas is a lonely endeavor, particularly in cities like Newark, with its transactional, machine politics.

Well-resourced forces of opposition spent a considerable amount of money spreading misinformation and actively attacking me personally. They made expensive sandwich boards, posters, and fliers with my face on it. In one image, the word "Liar" was printed as if it were

carved into my forehead and dripping blood. It was an open secret that they hired a full-time blogger to write stories about our work and about me personally (some with twists of the truth and some with outright lies). The blog was well formatted, looked like a real newspaper, and generally contained kernels of truth that were leaked from inside. Ads were purchased to place those stories in actual newspapers so that they looked like real news. Canvassers were hired to distribute leaflets about false school closures, and social media stalkers posted where my family and I were eating dinner.

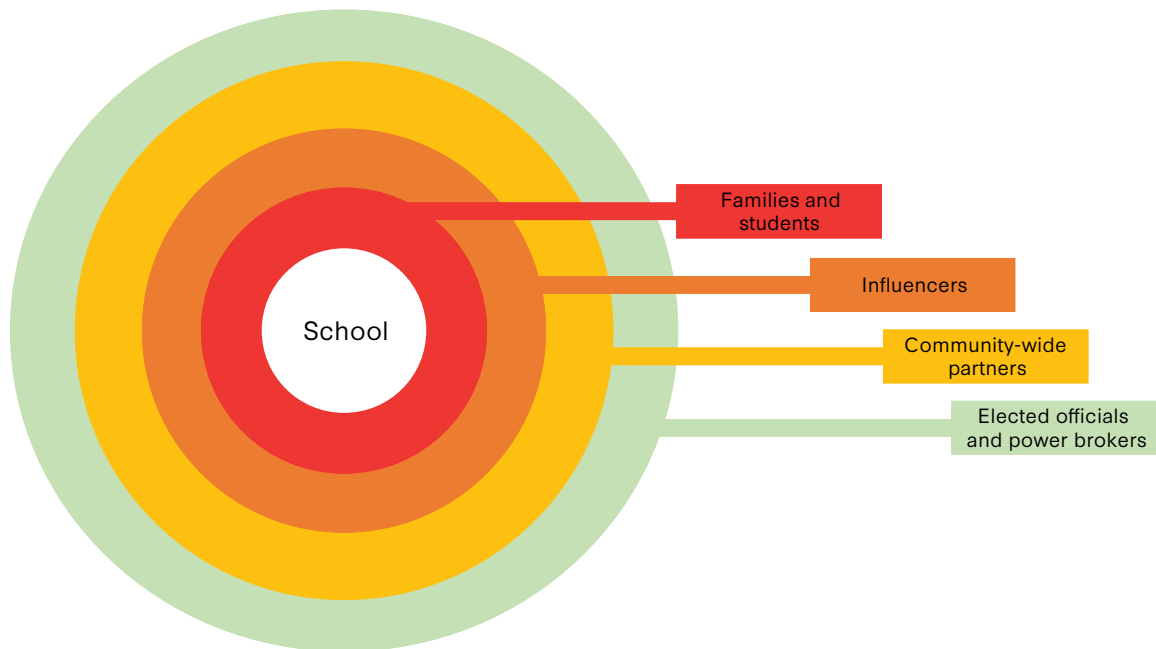
As important as we knew collective buy-in was to our success and as much time as we invested in it, our team was ultimately not successful in creating a groundswell quickly enough. We certainly had moments, but not enough. The One Newark plan should have been envisioned before the unintended consequences were at our doorstep. Maybe that would have given us more time. Surely, I was the wrong messenger: a White woman from out of town who represented the system. One Newark could have been a third-party entity with representatives from various sectors and a trusted, local leader. I thought this all along but failed to get stakeholders to agree and execute quickly enough. Meanwhile, we had to balance the budget and ensure quality in education.

I also clearly made mistakes. My messages were not straightforward and sticky enough. This work, as you can see, is complex and multifaceted, and I could have paid more attention to how to ensure good, proactive, community-friendly communication. I did not lead our team in good enough ways, small and large, to predict and combat misinformation that was rampant and that got even worse as social media exploded during my tenure. The forces for the status quo were organized and mobilized, and we were caught flat-footed. I didn't manage the flow of information with nearly enough precision, let alone attend to building my own brand. I made a classic mistake that many leaders have made before me: I presumed that if I did good work and led with authenticity, people would support progress.

More critically, I poured valuable energy into the community without focusing as much as I should have on the community influencers closest to our work at the school level. Since then, I've developed a more sophisticated understanding of how to see the community in relation to the system of schools. In figure 1, the center is the school, and the next level out is the families and students (red ring). The next ring is influencers (orange ring)—folks connected to the school who have direct influence on that specific school. The next ring is community-wide partners (yellow ring)—community-based agencies and other city agencies like police and child welfare. And the next ring out (green ring) is elected officials and power brokers—for instance, pastors of large congregations, thought leaders, and community-based organizations serving the city.

We knew it was critical to focus on our families and students, and we knew it was a tremendous amount of our work to build collective action focused on them. I give us high marks for our dogged and strategic work on the red ring. But in retrospect, we spent far too much time with folks in the outermost ring—the political and power class—and not enough with those in the orange and yellow rings. It wasn't until nearer the end of my tenure that we started to create a database for each individual school's yellow ring. We also made a much more

FIGURE 1 The community in relation to the system of schools



concerted effort to know the civil servants at all of our partner agencies. I came to realize a hard lesson—that while the politicians and power brokers confidently spoke for the community, they were often after a political win: a contract, a coveted spot in a school, a policy, or a job for a family member or friend. I wish I could take precious minutes I spent with those in the green ring and reinvest them in the yellow and orange rings.

The painful but informative experiences I had in Newark, along with a long career since then of working with systems leaders across the country, have convinced me that collective action is the missing link for change at the systems and community levels. Sadly, it is also the element I find most commonly reduced to uninspired bumper stickers and is wholly disconnected from the powerful and real work of reform. We talk about “community engagement” without honestly defining who the community is. We talk about consensus when real and hard calls have to be made every day about managing access to scarce resources, coveted high-quality seats, and community-based jobs. Many of those decisions can be in direct tension with building a system of great schools in every neighborhood by any means necessary. We interchange concepts of true grassroots organizing with community engagement and sidestep the obvious truth that power brokers and special interest groups have an organized, well-resourced, and often outsized influence on speaking for the community.

RECOMMENDATIONS FOR SYSTEMS LEADERS

While the work we did in Newark has been treated by many in education reform as a critical case study for emerging superintendents and district leaders, it is often told as a cautionary

tale of doing too much and ignoring community engagement. What is often lost are the lessons for building a successful “systems of schools.” Fundamentally, the story of the One Newark plan is a story of a district seeking to break out of the shell of its narrow school footprint and hold itself accountable for the educational futures of all its city’s children. The story of Newark should push all of us to define the role of the “system” and why it is so critical and yet so difficult to fulfill that mandate for an entire community.

Since *A Nation at Risk* was published, we’ve had important attempts at systemic reform that focused on specific pieces of the puzzle. Sadly, many of those efforts focused narrowly on individual “ingredients” (to use my earlier analogy) but not on the whole recipe, or the whole system.

We’ve had a lot of reforms focused on “community engagement” approaches, like the Annenberg Challenge.⁵⁸ But those initiatives failed to address the fundamentals of building better schools and didn’t wrestle with the extremely tricky work of defining community that is illustrated in the Newark example.

We have also seen many efforts to build better individual schools; most of the charter and “high-quality schools” movement has been about that: creating individual proof points without thinking about the layer above the school, let alone the community. New Orleans shows the profound limitations of this strategy, as the city still struggles to figure out the role of the system after watching a bunch of individual charter operators solve some problems and create new and complicated ones.⁵⁹ The Newark case study illuminates that while this approach can be vital for building “four-ingredient” schools, it will always be insufficient for establishing a holistic *system* of great schools.

We must focus on creating systems of great schools—not great school systems, not individual schools alone, not one piece of a puzzle, not some simplistic version of community engagement. We need to get clear on the roles of leaders at the systems level. In short: the system should manage the incentives, policies, guardrails, and resources to ensure that every child has access to a four-ingredient school by doing four things.

1. ENABLE “FOUR-INGREDIENT” SCHOOLS

As discussed above, we have promising practices when it comes to ensuring a game-changing principal in every school and an excellent educator in every classroom. We know the impact of high-quality instructional materials that are culturally competent. We have proven research on the importance of school culture and handling discipline. We know what conditions have to be in place to enable achievement. Systems leaders should set direction and advocate; procure best-in-class materials; set policy to incentivize districts, schools, and charter management organizations to implement what we know works; and sanction practices antithetical to student progress.

As one example, when Mississippi focused on the science of reading, providing best-in-class materials, training, and a way of measuring progress, kids across the state started reading

at unprecedented levels.⁶⁰ As another example, Nebraska adopted high-quality instructional materials statewide and provided options but also high-quality implementation support. This drove impressive gains in student outcomes.

Too often, cities select a superintendent or states select a commissioner because they are zealously focused on one single ingredient. Sometimes cities and states simply won't touch an ingredient because they don't want to fight with the union or other interest groups.

Policymakers have a responsibility to ensure that schools can obtain and mix best-in-class ingredients more quickly—trying to do so one school at a time doesn't make sense. A lot of this work happens at the district and network levels, but leaders at all levels must put people in place who understand and are committed to all four ingredients.

2. ENSURE QUALITY AND EQUITY

Our current system of districts versus charters sadly guarantees that many kids—particularly those with the most challenges—are left behind. Policymakers and community leaders should be held accountable for allowing kids and families to fall through the cracks.

Leaders need to step up and raise their hands for being accountable for all kids to access high-quality schools. They need to embrace good enough ways of measuring what that means—in terms of what students are learning and how they are feeling. Accountability systems need to help families hold schools to a standard of excellence for all kids, including those who consistently fail in all kinds of schools. These accountability systems should be family friendly and public. And they should explicitly shine a light on where inequities show up: fewer Black students having access to AP course and magnet programs, special education students in segregated classrooms with abysmally low student outcomes, inequitable criminalization of student behavior, and kids living in concentrated poverty too often getting the lowest-quality staff. Our new accountability systems should correct for some of the mistakes we made before, from focusing only on proficiency and meaningless graduation rates to treating growth, college-readiness, and retention as critical outcome measures.

Effective accountability systems have incentives that inspire schools and communities to step up. As one recent example, schools that were designated as failing were more than twice as likely to make big gains than those that weren't. Researchers surmised that this is because the label drove resources and supports where they needed to go—and rallied communities to do better. Good accountability systems drive decisions, sometimes hard ones, about redesigning schools, radically changing who runs them and how they are run, and even closing them.

3. BREAK BUREAUCRACY

A fundamental way to clear a runway for accelerated school improvement is to actively tear down past practices and federal, state, and local policies that block individual schools from innovating. As one example, school finance formulas are unnecessarily complicated and

opaque.⁶¹ Most states have an even worse and more complicated approach to funding facilities and infrastructure. We need more of a “whiteboard” approach than one that tweaks decades of dysfunction.

Traditional schools will never be able to “compete” with charters if we don’t actively tear down the unnecessary bureaucracy in existing schools. As one example, when Klein was New York City Schools chancellor, he had an entire team dedicated to creating one-stop compliance and communication approaches for principals in New York City over multiple years. They shrank thousands and thousands of pages of sometimes competing “must-dos” (most of them having nothing to do with serving students better) and still felt there was more to do.

Policymakers and community leaders need to wake up every day wondering what they can do to ensure that people running schools have the time to do the right thing as opposed to managing byzantine policies and procedures from competing departments. We certainly need oversight and compliance, but it must be streamlined and reexamined every year.

4. CREATE CROSS-SYSTEM AND COMMUNITY-BASED SOLUTIONS

The students who face the most challenges have generally been failed by multiple systems. I have a term for this: students whom systems have failed the most (SSFMs). Statistically, they are likely to be students of color. Too often they are labeled “special populations” and further marginalized out of classrooms and into separate and unequal programs. Very few schools of any governance structure meet the needs of these young people. Schools—and the systems in which they operate—are consistently failing 20 percent of their most vulnerable students.

Often these students and their families are connected to multiple systems: child welfare, public housing, homeless services, juvenile justice, criminal justice, immigration services, family services, and food programs. Some examples: nearly 90 percent of the juvenile justice population were in foster care at some point in their lives.⁶² About one-third of thirteen-to-seventeen-year-olds experience some sort of homelessness. These students struggle tremendously in school. In other words, we take the kids who—through no choice of their own—have been failed by one system and then fail them in another.

In order to truly reverse patterns for students that systems have failed the most, we need cross-agency and community-based solutions with school success at the core. Neighborhood-based collaboratives, like the Harlem Children’s Zone, have produced promising results that we started to scale during Arne Duncan’s tenure as secretary of education.⁶³ I’ve had the pleasure of working with teams creating memorandums of agreement between disparate agencies—the DA’s office, probation, public housing, and school systems—to share data and create common family support plans for young people and families connected to multiple systems. We need more out-of-the-box ideas to aggregate services and help students who are the most vulnerable succeed.

CONCLUSION

The insights and recommendations I've shared above are not based on any specific ideology. They were developed out of necessity and refined through years of application and practice across a wide variety of settings—from New York to California and many places in between, in both districts and charter networks, in small-school communities, and in the largest cities and states.

It may seem like a lot to tackle, and indeed it is. But if we are to truly transform our systems at scale, we can't simply cling to one specific ingredient or hew to a single governance ideology. The surest way to avoid bias and ensure a holistic strategy is to zoom out to the community-level goal. Make the community—not just one school, network, neighborhood, or district—the unit of change.

When I arrived in Newark, we started by asking how to ensure one hundred excellent schools to educate every child—including and especially those who are typically left out of the conversations about excellence. Moreover, how could we do it as quickly as possible (because kids only have one third grade)? Many education reformers at the time were looking for silver bullets—debating whether charters are better than traditional schools or what makes the perfect teacher or curriculum. Our approach of stepping back and asking what the shared goal was for the entire community led us down a fundamentally different path. That path was inherently third-way and therefore had us at odds with hard-line choice advocates, status quo defenders, and other rigid ideologues. But it also kept us focused on the community level and (on our good days) prioritizing those members of the community closest to the center of the circle.

Our efforts in Newark stood on the shoulders of emerging research and promising practices from around the country. Our team's focus on talent at every level was inspired by work in New York City, Baltimore, and Washington, DC, and in organizations like the New Teacher Project and New Leaders for New Schools. We took lessons from Student Achievement Partners, New York State, Charlotte-Mecklenburg Schools, and high-performing charters when it came to focusing on high-quality instructional materials. Our focus on high expectations and high support cultures was informed by research on achievement motivation, the intersection of adult expectations and student outcomes, and Chicago's work on measuring student and family satisfaction. Our citywide plan built on work from the new small schools movement in Chicago, New York, and Denver and universal enrollment in New Orleans and Denver. Throughout, we aimed to take the best of what was working elsewhere and ensure that it met the unique needs of our comprehensive citywide plan in Newark.

When leaders make the commitment to put community at the forefront of their work, they will encounter a number of challenges. In closing, I'll offer four guidelines to help us move forward.

First, think about the system of schools, not the school system. As I said before, this sounds like merely a semantic difference, but don't be fooled. Anytime we find ourselves putting aside what the broader community needs and instead focusing on what the district or network needs, we may achieve some short-term gains for the school system but often at the cost of the long-term goals for the community's system of schools. Continually reminding ourselves of this important distinction can go a long way toward preserving our focus.

Second, embrace a better, more honest definition of community. It is critical that the bounds of the community feel authentic to the members of that community—geographically, culturally, and politically. It is often convenient for politicians and others in positions of power to leave out members of the community to allow special interest groups to frame the conversation. In the planning and execution of systemic work, we have to put at the core of the work the very members who have been consistently failed by the system itself. When we are honest about who the real stakeholders are in our community and have clear priorities, the opportunity for real systemic change becomes possible.

Third, reject the idea that we have to start from scratch. In this chapter, I have identified many policies and practices that worked, some of which have been tossed out because of politics. The past five years have been a time of extreme polarization in all areas of public discourse, with education as no exception. An emerging playbook was beginning, but it has been all but obliterated. We need innovation for sure, and we also need a clear-eyed assessment of what didn't work, but we aren't starting from scratch. Kids don't have time for us to reinvent the wheel.

Fourth, accept that this is messy and that revolutions are never quiet or fast. As much as I've tried to codify my lessons and experiences into an actionable, coherent framework, nothing gets around the fact that transformative and disruptive systemic change will never be quick or tidy. This is not outpatient laser surgery that leaves no scar; this is chest-opening, quadruple bypass surgery with a lot of risks and long-term effects. But like a good doctor's goals, our charge is to help the patient lead a long, healthy life. We also cannot expect immediate results. Real change takes time.

The ideas and epiphanies in this chapter I share humbly and with tremendous gratitude to the countless friends, colleagues, and mentors in this sector who helped shape my beliefs about this work. It's been more than a decade since I arrived in Newark and forty years since *A Nation at Risk*. My hope is that we've all gained a bit of useful perspective and are ready to roll up our sleeves and put the lessons we've learned into action.

HESI PRACTITIONER COUNCIL RESPONSE

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.

Reading Cami Anderson’s description of the “four ingredients” that were so critical to her district-level turnaround work, I was struck by how applicable that list is to state education agencies (SEAs) as well.

Getting and keeping good people is key in any organization. One of the novel approaches I’ve seen SEAs use is the creation of year-long residencies for school and district personnel. Having a lead teacher or curriculum director work at the state department for a year not only gives them insights into state education policy, it also brings an important voice and perspective to the SEA.

High-quality instruction materials are increasingly a focus for SEAs, many of whom are using incentives to encourage schools and districts to adopt more impactful educational content. SEAs are also supporting their districts in new ways around data collection and analysis, providing a critical set of supports for student-centered learning.

From my time working with state chiefs, I know that shifting SEAs from a culture of compliance to a culture of responsiveness and innovation has long been a priority. SEAs were largely built (and funded) to ensure compliance with state and federal laws, and old habits die hard. Nevertheless, state leaders are working to create the conditions (the fourth of Anderson’s four ingredients) needed for real culture change by engaging stakeholders, developing school- and district-focused strategic plans, and improving internal capacities to increase impact.

In short, there is plenty any organization can learn from Anderson’s leadership in Newark.

—Stephen L. Bowen, executive director of HESI,
former commissioner of education for Maine

NOTES

1. “Mark Zuckerberg Announces \$100 Million Dollar Grant,” Oprah.com, September 24, 2010, video, <https://www.oprah.com/own-oprahshow/mark-zuckerbergs-big-announcement-video>.
2. John O’Boyle, “Newark Superintendent Clifford Janey’s Contract Is Not Renewed,” *Star-Ledger* (Newark), September 1, 2010, https://www.nj.com/news/2010/08/clifford_janey_is_ousted_as_ne.html.
3. O’Boyle, “Clifford Janey.”
4. Newark Public Schools, “Great Expectations: 2009-2013 Strategic Plan,” Newark Board of Education, 2007, <https://www.nps.k12.nj.us/archives/strategic-plan-great-expectations>.

5. Jesse Margolis and Eli Groves, "A New Baseline: Progress in Newark's District and Charter Schools from 2006 to 2018," MarGrady Research, 2019, 10, <https://njchildren.org/wp-content/uploads/2019/06/A-New-Baseline-in-Newark.pdf>.
6. Essex-Newark City Advertised Enrollments, March 25, 2011, <https://www.nps.k12.nj.us/wp-content/uploads/2014/10/2011-2012UserFriendlyBudget.pdf>.
7. Victoria St. Martin, "Newark High School Students Walk Out in Protest of Filthy, Unsafe School Environment," *Star-Ledger* (Newark), October 8, 2010, https://www.nj.com/news/2010/10/newark_high_school_students_wa.html.
8. "Irene Flooding to Force Newark Students Out of School Building," CBS New York, September 5, 2011, <https://www.cbsnews.com/newyork/news/irene-flooding-to-force-newark-students-from-school-on-first-day-back>.
9. Richard Perez-Pena, "Schools Run by Newark May Yet Reflect Christie's Vision," *New York Times*, September 23, 2010, <https://www.nytimes.com/2010/09/24/nyregion/24newark.html>.
10. "Reforming Urban Schools," CSPAN, June 4, 2008, video, <https://www.c-span.org/video/?205842-1/reforming-urban-schools>.
11. Newark Public Schools, "Great Expectations."
12. Daniel Weisberg, Susan Sexton, Jennifer Mulhern, and David Keeling, "The Widget Effect: Our National Failure to Acknowledge and Act On Differences in Teacher Effectiveness," New Teacher Project, June 8, 2009, <https://tnpt.org/publications/view/the-widget-effect-failure-to-act-on-differences-in-teacher-effectiveness>.
13. Isaac M. Opper, "Teachers Matter: Understanding Teachers' Impact on Student Achievement," RAND Corporation, 2019, <https://www.rand.org/education-and-labor/projects/measuring-teacher-effectiveness/teachers-matter.html>.
14. Amanda Ripley, *The Smartest Kids in the World (and How They Got That Way)* (New York: Simon & Schuster, 2010).
15. Byron Auguste, Paul Kihn, and Matt Miller, "Closing the Teaching Talent Gap," McKinsey & Company, September 1, 2010, <https://www.mckinsey.com/industries/education/our-insights/closing-the-teaching-talent-gap>.
16. Jason A. Grissom, Anna J. Egalite, and Constance A. Lindsay, "How Principals Affect Students and Schools: A Systematic Synthesis of Two Decades of Research," Wallace Foundation, February 2021, <https://www.wallacefoundation.org/knowledge-center/Documents/How-Principals-Affect-Students-and-Schools.pdf>.
17. Winnie Hu, "Newark Is Betting on a New Wave of Principals," *New York Times*, September 15, 2011, <https://www.nytimes.com/2011/09/16/education/16newark.html>.
18. Hu, "New Wave of Principals."
19. Hu, "New Wave of Principals."
20. Hu, "New Wave of Principals."
21. Kimberlee C. Everson, "Value-Added Modeling and Educational Accountability: Are We Answering the Real Questions?" *Review of Educational Research* 87, no. 1 (2015): 35-70, <https://doi.org/10.3102/0034654316637199>.
22. John Mooney, "Newark Teachers Contract, a Game-Changer for NJ's Often-Troubled School District," WNYC New York, October 19, 2012, <https://www.wnyc.org/story/244879-newark-teachers-contract-game-changer-njs-often-troubled-school-district>.
23. Steven Sawchuk, "Plan to Revamp Newark District Hits a Nerve," *EducationWeek*, March 17, 2014, <https://www.edweek.org/leadership/plan-to-revamp-newark-district-hits-a-nerve/2014/03>.
24. Susan Moore Johnson, "The Workplace Matters: Teacher Quality, Retention, and Effectiveness," Harvard Graduate School of Education, July 2006, <https://files.eric.ed.gov/fulltext/ED495822.pdf>.

25. Matthew Chingo and Grover J. "Russ" Whitehurst, "Choosing Blindly: Instructional Materials, Teacher Effectiveness, and the Common Core," Brookings Institution, April 10, 2012, <https://www.brookings.edu/articles/choosing-blindly-instructional-materials-teacher-effectiveness-and-the-common-core>.
26. Yasmeen Khan, "Common Core Supporters Speak Out," WNYC New York, December 10, 2013, <https://www.wnyc.org/story/supporters-common-core-speak-out>.
27. Anthony Bryk and Barbara Schneider, *Trust in Schools: A Core Resource for Improvement* (New York: Russell Sage Foundation, 2002).
28. Ronald F. Ferguson, "Can Student Surveys Measure Teaching Quality?" *Phi Delta Kappan*, November 1, 2012, https://www.researchgate.net/profile/Ronald-Ferguson/publication/262091843_Can_Student_Surveys_Measure_Teaching_Quality/links/5c21304aa6fdccfc7066ffac/Can-Student-Surveys-Measure-Teaching-Quality.pdf.
29. Ferguson, "Student Surveys."
30. Bill and Melinda Gates Foundation, "Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project," MET Project Research Paper, December 2010, <https://usprogram.gatesfoundation.org/-/media/dataimport/resources/pdf/2016/12/preliminary-findings-research-paper.pdf>.
31. Jeannie Oakes, *Keeping Track: How Schools Structure Inequality* (New Haven, CT: Yale University Press, 1985).
32. Ryan Felton, "Black Students in US Nearly Four Times as Likely to Be Suspended as White Students," *The Guardian*, June 8, 2016, <https://www.theguardian.com/education/2016/jun/08/us-education-survey-race-student-suspensions-absenteeism>.
33. National Center for Education Statistics, "Racial/Ethnic Enrollment in Public Schools," last updated May 2023, <https://nces.ed.gov/programs/coe/indicator/cge/racial-ethnic-enrollment#:~:text=Between%20fall%202010%20and%20fall,from%2016%20to%2015%20percent>.
34. Rebecca Epstein, Jamilia J. Blake, and Thalia Gonzalez, "Girlhood Interrupted: The Erasure of Black Girls' Childhood," Georgetown Law Center on Poverty and Inequality, 2017, <https://genderjusticeandopportunity.georgetown.edu/wp-content/uploads/2020/06/girlhood-interrupted.pdf>.
35. Rhonda Nese, Joseph F. Nese, Connor McCroskey, Paul Meng, Danielle Triplett, and Eoin Bastable, "Moving Away from Disproportionate Exclusionary Discipline: Developing and Utilizing a Continuum of Preventative and Instructional Supports," *Preventing School Failure: Alternative Education for Children and Youth* 65, no. 4 (2021): 301-11, <https://www.tandfonline.com/doi/abs/10.1080/1045988X.2021.1937019>.
36. Carrie Spector, "Racial Disparities in School Discipline Are Linked to the Achievement Gap between Black and White Students Nationwide, According to Stanford-Led Study," Stanford Graduate School of Education, October 16, 2019, <https://ed.stanford.edu/news/racial-disparities-school-discipline-are-linked-achievement-gap-between-black-and-white#:~:text=The%20researchers%20found%20that%20a,achievement%20gap%2C%E2%80%9D%20Pearman%20said>.
37. Cami Anderson, "Boldly Breaking Patterns," *HuffPost*, January 16, 2015, https://www.huffpost.com/entry/boldly-breaking-patterns_b_6433984; Newark Public Schools, "The One Newark Plan, February Update, Frequently Asked Questions," <https://www.nps.k12.nj.us/wp-content/uploads/2014/08/One-Newark-Plan-FAQs.pdf>.
38. Donald Cohen, "Getting Education Reform Right," *Washington Post*, January 5, 2022, <https://www.washingtonpost.com/education/2022/01/06/getting-education-reform-right>.
39. Rob Dekanski, "Newark Economy: Top Industries, Biggest Employers, & Business Opportunities [2023]," RE/MAX blog, <https://www.newjerseyrealestatenetwork.com/blog/newark-economy>.

40. Martin Carnoy and Susanna Loeb, "Does External Accountability Affect Student Outcomes? A Cross-State Analysis," *Educational Evaluation and Policy Analysis* 24, no. 4 (Winter 2002), <https://journals.sagepub.com/doi/abs/10.3102/01623737024004305>.
41. FairTest: National Center for Fair & Open Testing, "'No Child Left Behind' after Two Years: A Track Record of Failure," January 2004, <https://nepc.colorado.edu/sites/default/files/EPRU-0401-55-OWI.pdf>.
42. "Newark Schools Open, but Some Parents Boycott over New Enrollment Program," CBS New York, September 4, 2014, <https://www.cbsnews.com/newyork/news/newark-parents-planning-to-boycott-first-day-of-school-over-new-enrollment-program>; Steven Sawchuk, "Plan to Revamp Newark District Hits a Nerve," *EducationWeek*, March 17, 2014, <https://www.edweek.org/leadership/plan-to-revamp-newark-district-hits-a-nerve/2014/03>.
43. Betheny Gross, Michael DeArmond, and Patrick Denice, "Common Enrollment, Parents, and School Choice: Early Evidence from Denver and New Orleans," Center for Reinventing Public Education, May 2015, <https://crpe.org/common-enrollment-parents-and-school-choice-early-evidence-from-denver-and-new-orleans>; Arianna Prothero, "Denver, New Orleans Offer Lessons on Streamlining Enrollment Between Sectors," *EducationWeek*, May 12, 2015, <https://www.edweek.org/leadership/denver-new-orleans-offer-lessons-on-streamlining-enrollment-between-sectors/2015/05>.
44. Naomi Nix, "5 Numbers to Know about Newark's Universal Enrollment System," NJ.com, April 15, 2015, https://www.nj.com/essex/2015/04/4_numbers_to_know_about_newarks_universal_enrollme.html; Kimberly Austin, Lucero Batista, Mahua Bisht, Andrew Karas, Jesse Margolis, and Andy Sonnesyn, "Newark Enrolls: A Principled Approach to Public School Choice," Columbia Law School, April 2018, https://cpri.law.columbia.edu/sites/default/files/content/newark_enrolls_report.pdf.
45. Austin et al., "Newark Enrolls."
46. Lia Eustachewich, "\$350K Shared Campus Grant Announced," Patch.com (Newark, NJ), October 17, 2011, <https://patch.com/new-jersey/newarknj/amp/2527548/350k-shared-campus-grant-announced>.
47. Margolis and Groves, "A New Baseline," 3; Linh Tat, "Cerf, Booker Hail Improvements to Newark Schools as District Prepares to Regain Control," *POLITICO Pro*, January 29, 2018.
48. Newark Board of Education press release, "Newark Public Schools Receives \$7 Million Head Start Grant from US Department of Health and Human Services to Improve Early Childhood Program Quality and Access," September 2, 2014, <https://www.nps.k12.nj.us/press-releases/newark-public-schools-receives-7-million-head-start-grant-us-department-health-human-services-improve-early-childhood-program-quality-access>.
49. State of New Jersey Department of Education, "Christie Administration Announces the Renewal of Newark Superintendent Contract," press release, February 26, 2015, <https://www.nj.gov/education/news/2015/0226new.pdf>.
50. Christine Campbell, "Measuring Up: How American Cities Are Shortchanging Black Students and What We Can Do About It," Center for Reinventing Public Education, October 2015, <https://crpe.org/measuring-up-how-american-cities-are-shortchanging-black-students-and-what-we-can-do-about-it>.
51. Margolis and Groves, "A New Baseline," 4.
52. Mark Chin, Thomas J. Kane, Whitney Kozakowski, Beth Schueler, and Douglas Staiger, "Assessing the Impact of the Newark Education Reforms: The Role of Within-School Improvement vs. Between-School Shifts in Enrollment," Center for Education Policy Research at Harvard University, October 2017, https://cepr.harvard.edu/files/cepr/files/newark_ed_reform_report_final.pdf.
53. Grover J. "Russ" Whitehurst, "The 2016 Education Choice and Competition Index," Brookings Institution, March 29, 2017, <https://www.brookings.edu/articles/the-2016-education-choice-and-competition-index>.

54. Margolis and Groves, "A New Baseline," 4.
55. Jeffrey R. Backstrand and Kristi Donaldson, "Post-Secondary Outcomes of Newark High School Graduates (2011-2016)," Newark City of Learning Collaborative & the School of Public Affairs and Administration, Rutgers University-Newark, 2018, <https://files.eric.ed.gov/fulltext/ED590999.pdf>.
56. Eleanor Fulbeck, Martyna Citkowicz, Candace Hester, and David Manzeske, "Newark Public Schools and Newark Teachers Union Teacher Contract Evaluation," American Institutes for Research, February 26, 2016, <https://www.air.org/resource/report/newark-public-schools-and-newark-teachers-union-teacher-contract-evaluation>.
57. Austin et al., "Newark Enrolls."
58. Annenberg Foundation, "The Annenberg Challenge," <https://annenberg.org/initiatives/the-annenberg-challenge>.
59. "How Many More Years of a Failed Reform Must New Orleans Parish Endure?" *New Orleans Tribune*, November 18, 2022, <https://theneworleanstribune.com/2022/11/18/how-many-more-years-of-a-failed-reform-must-orleans-parish-endure/#:~:text=Nearly%2054%20percent%20of%20the,A%20or%20B%20letter%20grade>; Susan Moore Johnson, "The Workplace Matters: Teacher Quality, Retention, and Effectiveness," National Education Association, July 2006, <https://files.eric.ed.gov/fulltext/ED495822.pdf>; Frank Adamson, Channa Cook-Harvey, and Linda Darling-Hammond, "Whose Choice? Student Experiences and Outcomes in the New Orleans School Marketplace," Stanford Center for Opportunity Policy in Education, 2013, <https://htv-prod-media.s3.amazonaws.com/files/stanford-study-1644255283.pdf>.
60. Sharon Lurye, "Kids' Reading Scores Have Soared in Mississippi 'Miracle,'" *PBS NewsHour*, May 17, 2023, <https://www.pbs.org/newshour/education/kids-reading-scores-have-soared-in-mississippi-miracle>.
61. "How Do School Funding Formulas Work?" Urban Institute, November 29, 2017, <https://apps.urban.org/features/funding-formulas>.
62. "Charlotte-Mecklenburg Board of Education Theory of Action for Change," revised March 2012, <https://www2.cms.k12.nc.us/boe/0809%20School%20Board%20Meetings%20and%20Schedules/Theory%20of%20Action%20for%20Change%20March%202012.pdf>.
63. Anna Maier, Julia Daniel, Jeannie Oakes, and Livia Lam, "Community Schools as an Effective School Improvement Strategy: A Review of the Evidence," Learning Policy Institute & National Education Policy Center, December 2017, <https://files.eric.ed.gov/fulltext/ED606765.pdf>.



Conclusion

Times Have Changed. The School System? Not So Much.

Margaret E. Raymond

Executive Summary

A Nation at Risk + 40 brought together twelve exceptional scholars and thought leaders to follow the National Commission on Excellence in Education's challenge to the nation by tracking subsequent efforts and effects since the release of *A Nation at Risk* in 1983. The resulting essays offer a unique comprehensive record of reforms and innovations designed to improve student learning in the United States.

The majority of the reforms focused on changes to system inputs, but half of these showed little or no impact. Reforms to processes such as site-based management and teacher professional development are more complex and therefore harder to study; the evidence favors reforms that are focused and sustained, though political pushback is more common with these types of reforms. Whole-system reforms attempt a comprehensive overhaul of both inputs and processes simultaneously or seek to provide alternatives to the system altogether; their results are positive for student learning and endure to the extent that they can weather or avoid backlash from entrenched interests.

As documented in a variety of measures, forty years of scattershot reforms have, on the whole, failed to improve student learning. Still, the catalog of activity amassed by the authors supports a few observations about future directions for elementary and secondary education in our country. We can characterize the historical record of reform efforts with six I's that serve as lodestars—albeit in the negative—with which to assess new proposals to deliver strong education to our nation's students:

- **Impulsive:** Most of the reforms were adopted at full scale based on similar efforts elsewhere—across an entire state or the nation—and many efforts to push programs across states or regions had roots in advocacy pressure to move reforms quickly.

- **Incremental:** The most pervasive attribute is the incremental nature of the interventions. As a result, they lacked the scope or initial scale necessary to drive needed system changes.
- **Incoherent:** Beyond the inadequacy of incremental change as a lever for system change, most of the changes undertaken over the past decades were launched with no consideration of how the reform would interact with the rest of the K-12 system.
- **Impatient:** A separate issue that permeates the essays is the (often unstated) expectation that improvement efforts produce large, demonstrable results almost immediately and without regard to the differing time requirements of the changes being made.
- **Intransigent:** The possibility exists that the summative effect of all the efforts over the years has led to reform fatigue and fostered a resiliency to *any* improvement efforts—an adaptive state of resistance to change of its core activities.
- **Ineffective:** Even before the blow to student learning of COVID-19 school closures, the long-run reports noted that US student performance was stagnant or in decline, which we appear to tolerate with indifference.

• • •

In 1983, the National Commission on Excellence in Education (NCEE) released *A Nation at Risk* (ANAR), which issued a wake-up call, named the state of US education a crisis, and presented thirty recommendations for action. It bears noting that the Commission’s recommendations were targeted in focus and scope, leaving the prevailing “one best” district-based education model intact. We will never know whether larger-scaled interventions were considered or not. Whatever the genesis, the final recommendations left education policymakers with an organizational checklist, and as the essays in this series have demonstrated, they responded accordingly.

A Nation at Risk + 40 brought together twelve exceptional scholars and thought leaders to review the nation’s response to the Commission’s challenge. At the outset of this research collaboration, compiling the record of forty years of school improvement efforts and summarizing the available evidence of their respective impacts on student outcomes appeared straightforward, if even a bit tedious. It turned out to be anything but that.

Each of the twelve essays fulfilled its assignment. In each strand of investigation, the authors documented the evolution of improvement activity and—where it exists—described the degree to which the efforts paid off. On its own, every one of the essays makes an important contribution to our ongoing national conversation about the critical state of the public K-12 education sector. While we make no claim that the scope of inquiry was definitive, the separate reviews cover billions of dollars in major programs and initiatives pursued by districts, states, and

philanthropy. Many of these initiatives were incentivized by Congress and span Republican and Democratic presidential administrations. Our authors offer their own recommendations that, if followed, hold promise to improve conditions in the spheres they examined.

The research collaborative delivered an even more valuable asset, as the result is far more than the sum of the parts. Until the essays were gathered into a collection, the aggregate record of attempts to improve the K-12 education system in the United States was uncharted and unrecognized. We know of no other compilation that illuminates the sheer breadth of reform activity.

For the first time, we can compare the impacts across different areas of investment. Beyond this, taking the full collection as a whole augments the strand-specific recommendations with several crosscutting observations to inform future action.

WHAT DID WE DO?

There can be no dispute that, as a nation, we certainly tried hard to fix the problem. Practically speaking, we addressed every node that was mentioned by the Commission and several that weren't. It is remarkable how doggedly educators, policy leaders, advocates, and funders have augmented policy and practice with interventions. The sheer volume and spread of reform efforts are worth examining, as they begin to shed light on the situation we currently face in public K-12 education.

Other scholars (Hattie 2023) have used evaluations and other research to rank the impact on student performance of various reforms. The impact estimates are drawn from a vast collection of meta-analyses, yielding a super-meta-analysis that rank-orders reported results across different interventions. The rankings are widely interpreted as the definitive, adjudicated, and authoritative guide to improving student performance. In statehouses, state education agencies, and school districts, the rankings have taken on mythic proportions in guiding policy decisions about school improvement.

It is easy to see the appeal. The aim is noble, and the appetite is intense. Sadly, deeper inquiry into the rankings shows significant problems with the work: the desire to be expansive sits in tension with the need to apply stringent criteria about which meta-analyses are fed into the rankings. We learned that the underlying quality of the reform interventions themselves and the rigor of the research about their effects varied widely. To illustrate with a hypothetical: in the rankings, one thousand low-quality interventions with medium-strength evidence receive higher weight than one hundred high-quality interventions with a high-quality evaluation.

The concerns go beyond the problem of the quality of evidence. The implication for policy-making and educator practice is that the rankings encourage devotion to one or two marginal adjustments to schooling at the expense of lower-ranked options. The greatest risk lies in overlooking emerging successes for years until the next update to the rankings occurs.

Wishing to avoid a similar result, we chose a different approach to exploring the body of evidence. Beyond the notable volume of reform efforts attempted over the past forty years, it is useful to consider the points of the system that the various reforms were designed to change. This is important because many of the checklist items from ANAR’s recommendations aim at strengthening only one facet of the K-12 system, and the Commission did not offer recommendations on mixing, matching, or stacking multiple reform efforts.

The stability of the basic model of US K-12 public education over four decades is advantageous for our purposes because it supports a generalized theory of action, sometimes called a “logic model.” Theories of action specify the types of capital, staffing, and other resources that are needed to provide K-12 education. Theories of action also detail the policies and practices that are followed. Inputs and processes combine to produce a near-term result referred to as “outputs.” The eventual value of the results is identified as “outcomes.” With this lens, we classify the policies, programs, and initiatives discussed by the essay authors in order to learn about the targets and yields of reform activity. To be clear, some improvement efforts span our classification categories (e.g., some professional development includes input and process features); these are assigned by their most prevalent attributes.

Our authors are highly sensitive to the availability and caliber of research and evaluation. In many areas, such as public school choice and inclusion of master teachers in educator preparation programs, no evidence exists. In other areas, impact information is hindered by studies involving few examples, fuzzy specifications, or weak counterfactuals. Evaluative studies of school-based health centers and socio-emotional learning are examples where evidence of impact is lacking. The field of impact studies has evolved in constructive ways, but it still hinges critically on a weak commitment to objective assessment of impacts and the discipline to incorporate the insights into practice.

INPUTS

A preponderance of the improvement efforts identified by the authors sought to adjust the inputs used by the education system. These include teacher-focused efforts such as alternative certification and incentive pay arrangements, adding school-based health centers, strengthening early childhood programs, and overhauling curriculum. System-focused input changes seek to expand the variety of inputs or the overall structure of the system, whereas marginal input reforms seek to improve the quality of the selective resources within the existing stock.

Taken together, these efforts aimed to enrich the ingredients in the “recipe” for K-12 education. Focusing reform attention on adjusting the quantity, quality, or intensity of a factor *before* it is used keeps the reform at arm’s length from the actual production of education. Think of upgrading tires on a race car—the improvement to the equipment takes place offline and then is brought online in the hopes of improved performance.

The evidence shows that the range of impacts for inputs-focused reforms run from zero to as much as three-quarters of a year of additional achievement for students. About half the input reforms have negligible or no effect on student academic achievement. The options that show no impact share the attribute of shallow or isolated treatment—a few hours of professional development or play-based preschool. For both system-focused and marginal input reforms, positive results point to interventions that have significant weight, scale, and duration to create and sustain the momentum for change. As examples, we see this in the small-schools movement (systems focused) and in laser-focused teacher professional development (marginal adjustments).

Input reforms assume that the rest of the system will respond organically to the change in the treated input. As the evidence shows, many efforts provided too little leverage to lift the rest of the operation. Worse, an exclusive input focus ignores the possible interactions with other components that may react in different ways than expected.

PROCESSES

Process reforms aim to change the way education is created, delivered, and monitored by schools and their oversight bodies. To extend the recipe analogy, processes are the mixing and cooking instructions. Marginal process reforms attempt to mix inputs in new ways or interact inputs with new policies or protocols. Systemwide process changes try to ubiquitously reengineer old ways of doing things to produce better results, such as the experience of adopting the IMPACT teacher evaluation and compensation initiative in Washington, DC, or implementing a digital learning platform across all the middle schools in a district.

Given the challenges of designing and implementing new programs, it is little wonder that our authors found fewer process reform examples in their scans. Across the essays, the authors identified three general areas of process reforms.

Teacher professional development falls largely into the process category—selected areas of knowledge and skills are targeted to expand the capacity of teachers to perform their duties. This differs from input reforms, which are directed toward improving the number or quality of candidates at the point of hiring. The available evidence suggests that for much of the past forty years, there was little or no effect from a large proportion of professional development. Recent evidence, however, shows positive impacts when the programs are strictly focused, multifaceted, and sustained, producing between one and four months of extra achievement.

Incentive programs for higher teacher performance have strong impacts on student academic achievement for their duration, from about two months to an extra year of added achievement. However, these impacts are largely one sided; they did not induce low-performing teachers to move up or move out. Rather, they provided financial and work assignment flexibility incentives for teachers. Similar programs that trade extra compensation for teaching in the most

challenging settings also produce strong student gains of similar magnitudes. Both types of reforms are highly vulnerable to political disruption at all points of the program, especially if teachers' participation requires evaluation of their performance.

Technology adoptions can also be classified as process reforms. Once technology has been purchased and distributed, it serves a process function. The evidence of impact from the broad provision of education technologies has, for the most part, been disappointing, showing no impact and substantial stranding of investments. Despite that general trend, however, a number of significant and strongly positive examples of technology-supported education have emerged as promising proof points.

The third area of process reforms occurs at the governance level of the system. Since ANAR's release, states have changed the way they fill key positions on their boards of education and within the Council of Chief State School Officers. The change in appointment mechanisms is a process change whose influence is systemwide. Likewise, changes in district school boards to a portfolio management model also flow across the district system. The evidence on these governance changes has been mixed.

It is clear that important differences exist between systemwide process changes and those that are marginal in nature. Some process reforms can work only if introduced systemwide, such as adoption of student safety protocols or school-based disciplinary programs; a "half a loaf" approach won't work. Alternatively, marginal process change can be narrow in scope, in terms of either the focus of the reform or the organizational level that is targeted. Pilot programs are a clear example. In marginal process reforms, the rest of the schooling equation remains untouched. The balance between systems and marginal processes can shift either way depending on the interplay of cost, the scope of the planned innovation, friction with adjacent policies or practices, and political resistance.

Moreover, estimating the effects of process changes is technically and practically more difficult than measuring the effects of input shifts. The interactions of new processes with other factors and their dynamic nature over time create complexity that is difficult to measure. The body of evidence is therefore smaller than exists for input-focused changes. New instructional models such as discovery or expeditionary learning are process changes. The evidence on these is thin, except for personalized learning modalities, which show strongly positive effects on learning gains and graduation rates.

Likewise, the expansion of technology—equipment, connectivity, and content—in schools is a process change that has altered the way curriculum and instruction are organized and deployed. The impacts are sobering: unused resources cannot advance learning, but where strong implementation occurs, we also see improved student academic achievement.

The final set of process changes can be grouped as "infusion" efforts. Extended school years appear not to improve student results, but additional time in focused instruction helps; the extra time matters only if it is used well. Similarly, teacher and leader professional learning

programs are seen as a mixed bag. As with extra time in school, the evidence shows that focused and targeted experience can produce positive impacts on student learning, but those conditions do not appear to be the norm.

Although they have a smaller evidence base, process reforms deal with larger segments of the education enterprise than inputs. Those that work share the attribute of internal design coherence, even if they do not fit well into the rest of the system. Finally, the larger the process reform, the more of a political target it offers to opponents.

OUTPUTS

When we consider the near-term results of elementary and secondary education or the milestones on the way to reach these results, we are discussing outputs. These are the immediate products that reflect the end state that inputs and processes have created. In K-12 education, common outputs include meeting learning benchmarks for grade promotion, satisfying graduation requirements, and implementing performance measures for teachers and leaders. It bears noting that outputs are agnostic to inputs and processes: many combinations are possible to create a particular output.

Systems-oriented improvement efforts have been judged by both outputs and outcomes. In Cami Anderson's essay on the results of districtwide reform strategies in Newark, New Jersey (chapter 12), early childhood enrollment increases of 35 percentage points were one output. Another was the rise of 20 points in the percent of Black students enrolled in above-average schools, followed by significant early gains in reading achievement and eventual gains in math. Ironically, the impressive improvements in Newark were not tallied to be a successful outcome, largely because of friction in the community and with elected leaders. Similar efforts under the US Department of Education School Improvement Program did not create positive results.

There are other examples of reforms that aim to change outputs. Redirecting school board activity to prioritize academics and student learning has been shown to produce positive movement on outcome measures for schools and districts.

The largest efforts to move outputs of elementary and secondary schooling lie in the national adoption of accountability programs. The consequential approach to school-based accountability advanced by the No Child Left Behind Act (NCLB) improved learning by one half per year of student achievement and narrowed achievement gaps between groups of students. High school graduation rates increased by 15 percentage points with concomitant increases in college enrollments. These improvement trends persisted through 2015, but they have all but reversed over the past eight years, with student learning falling dramatically over the course of the COVID-19 global pandemic.

Other efforts to affect teacher preparation programs also looked at outputs, but to no avail: current teacher certification exams are unable to predict future variations in teachers' performance once they are in the classroom. Other common indicators, such as academic credentials or years of experience (also inputs), are similarly disconnected from future teacher performance.

Finally, some reform activities deliberately circumvent mainstream institutions and channels in an attempt to create better outputs. Extra-system initiatives can take the form of inputs or processes, or they can combine the two. Some options that have shown positive impacts for student results include mayoral control (significant gains in achievement and better fiscal controls) and gubernatorial appointment of state board members (better performance on the National Assessment of Educational Progress assessments).

As noted by other scholars, school choice can arise within, across, or outside of school systems (Lake 2020). Intradistrict school choice redistributes seats in schools by changing the way students are assigned to schools; it aims to improve the outputs for the students who access better classrooms. As a process reform, it is associated with stronger achievement in math for minority students. Interdistrict choice is rare, and its effects are not well studied. Charter schools operate in a separate policy stream and deliver stronger growth and achievement in reading and math, especially in urban charter school networks (CREDO 2023). For vouchers, the impact for students on balance has not been positive; the evidence on vouchers shows weaker achievement for enrolled students even as they create positive spillover impacts on public schools. Other efforts that move outside the usual institutional arrangements are less understood. Newer options such as education savings accounts (ESAs) and microschoools have yet to be examined in depth.

OUTCOMES

In an education theory of action, outcomes are the final results of the entire enterprise. Outcomes differ from outputs because they apply external standards and criteria to the nominal outputs to make judgments about what is "good enough." So, while outputs may be expressed as test scores, CTE credentials, or course completions, when we apply evaluation standards such as postsecondary readiness, we are making judgments about the performance that was produced.

Since *ANAR* was released, we have gained clarity, if not conviction, about what we intend our schools to produce. Performance frameworks that illustrate the results that stakeholders deem desirable have grown in number and complexity. Across the country, charter school authorizers and state and local school boards use performance frameworks as central elements of school and district oversight and accountability. Newer examples of our collective expectations are seen in the work in some states to define the profile of a graduate, setting explicit criteria for what a high school diploma should represent.

By law, every state reports publicly on how its students and schools are performing. State-issued “report cards” for districts and schools generally include demographic information for teachers and students, operational and financial information, and student academic performance information. States set thresholds for student and school performance expectations, though these thresholds vary a lot. Whatever their aspirations, we are not in vastly different territory today than in 1983. Disappointing outcomes (e.g., high school math performance) have even prompted attempts to improve the optics by diluting some of the criteria (such as watering down the instructional frameworks or course requirements), but such maneuvers do nothing to alter the underlying reality.

INSIGHTS FROM THE EVIDENCE

As Walt Kelly’s cartoon character Pogo said, “We have met the enemy and he is us.” Indeed, the staggering array of treatments, interventions, redesigns, and innovations that our authors identified makes it a challenge to rationalize our collective experience into any semblance of order. If we had aimed for chaos at the outset, it is hard to imagine a better result.

Despite the cacophony, the catalog of activity amassed by the authors supports a few observations about our forty-year effort to reform that hold potential for illuminating future directions for elementary and secondary education in our country. After identification, we can characterize the record of reform efforts with six I’s: impulsive, incremental, incoherent, impatient, intransigent, and ineffective, as discussed below.

IMPULSIVE

Most of the reforms were adopted at full scale—across an entire state or the nation. Many efforts to push programs across states or regions had roots in advocacy pressure to move reforms quickly. Many state leaders were game to bring new policies to their state if they were perceived as having been successful elsewhere, as it reduced the perception of risk and provided an existing model to copy.

Doing the “here, too” dance hobbled the new adopters in two ways. It skipped over analysis of the “fit” of the reform in the local context—and the important variation in local contexts—on the receiving end. It is impossible in hindsight to determine how many of the “mixed result” outcomes stemmed from differences in the settings on the ground, but it seems safe to say local contours were likely overlooked as most of the programs or policies were advanced. It is also true that jurisdiction-wide adoption curtailed the ability to evaluate implementation and impacts in real time, so valuable learning was lost at the get-go.

INCREMENTAL

The most pervasive attribute is the incremental nature of the interventions. This stems in part from the original recommendations of the ANAR Commission, framed as commonsensical

and achievable changes. The commitment to incrementalism continued even when earlier efforts proved ineffective. One might argue that it made sense to aim small to soften implementation friction. The record suggests otherwise. Because the interventions were mostly narrowly focused, not only did they lack the scope or initial scale necessary to drive needed system changes, but in their sheer volume—so many reforms in so many areas—they led to a reform fatigue that lasts to this day.

It is important to note that the essays identified examples of successful reform that did not involve incremental adjustments. Systemwide efforts as described for Newark and new systems building as seen with charter schools have larger blueprints and therefore greater areas for change.

INCOHERENT

A third observation is that most of the changes undertaken over the past decades were launched with no consideration for how the reform would interact with the rest of the K-12 system. Changes to piece parts were designed and adopted as autonomous endeavors. This partially explains why many innovations fail to scale effectively.

This does not mean that things were only tried one at a time. Many examples exist of multiple incremental reforms launched simultaneously without an understanding of the interplay between them or with the rest of the equation. Reforms were “bolted on,” one after another, without regard for how they fit together. And each one that was added “diluted” the impact of the others. The resulting lack of coherence often led to unintended consequences that were never even considered, much less planned for.

One important implication of incoherence is a lost opportunity to ensure that stakeholders—especially the ground-level personnel—function with an understanding of the way the system works and how they belong in it; a well-crafted plan of action can provide that. A second implication is that it is difficult to objectively learn from experience, especially from unsuccessful ventures. When the general model is unorganized, it is hard to assign causality, for example, between lack of implementation fidelity of a sound design and a design that does not fit the context it is meant to improve.

IMPATIENT

A separate issue that permeates the essays is the (often unstated) expectation that improvement efforts produce large demonstrable results almost immediately and without regard to the time requirements of the change being made. Changes to organizational culture need to occur rapidly, but other changes take time. Shifts in instructional methods often require more than a single year to stabilize enough to know how well they work. Incorporating new systems such as new-teacher onboarding can take even longer to reveal their true value and impact.

The expectation of quick results creates multiple harms. It doesn't give the good parts time to take root or provide the space to iterate toward success. Moreover, it seeds unrealistic expectations about the diligence needed to give new approaches their due. From a political vantage, it gives the doubters and pouters a head start on declaring new reforms a failure. It also contributes to the "carousel," as one teacher described it: "I don't have to do anything but wait—in three years there will be something new."

Compounding the problem, the governance side of the equation needs strong and enduring leadership to be patient with complicated, multifaceted reform efforts and to plan and invest for the long term. Even if the enabling conditions are understood and a proven scaling strategy is in place—such as with charter management organizations—when the reform in question needs ten to twenty years to come to fruition, rapid turnover cycles of education leaders lose important institutional knowledge, and politicians are short on patience (or incentive) to see it through.

All too often, the time needed to see results is longer than the amount of time politicians have in their seats, and it does not line up with the cyclical campaign and election cycle. Short-run wins are coveted by political actors seeking to establish a record of success on which to build advancement. The bias toward quick returns and the lack of political will or appetite to invest in long-run solutions have a serious trickle-down effect: (1) a constant churn of reform that does not give space or time to realize success and (2) systems that learn to wait out the current wave of reforms, as "this, too, shall pass." When the need for improvement is glaring but the actors in legislatures and education agencies prioritize their own short-run interests, we face compound system failure.

INTRANSIGENT

The authors carefully identified examples of reforms that produced positive student learning impacts, but many were subject to political interference or failed to perform at scale. Still, the examples show what may be possible. What they do not show is the complementing picture of the myriad reforms that went nowhere and evaporated into history. There is no tally of their number.

But anecdotal reports have consistently told the story of reform churn. Charles Payne's phrase, "So much reform, so little change," seems to apply. Instead of forty years of sustained and coherent reform, we have forty short-run reforms that each last three years. School teams are introduced to new practices during the professional development days that accompany the start of school each fall, with short windows of time to prepare for deployment and little implementation support during the year. The school teams learn about impacts indirectly—and often too late to try modifications. Decisions about continuing or terminating the effort usually do not include input from those on the front line. More often than not, new initiatives are quietly abandoned, with the cycle left to repeat itself the following year.

It is notable that, despite this endless churn of reforms, the prevailing institutional structure of “SEA, LEA school board, district administration, school leadership, grade/class grouping, teacher” remains largely unchanged, despite repeated pressures on it to adapt. The possibility exists that the summative effect of all the efforts over the years has fostered a resiliency to *any* improvement efforts—an adaptive state of resistance to change of its core activities. It may help to explain the tendency to shift focus to other facets of students, teachers, or teaching where ground may be more fertile for positive experience. There is no way to test this idea empirically, but it fits the pattern of the evidence and explains the abundant cynicism and burnout.

INEFFECTIVE

The strongest case for learning from our experience lies in our national trends on student performance. Given the authors’ reports, it is little wonder that, even before the blow to student learning of COVID-19 school closures, the long-run reports noted that US student performance was stagnant or in decline.

Two considerations help to explain our current state. Part of the problem is that, apart from formal pilots, most reforms launch without considering how to learn from them. We are seriously underresourced across the sector in measuring local conditions and reform effectiveness.

In addition, even after forty years, the system has significant internal inconsistency—it lacks a “unified theory” of how reform should be done. This essay collection recounts how many reforms were launched without a sufficient discussion of which level of the system (e.g., state, district, school) might be the most effective to lead the transformation efforts.

CONCLUSION

We face an even more daunting challenge today, which is that forty years of reform have exhausted *everyone* involved. The one thing we may have conclusively proven is that the system, as presently constituted, has been resilient to reforms at scale. A modern ANAR report might not fall on deaf ears—the need for school reform is real—but it would fall on ears that are tired of hearing about it.

What is clear is that we have a thin collection of reforms that have been shown to work and that can scale. None of the proven reforms seek to integrate with other proven reforms to concentrate their success. The larger the scale of innovation/reform, the larger the political target it presents for opponents of change.

What we *do* have is an impressive record of what *not* to do. We can’t assume that ideas that have been proven effective in one setting will be effective in every setting. We can’t expect change at the margins (no matter how well they are done) to be able to leverage an entire school model. We can’t impose reforms that ignore how the change affects other parts of the

enterprise. We should accept these lessons as a form of learning in itself and perhaps the best final message of this exercise. Drawing on the six I's—impulsive, incremental, incoherent, impatient, intransigent, and ineffective—may provide lodestars by which to assess new proposals toward more effective approaches to delivering strong education to our nation's students.

REFERENCES

Center for Research on Education Outcomes (CREDO), "As a Matter of Fact: The National Charter School Study III 2023," June 16, 2023.

Hattie, John. *Visible Learning: The Sequel*. New York: Routledge, 2023.

Lake, Robin. "When School Districts Let Parents Choose." Hoover Education Success Initiative, Hoover Institution, July 2020.

National Commission on Excellence in Education (NCEE), *A Nation at Risk: The Imperative for Educational Reform* (Washington, DC: US Department of Education, 1983).

ABOUT THE AUTHORS

CAMI ANDERSON

Cami Anderson, CEO of ThirdWay Solutions, supports leadership teams across sectors to achieve goals in pursuit of racial equity and social justice. Superintendent of schools for nearly a decade, Anderson was named by *Time* as one of the world's one hundred most influential people. She received a legacy achievement award from the King Center for her work in racial justice.

ERIC BETTINGER

Eric Bettinger is a senior fellow at the Hoover Institution and the Conley-DeAngelis Family Professor in the School of Education at Stanford University. He is a research associate at the National Bureau of Economic Research. His research interests include the economics of education; student success in college; and the effects of school choice on both academic and nonacademic outcomes.

STEPHEN L. BOWEN

Stephen L. Bowen is the executive director of the Hoover Education Success Initiative (HESI). He began his career in education as a middle and high school social studies teacher, is a former state legislator, and served as the Maine Commissioner of Education. He came to Hoover from the Council of Chief State School Officers, where he was deputy executive director.

THOMAS S. DEE

Thomas S. Dee is the Barnett Family Professor at Stanford University's Graduate School of Education and a senior fellow (courtesy) at the Hoover Institution. Dee is also a senior fellow at the Stanford Institute for Economic Policy Research and a research associate with the programs on education, children, and health at the National Bureau of Economic Research.

MARIA D. FITZPATRICK

Maria D. Fitzpatrick is professor of economics and public policy in the Brooks School of Public Policy at Cornell University. She is codirector of the National Data Archive on Child Abuse and Neglect, research associate at the National Bureau of Economic Research, and an affiliate in the CESifo Research Network. Her research focuses on child and family policy, particularly education.

MICHAEL HANSEN

Dr. Michael Hansen holds the Herman and George R. Brown Chair in Governance Studies and is a senior fellow in the Brown Center on Education Policy at the Brookings Institution. A labor economist by training, he has conducted original research on a wide array of education policy issues, with a specialization in K-12 teacher policy.

ERIC A. HANUSHEK

Eric A. Hanushek, the Paul and Jean Hanna Senior Fellow at the Hoover Institution, is internationally recognized for his economic analysis of educational issues. Recipient of the 2021 Yidan Prize for Education Research, he has authored or edited twenty-six books. He is a distinguished graduate of the US Air Force Academy and completed his PhD in economics from MIT.

MICHAEL T. HARTNEY

Michael T. Hartney is an associate professor of political science at Boston College, a Hoover Fellow at the Hoover Institution, adjunct fellow at the Manhattan Institute, and research affiliate at the Program on Education Policy and Governance at Harvard University. He received his PhD in political science from the University of Notre Dame.

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*.

ROBERT PONDISCIO

Robert Pondiscio is a senior fellow at the American Enterprise Institute, focusing on K-12 education, curriculum, teaching, school choice, and charter schooling. He previously worked at the Thomas B. Fordham Institute and the Core Knowledge Foundation and is the author of *How the Other Half Learns*. His became involved in teaching and education policy after a twenty-year journalism career.

ABOUT THE AUTHORS

MARGARET E. RAYMOND

Margaret “Macke” Raymond is the director of the Center for Research on Education Outcomes (CREDO) and program director for education at the Hoover Institution of at Stanford University. CREDO analyzes education reform efforts around the country and is a leader in studying US charter schools, school reform policy, measurement and performance, and accountability systems and incentives. Dr. Raymond holds a PhD in political science from the University of Rochester.

JOHN D. SINGLETON

John D. Singleton is an associate professor of economics at the University of Rochester and a faculty research fellow of the National Bureau of Economic Research. His work on school choice has appeared in leading economics journals. He was previously a dissertation and post-doctoral fellow of the National Academy of Education. He received his PhD from Duke University.

DEBORAH STIPEK

Deborah Stipek is the Judy Koch Emeritus Professor of Education and the former dean of the Graduate School of Education at Stanford University. Her doctorate is from Yale University in developmental psychology. She currently works closely with the California Department of Education on early childhood policy and with California Education Partners helping school districts develop preschool-third grade coherence.

TOM VANDER ARK

Tom Vander Ark is an advocate for innovations in learning. As CEO of Getting Smart, he is an advisor to schools, foundations, and learning organizations. The author of numerous books and blogs, Vander Ark previously served as the first executive director of education for the Bill & Melinda Gates Foundation and was a public school superintendent in Washington State.



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 authors 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 EDITORS



STEPHEN L. BOWEN

Stephen L. Bowen is the executive director of the Hoover Education Success Initiative (HESI). He began his career in education as a middle and high school social studies teacher, is a former state legislator, and served as the Maine Commissioner of Education. He came to Hoover from the Council of Chief State School Officers, where he was deputy executive director.



MARGARET E. RAYMOND

Margaret “Macke” Raymond is the director of the Center for Research on Education Outcomes (CREDO) and program director for education at the Hoover Institution of at Stanford University. CREDO analyzes education reform efforts around the country and is a leader in studying US charter schools, school reform policy, measurement and performance, and accountability systems and incentives. Dr. Raymond holds a PhD in political science from the University of Rochester.

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.

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

