School Accountability—Past, Present, and Future

A POLICY ANALYSIS FROM THE HOOVER EDUCATION SUCCESS INITIATIVE

by Chester E. Finn, Jr.

A PUBLICATION OF THE HOOVER INSTITUTION

A POLICY ANALYSIS FROM THE HOOVER EDUCATION SUCCESS INITIATIVE

School Accountability—Past, Present, and Future

FINDINGS AND RECOMMENDATIONS FOR STATE AND DISTRICT POLICYMAKERS

CHESTER E. FINN, JR.

November 2020

Hoover Education Success Initiative

EXECUTIVE SUMMARY

American public education has been grappling with school accountability for several decades, propelled both by federal mandates and by state and local leaders bent on boosting student achievement; narrowing achievement gaps; advancing transparency, equity, and excellence; and revving school effectiveness. The country is now five years into accountability's "ESSA (Every Student Succeeds Act) phase" and it's time to examine how that's working, how it can be made to work better, and what may lie ahead. That's what this paper does.

Part I recaps the evolution of accountability over the past half century. Where did it come from and why? Why was it important then, and why is it still important?

Part II unpacks core issues surrounding school accountability and appraises the evidence to date regarding its efficacy. It concludes that well-crafted, outcomes-based accountability systems that rely on solid data and incorporate consequences generally lead to stronger achievement.

Part III describes changes wrought by 2015's Every Student Succeeds Act, delves into how states are responding to it, and makes recommendations for maximizing its value. Similar to the Council of Chief State School Officers' "Roadmap for Next-Generation State Accountability Systems," we recommend that a state's ESSA plan include these key elements:[1]

- *Supply transparent information*, readily accessible to the public, including clear summative ratings for schools.
- *Provide comparisons* by which school scores are placed in context, that is, viewed alongside those of other schools and districts within the state, with other states, and perhaps internationally.
- *Emphasize measures that are amenable to gains* and primarily within the school's control, including achievement and achievement growth.

- *Employ indicators that are hard to manipulate.*
- *Ensure the transparent treatment of key subgroups.*
- *Report both proficiency and growth, equally weighted.*
- *Carefully select and deploy valid growth measures,* whether criterion- or norm-referenced.
- Insofar as it can be accurately and fairly tracked, *take account of student absenteeism,* particularly chronic absenteeism.

> The testing hiatus and "data hole" resulting from school closures and federal waivers in spring 2020 will make calculation of year-to-year achievement growth difficult or impossible in the short run. Variability in school operations and attendance during the 2020–21 year will also bedevil such calculations. Schools may have to substitute two-year growth (i.e., spring 2019 to spring 2021) as best it can be gauged, provided of course that they remain assiduous about 2021 assessments. The unevenness of data and fragility of calculations based on them may lead some states to suspend their summative school ratings for a year. We urge careful improvisation in the near term and a resumption of familiar calculations, comparisons, and ratings as soon as possible.
>
> We agree with the Data Quality Campaign, the Alliance for Excellent Education, the Collaborative for Student Success, the Education Trust, and a host of civil rights organizations that "states can and should continue to measure student growth in 2021. . . . By measuring student progress between the 2019 and 2021 annual assessments, state leaders can still get the vital insights they will need to understand and continue to support student learning."[2]

States should consider issuing both data-rich (but easily understood) report cards and more comprehensive information "dashboards" for individual schools.

Because ESSA does not prescribe specific consequences for poorly performing schools, state leaders should understand that without careful implementation of their accountability plans *and* deployment of well-considered consequences for such schools, little improvement is apt to occur.

Part IV looks to the future of school accountability, conscious that resistance to testing in particular and consequential school accountability in general have led some elected officials and education leaders to ease back on such things. We urge state (and federal) officials not to forsake results-based accountability or to shun high-quality assessments of student

learning as an indispensable source of essential information as to where and how well those results are being achieved. To forgo them would cause K–12 education in America to fly blind, akin to a plane in the fog without instruments.

"States can and should continue to measure student growth in 2021. By measuring student progress between the 2019 and 2021 annual assessments, state leaders can still get the vital insights they will need to understand and continue to support student learning."

In Part IV, we assume that ESSA is not immortal and that, after five years of experience with it, we should identify elements of a better approach to school accountability, including, for example, a somewhat changed assessment regimen and fresh thinking about consequences. We believe that tomorrow's accountability systems should be geared less to short-run gauges of "proficiency" and more to students' true readiness for college, career, citizenship, and adulthood. Here, in summary form, are our key recommendations:

Assessment

- *Kindergarten readiness*. Although this measure ought not be used for elementary school accountability, entering kindergartners should be assessed on their readiness to succeed in school, including but not limited to their preparedness to undertake reading and arithmetic. Such information creates an essential baseline for all that follows.

- *English language arts (ELA) and math prowess*. States should test students' ELA and math prowess at least in grades two or three, four, six, and eight, with the fourth- and eighth-grade assessments coincident with National Assessment of Educational Progress (NAEP) testing. Annual testing yields more precise calculations of achievement growth, but states may prefer to lighten the annual testing burden in light of the additions recommended below.

- *End-of-course exams*. States should add capstone (end-of-course, or EOC) exams in other core subjects during middle and high school, aligning these with high-quality curricula in those subjects, including career and technical subjects as well as the traditional academic core.

- *Diplomas that mean something*. With EOCs in place for key high school courses, requiring that they be passed at a satisfactory level becomes an excellent way of ensuring that diplomas attest to actual accomplishment.

For accountability purposes at the high school level, states should weigh both a school's success in getting all students to the passing level on the EOCs (and thence to graduation) and also its success in getting as many as possible to the college/career readiness threshold and beyond.

Analysis and Reporting

Achievement and growth should remain the core criteria by which school performance is evaluated. We recommend continuing with ESSA-style data gathering on student learning, based primarily on external exams developed by states and aligned with their academic standards. Results should be disaggregated by student group and reported at the school, district, and state levels, as should growth whenever that can be calculated. Schools should continue to be assigned ratings based on their performance, and easily grasped information about that performance should be made public on websites, report cards, and dashboards.

We recommend continuing with ESSA-style data gathering on student learning, based primarily on external exams developed by states and aligned with their academic standards.

Consequences and Capacity

What to do about troubled schools? Two approaches—and combinations of them—deserve consideration. Both depend on quality data and transparency about student and school performance, but they point in different directions. One strives to improve upon the familiar thrust of federal policy, which is to intervene in low-performing schools with an eye to making them better. The other is to stop trying to "fix" troubled schools and instead rely on informed choices by parents to place their children in better schools.

Particularly in view of mounting resistance to test-based accountability of the traditional kind, states should also attend to issues of school and district capacity (as well as their own) and adopt strategies to assist troubled schools in making the improvements they need.

INTRODUCTION

As American K–12 education lives with the five-year-old Every Student Succeeds Act (ESSA) and its decentralized approach to school accountability, criticism floods in from many directions: it isn't working; it's just No Child Left Behind (NCLB)–lite, with many of the same shortcomings; its weak-kneed devolution to states lets schools off the hook and leaves many kids behind; it's part and parcel of America's oft-criticized reliance on standardized testing, now disrupted anyway by COVID-19; it's an invitation to squishy metrics and faux interventions that conceal and excuse failure; and, perhaps most often, this sort of test-driven accountability was never a sound strategy for improving American education, which instead needs more TLC, ampler resources, better-paid teachers, and renewed trust in the professional judgment of educators.

Is accountability dead? Will the next round of federal legislation bury its remains? Or is the plague-induced testing pause a great opportunity to take stock of lessons from the past, to review how ESSA is or isn't working, and to consider what the future may and perhaps should hold? Parts III and IV of this paper examine today and tomorrow. First, though, Part I recounts the multidecade saga that led to ESSA, and Part II surfaces enduring issues that beset almost every discussion of accountability and reviews research on its effectiveness.

PART I: HOW WE GOT HERE

Once upon a time, "accountability" had little to do with whether children actually learned anything in US schools. In fact, the term was scarcely ever encountered in the K–12 context. Insofar as schools had to "account" for anything, it meant operating as intended, spending resources as budgeted, complying with laws and rules as expected, and generally doing what they were supposed to. Americans pretty much took for granted that if their schools did those things diligently, their pupils would emerge with whatever learning was needed to go on to college or get a job and function as citizens. They would have been schooled, which meant they were educated.

Those comfortable assumptions began to change in the mid-1960s. US Education Commissioner Francis Keppel, greatly aided by the Carnegie Corporation of New York, recognized the need for better information about student achievement and put into motion what emerged by decade's end as the National Assessment of Educational Progress (NAEP). Even more consequentially, in 1966 the distinguished sociologist James S. Coleman and a team of analysts reported, on the basis of a huge study, that variations in traditional school "inputs" didn't have much of a relationship to student achievement outcomes. Although this mega-finding was intensely controversial among educators—and embarrassing to the Johnson administration, which the previous year had persuaded Congress finally to enact a major program of federal financial assistance for the education of disadvantaged children—

subsequent reanalyses of Coleman's data and related studies came to much the same conclusion.[3]

Implicit in these analyses was the message that, if people weren't satisfied with what was coming out of US schools, they had better focus laser-like on those outcomes, not just on school resources, offerings, and operations.

Dissatisfaction with school outcomes could be found even earlier, initially from individual authors—Rudolf Flesch's *Why Johnny Can't Read* dates to 1955—and issue-focused organizations, such as the Council for Basic Education, founded in 1956, which rued the dearth of traditional knowledge in the standard K–12 curriculum. That situation changed in 1957—and disquiet with the outcomes of American schooling grew louder and more widespread—when the Soviet Union's successful Sputnik launch prompted widespread concern about the school system's failure to equip its graduates with sufficient prowess in science, technology, engineering, and mathematics (STEM) and foreign-language subjects, prompting Congress to pass the National Defense Education Act (1958).

In the two decades that followed, discontent with the schools mostly took different forms as the United States struggled with desegregation and other equity issues. Even then, however, some states required students to pass exams in order to graduate—most famously the New York Regents—and others, responding to concern that some high school seniors could barely read their own diplomas, mandated "minimum competency tests" that they had to pass. This was, in fact, a form of outcomes-based accountability for education, but it focused on students rather than the schools they attended.

A big shift began in the early 1980s, this time focused on the weak academic results emerging from the schools themselves, which the College Board had foreshadowed in 1975 when it disclosed that the average score on the Scholastic Aptitude Test (SAT) had been declining for the previous eleven years.[4] But the key event came in 1983 when the National Commission on Excellence in Education issued its famous, critical report on American K–12 education, titled *A Nation at Risk*.[5] It was swiftly joined by a number of other reports, studies, and warnings saying more or less the same thing: American schoolkids weren't learning enough. Their schools needed to produce better results or the country was in serious trouble.

What would such an effort mean in practice—and how could such a demand fit into our system of education? For starters, it meant becoming clearer and more explicit about the outcomes that Americans want from their schools—in essence, a form of goal setting. That's standard practice in many realms of business and government, but, in the decentralized structure of American K–12 education, decisions about learning goals had long been left to individual teachers, schools, and districts. That familiar arrangement wasn't going to cut it for a "nation at risk," however, or for a growing number of fretful governors, particularly

in the South, who worried about their states' economic future. Loosely coordinated by the Southern Regional Education Board, and in time turned into a wider effort by the National Governors Association (NGA), they came to understand that—as Governor Lamar Alexander repeated to all who would listen while promoting the ambitious education reform package that he unveiled months before *A Nation at Risk*—"Better schools mean better jobs for Tennesseans." They also came to insist, in a phrase that the NGA employed as the title for its multiyear advocacy effort (also spearheaded by Alexander), that it was "time for results."[6]

Specifying the desired results would not, however, amount to much without some way of knowing whether those results were being produced. Nor was it evident that the results being sought from schools were genuinely "marketable" in the post-K–12 world of college and the workplace.

Specifying the desired results would not, however, amount to much without some way of knowing whether those results were being produced.

Suppose, too, that a state did, say, spell out the results it wanted and developed metrics for gauging progress toward them. Why would that lead to *better* results unless there was also some mechanism for inducing change when the results weren't satisfactory? Consequences, one might say. Thus began to emerge an "accountability tripod" consisting of targets and standards, tests and other performance metrics, and transparency and consequences.

Yet discontent with student achievement and tough-minded accountability are not the entire story. Part of the backdrop to this saga is the quest for the continuous improvement of a vital enterprise, an approach that's taken for granted in many other realms where organizations may be healthy enough to meet today's challenges but not robust enough for tomorrow's. In education, it's often termed "capacity building," referring to the readiness of schools, districts, and states to make the complex, interlinked changes that are generally required to make systems more productive. In retrospect, many commentators argue that the accountability tripod needed a fourth leg from the beginning, thereby transforming it into a sturdy table that joined the demand for better results with the expertise, resources, and will to produce them.

Analysts such as Marshall Smith, Jennifer O'Day, and the late Ronald Edmonds observed that much could be learned from highly effective schools—of which the United States has long had a goodly number—and the systems and policies that make them possible. Moreover, for results-based accountability to work as intended, it should—they insisted—be part of systematic thinking about what schools and districts require by way of learning opportunities, expertise, training, and resources in order to produce today's desired results and be well positioned for tomorrow's. By the 1980s, the "effective schools" movement, piloted by Edmonds, had begun to get traction, gradually evolving into "systemic" thinking about school reform.[7] Such systems include high expectations, clear standards,

transparent metrics, active feedback loops, and a culture of improvement. In that context, "accountability" is a helpful element of a forward-moving system, not a scary monster waiting to pounce.

Charlottesville and Beyond

By the time newly elected president George H. W. Bush—America's first self-declared "education president"—convened state governors in Charlottesville in 1989, one important step had already been taken: the NAEP had been revamped, such that achievement data would henceforth be produced for grades four, eight, and twelve, opening the possibility that such data would be supplied for individual states as well as for the nation as a whole, and allowing NAEP's new, semi-independent governing board to develop "achievement levels" that would ultimately function as benchmarks by which to report the adequacy of that achievement. This overhaul emerged from a bipartisan effort between the Reagan administration and the Democratic-majority Congress, an effort impelled in part by governors' frustration at the inadequacies of available achievement data as well as the "Lake Wobegon" discovery in the late 1980s that the normed tests in common use by state and local school systems were masking the weak performance of many children and schools.[8]

Gathered at the University of Virginia in September 1989, eight months after Bush's inauguration, the White House and forty-nine governors—led by Iowa's Terry Branstad, then chair of the National Governors Association, as well as Arkansas's Bill Clinton and South Carolina's Carroll Campbell—agreed on something unprecedented for the United States: national education goals, set for a decade hence, wildly ambitious, and expressed primarily in terms of outcomes, including America becoming "first in the world in math and science by 2000," and this one, which held big implications for school accountability:

> By the year 2000, American students will leave grades 4, 8, and 12 having demonstrated competency in challenging subject matter, including English, mathematics, science, history, and geography, and every school in America will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy.[9]

In short order, the goals were joined by a new National Education Goals Panel to monitor and report on progress toward them.

In hindsight, the goals as framed were moon shots, unaccompanied by rocket building and astronaut training. Nor did they have "marketplace validity" inasmuch as the targets, however grandly stated, were nebulous and not anchored to real-world standards such as true readiness for college without remediation or for decently

compensated roles in the workforce. Still, the "goals process" set the stage both for state-specific academic standards in key subjects and for a series of moves by the federal government as well as many individual states to employ those standards to drive change in US schools.

Timeline of Key Events

1963: Education Commissioner Keppel launches what becomes NAEP
1966: Coleman report
1970s: Many states develop new tests
1983: *A Nation at Risk*
1985: National Governors Association launches "Time for Results"
1980s: Some states (e.g., North Carolina, Texas) create accountability systems
1988: New NAEP legislation paves way for state-level results, standards
1989: Charlottesville Summit; National Education Goals set
1991: President George H. W. Bush proposes "America 2000"
1994: Congress passes Goals 2000 and Improving America's Schools Acts
2002: Congress passes No Child Left Behind Act
2005: Education Secretary Spellings begins to issue NCLB waivers
2009: Launch of Race to the Top
2015: Congress passes Every Student Succeeds Act

At the state level, say Eric Hanushek and Margaret Raymond, "the decade of the 1990s began the age of accountability" and their analysis yielded "consistent evidence that introduction of state accountability had a positive impact on student performance."[10]

At the national level, the first move toward accountability was the "America 2000" initiative developed by Lamar Alexander—by then education secretary—for President Bush.[11] It was to include—along with much else—"world class standards" in five core subjects and a voluntary program of national tests keyed to those standards. Versions of the former did come to pass, but not the latter, and in barely more than a year Bush was succeeded in the Oval Office by Bill Clinton, whose team worked with a more congenial Congress (which had been excluded from the original "national goals" initiative) to pass the Goals 2000: Educate America Act in 1994, thus creating a legislative basis for the goals process while adding even more goals to the already-ambitious list.

In the same spirit, but more consequential for education accountability, 1994 brought another piece of legislation: the Improving America's Schools Act (IASA), key provisions of which were summarized thus by *Education Week*:

> In exchange for Title I grants, states must develop school-improvement plans—with input from local district officials, teachers, parents, and others—that establish high content and performance standards in at least mathematics and reading or language arts. . . .
>
> Assessments aligned with the content standards must be administered "at some time" between grades 3 and 5, again between grades 6 and 9, and again between grades 10 and 12. They should include "multiple, up-to-date . . . measures that assess higher-order thinking skills and understanding" and "provide individual student interpretive and descriptive reports" as well as disaggregated results within states, districts, and schools by gender, race, limited-English-proficient status, migrant status, disability, and economic status. States have one year after receiving their allocations for fiscal 1995 to develop standards and assessments. If they do not, they must adopt approved standards drafted by another state.[12]

IASA thus took long strides toward requiring states to develop academic standards and assessments, as well as analyses that disaggregated the assessment results on multiple dimensions. But it also established a congressional habit of employing terms such as "proficiency," "high standards," and "higher-order skills" instead of attaching the targets to bona fide readiness for life after high school. Nor did it do much for the consequences leg of the tripod. Its working premise had more to do with ensuring that all kids in a state would be pointed toward the same standards and deploying the theories of "systemic" school reform—a blend of standards, resources, and local flexibility—than with sanctioning or forcibly intervening in low-performing schools. One might say it loaded the accountability gun but never pulled the trigger.[13] Hoover Education Success Initiative (HESI) contributing author Thomas Dee describes the situation seven years after enactment of IASA thus:

> In 2001, only 17 state accountability systems rated all their schools. Ten additional states identified only their lowest-performing schools. And, critically, the attention to "subgroup" performance (e.g., by race and ethnicity) under these state accountability systems was even more rare (i.e., in 5 states). Similarly, only a distinct minority of states had clear authority for any sort of sanctions like closing failing schools, replacing individual principals or teachers, permitting students to enroll elsewhere, or revoking accreditation. And most state accountability systems did not articulate any school supports such as technical assistance or extra funds.[14]

No Child Left Behind

Self-assured is one characterization of George W. Bush as he moved from the Texas governorship to the White House in January 2001. Arriving with him—and a key plank in his campaign platform—was a track record of marked improvement in the performance of Lone Star schools and minority students during his time in Austin. In the midst of the presidential campaign, in fact, the RAND Corporation published a report that lauded both Texas and North Carolina for the education gains those states had made—and asserted

that "accountability" had been part of the formula.[15] Although neither state had engaged in forceful interventions in its faltering schools, both had made good use of transparency as an incentive, and the Tar Heel State had sent teams to assist low-performing schools to up their game.

Bush was keen to extend the Texas successes to the entire nation. (A GOP colleague recalls him as "wanting to be governor of the United States.") He was a zealous education reformer who deplored what he termed "the soft bigotry of low expectations." He was not averse to toughening the federal requirements associated with the big Title I program, which focused on schools serving predominantly disadvantaged youngsters but which, in thirty-five years, hadn't recorded much success in ramping up their achievement. It didn't seem to matter to him that conservative Republicans had long resisted the "big government" moves that would be required to press states and districts in this (or any) direction. Instead, he and his team made common cause with Massachusetts Democrat Edward M. Kennedy, who chaired the key Senate committee, and in remarkably short order—accelerated after the 9/11 attacks—a bipartisan agreement was forged. Professor Dee described the agreement thus:

> The federal NCLB Act, which was signed in January 2002, marked a dramatic expansion in the scale and ambition of these earlier state-level school accountability systems. *In particular, NCLB brought test-based accountability to scale across the U.S. with an emphasis on both consequences and subgroup performance*. Specifically, NCLB required public schools receiving Title I funding to test students in reading and mathematics in grades 3 through 8 and once in high school. NCLB also required public reporting of school-level test results, both overall and for various subgroups (e.g., English learners, low income, special education, racial minorities). NCLB gave states flexibility in the design of their tests and their standards but also required state participation in NAEP assessments as a form of auditing.[16] [emphasis added]

In one important sense, however, NCLB had it backward: it charged the states with identifying the learning standards and testing regimes that they would use to describe student proficiency, while the federal government dictated what actions should be taken if progress was insufficient. Lacking knowledge of the demands on or capacities of America's tens of thousands of diverse schools, Washington was ill prepared to dictate how all schools should provide education. At the same time, individual states have little capacity to set standards aligned to the national and international labor markets that they are preparing their young people for. And because truly linking their standards to readiness for what followed and then attaching high stakes to them would mean, for years to come, that millions of young Americans would be denied grade-to-grade promotions and high school diplomas, while also aggravating and embarrassing innumerable educators, states had an incentive to set low bars, maybe even (in Daniel P. Moynihan's memorable phrase) to "define deviancy down."

NCLB had it backward: it charged the states with identifying the learning standards and testing regimes that they would use to describe student proficiency, while the federal government dictated what actions should be taken if progress was insufficient.

Backward though it may have been, and flawed as it turned out to be in other ways, NCLB's stepping up of the federal role in American K–12 education was unprecedented. And almost from the outset, it felt too rigid and uniform for many states, some of which—including Florida, led by the president's education-reform-zealot brother Jeb—had preexisting accountability regimens of their own. Remarkably, though, as we'll see below, NCLB did produce some gains in student learning and gap narrowing.

The Waiver Era

By Bush's second term, many states were seeking waivers from NCLB's strictures and Education Secretary Margaret Spellings started to hand them out, a process that accelerated—and grew more formal and complicated—during the Obama administration. Dee describes what happened thus:

> In November of 2005, Secretary of Education Margaret Spellings invited states to submit proposals for "growth models" that measured student achievement through gains rather than relying on proficiency thresholds. This allowed the Bush Administration to avoid a messy legislative overhaul of a key feature of its domestic agenda as the law approached its 2007 expiration date. . . .
>
> In 2011, the Obama Administration unilaterally invited states to apply for waivers from key NCLB requirements through a formal process in which they would articulate new accountability plans consistent with the Administration's vision for reauthorization. Most states applied for and all but 7 states received these federal waivers.
>
> In general, the design of NCLB waivers gave states substantially increased flexibility but did so within the structure of two broad school-accountability features. First, these federal waivers required states to define and implement "college and career ready" standards and school-level performance reporting. This new guidance continued to require state content standards in reading and mathematics as well as aligned "high quality" assessments. However, states were no longer required to achieve universal student proficiency on these test measures. Instead, NCLB waivers allowed states to articulate "ambitious but achievable" goals for school improvement. And the measurement of school performance under waivers no longer focused exclusively on test-based proficiency thresholds and the performance of multiple subgroups. Instead, waivers allowed states to measure school performance using more complex performance indices. . . .
>
> Second, NCLB waivers required states to develop and implement a system of "differentiated accountability" that targeted a distinct minority of each state's schools

> for unique identification and reform. Specifically, states were asked to identify two separate groups of schools for targeted interventions and supports. . . . The waiver process required states to implement one of several federally prescribed reforms in these schools (e.g., transformation, turnaround, restart, or closure). . . . Notably, under these waivers, schools identified for improvement were no longer required to offer public school choice or supplemental services to their students.[17]

Because the Obama waivers were closely linked to a parallel federal education initiative, the Race to the Top program funded with stimulus dollars meant to alleviate the Great Recession, states that found themselves with greater flexibility on some fronts were simultaneously being pressed to embrace other Washington-dictated practices, including the ultimately controversial "Common Core" academic standards and efforts to connect teacher evaluations to student test scores. This fed an antitesting movement on the part of both teachers and parents, which deepened alongside the country's worsening political polarization.

In Part III, we will bring the school-accountability saga up to the present by examining the Every Student Succeeds Act and its implementation. First, though, we dig deeper into generic accountability issues and a review of the research evidence.

PART II: ISSUES AND EVIDENCE

What's Accountability, Anyway?

Accountability in education, as noted above, may be thought of as a tripod, although many prefer the stabler version that adds the fourth leg, called capacity building.

The first leg sets forth the desired outcomes, typically in the form of academic standards that spell out the skills and knowledge that students should acquire at key stages of their progress through the schools. In recent years, due to federal requirements, this has most commonly entailed grade-by-grade learning expectations in core subjects, beginning with English language arts and math but often including science, social studies, and perhaps more. Other desired outcomes, such as high school graduation or college/career readiness, may be added or substituted.

The second leg is the creation and use of measures by which to gauge how well those desired outcomes are (or aren't) being achieved—by individual students and by classrooms, schools, districts, states, and the nation as a whole. This has most often taken the form of standardized testing of various kinds, although other metrics and indicators are also employed. While noting that formative assessments, teacher-conferred grades, and promotion/graduation rates are important information sources for educators and parents alike, in recent years the tripod's second leg has consisted mostly of end-of-year assessments external to the school itself, most often (under federal law) administered by the state and aligned with its academic standards.

The third leg of the tripod includes the rewards, benefits, sanctions, changes, and interventions that follow from the information on how well a student, teacher, school, or district is achieving the desired results. Consequences take many forms, sometimes spontaneous (as when a teacher or school moves to alter its practices or a parent moves her child into or out of a particular school) and sometimes imposed from outside: gold stars and blue ribbons on the one hand and restructurings, probations, and staff changes on the other.

Not every accountability system adds a fourth leg—often called "capacity building," sometimes called "opportunity to learn"—but when done this generally consists of efforts to enhance opportunities, foster equity, and strengthen performance, typically by adding expertise, resources, and better-prepared personnel.

Who's Accountable and How?

Underlying almost every discussion of accountability, but not always made explicit, is a basic question: What level of education is to be held to account for its results? Is it the individual student? The teacher? An entire school? The district? All these versions have been—and are being—tried in various ways around the country. Federal requirements primarily bear on schools themselves—particularly Title I schools—and to a lesser extent on districts, but states actually operate the accountability system. Uncle Sam has occasionally ventured into teacher accountability, and a number of states also focus on student accountability. The latter takes many forms—not just tests—and is an important topic in its own right but not one we have scope for here. Here we focus on school-level accountability, but as we do, let's stipulate that it's just one of four possibilities for results-based (or outcomes-driven—we'll use the terms interchangeably) accountability in K–12 education.

Many complexities arise in connection with school accountability, beginning with the blunt fact that those who lead and work in a school seldom like or want it.

Many complexities arise in connection with school accountability, beginning with the blunt fact that those who lead and work in a school seldom like or want it—and there's been pushback from many in the education community, beginning with denial of the problems identified in *A Nation at Risk* and continuing—even intensifying—into the present day.

Educators understandably see themselves as competent, trustworthy professionals who know what's best for children and do their best to deliver it. Nobody much likes having outsiders watch over them, particularly when the watchers are authority figures who also oversee budget, personnel assignments, and performance evaluations. Nobody with less-than-stellar results likes being embarrassed—humiliated, even, with reputations tarnished—by publicity about their results, and it's worse still when one disdains the metrics by which those results are tabulated.

The Pandora's Box of Consequences

Unflattering publicity may be just the beginning. A fundamental issue for accountability is what is done with—and follows from—that information about school performance. For starters, it informs the school's many constituencies, which include parents of children who attend it and others considering which school to send their kids to or which neighborhood to move into; educators who work in that school and other schools, some of whom may be considering where to work in the future; and the broader population of voters, taxpayers, civic leaders, journalists, shopkeepers, politicians, advocates, and government leaders.

A fundamental issue for accountability is what is done with—and follows from—that information about school performance.

This diffusion of performance information may itself trigger change. It may prompt those who lead and work in the school to take stock, see what needs fixing, and alter their staffing, curricula, instructional practices, and pupil expectations. It may lead the school's constituents—parents, neighbors, alumni/ae, others—to press for reforms. It may lead the district or network that the school is part of to require such changes. It may lead families to exit from—or flock into—the school, exercising (when available to them) any of myriad versions of choice, with further consequences for the school and those who work in it. And in the framework of a state accountability system, the information may lead to interventions dictated from the capital, even (as under NCLB) from Washington. These might include the dispatch of a flying squad of experts and professional developers to assist the flailing school to solve its problems, or might mean more forceful intrusions: engage a "turnaround" organization, change the school's leaders, replace its instructional staff, convert it to a charter school, entrust it to a special supervisory unit, even shutter it completely. In a few states, unsatisfactory academic results—and/or financial chicanery—have led to takeovers or receiverships for individual schools and entire districts.[18]

Such interventions have a mixed track record. After a close review of the ways they have been applied—usually with the least permissible disruption to the schools or districts in need of overhauls—the Bellwether Education Partners team concluded that "turnaround efforts have produced measurable gains for students only in places that engage in serious, dramatic reform efforts that meaningfully alter all or some parts of the school's curriculum, instruction, and staffing."[19] Elsewhere, there has been much wheel spinning, which more often than not means a low-performing school or district will stay that way.[20]

Like other organizations, schools *can* improve, of course, but that's not quite the same thing as "being improved" via outside intervention. More often—and, sadly, it's not very often—this requires a protracted cycle of "continuous improvement" based on evidence and the application of sound management principles.[21]

We tend to think of accountability systems as punitive, but they can be much more than that.

Bear in mind, too, that the consequences leg of the accountability tripod does not always bring embarrassment or unwanted interventions to a school. Data on a school's performance may, in fact, be so favorable as to inspire commendations for staff, a gold star or blue ribbon for the school—and an influx of additional pupils. We tend to think of accountability systems as punitive, but they can be much more than that. They can boost property values in the neighborhood, lead to VIP visits and grand celebrations, create promotions and career opportunities for teachers and principals, improve kids' college-admission prospects, splash all over the local media, draw accolades at national conferences, establish best practices for wider emulation, and perhaps figure in articles and TV shows about great schools.

Why All the Testing?

Transparency about school outcomes is no better than the information it's based on, which leads to the next issue: What *is* that information? What exactly is being measured—and why does so much of it usually boil down to student test scores? Every educator knows that tests, even the most sophisticated, are at best a partial indicator of what kids are actually learning and of what's worth learning, not to mention the many other functions that schools and educators are expected to perform for their young charges that don't show up on tests.

America is in the grip of a testing backlash, attributable in part to sometimes clumsy and overwrought accountability systems, inflamed by educators who don't like what the emphasis on tests is doing to their curricula and pedagogy and would just as soon the assessments and accountability structures disappear. They'd much rather have parents judge their children's school performance by traditional teacher-conferred grades and comments. And they bridle at having their classroom prowess evaluated on the basis of what their pupils did or didn't learn—a grievance that gains legitimacy when we consider students' differing circumstances and mobility rates as well as the relatively small fraction of most schools' instructional staffs to which year-to-year test-score changes are even applicable.[22]

Every educator knows that tests, even the most sophisticated, are at best a partial indicator of what kids are actually learning and of what's worth learning.

We must also be mindful of "Campbell's Law," named for the eminent social scientist Donald T. Campbell, who wrote in 1979 that "the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor."[23] In its mildest form, Campbell's Law says that when we judge—and reward and punish—schools on the basis of test scores, educators will inevitably place heavy emphasis on what's tested—and on drilling students

in advance of the tests—and pay less attention to everything else, however important those other things may be. In more extreme form, people working in schools will finagle those scores to benefit themselves, from arranging for likely low scorers to be absent on test day to actually altering pupils' answers on the test forms.

Yet standardized tests are heavily used in the accountability realm because they have clear advantages. They're relatively inexpensive to administer and their administration can be monitored to minimize cheating and ensure that students do their own work. Most can be scored by machines, which is fast and cheap and gives at least the appearance of objectivity and uniformity, which many see as equivalent to integrity, comparability, and fairness. (Open-ended, free-response, and essay elements keep being added to the tests, and artificial intelligence is showing early promise as an affordable and consistent tool for evaluating those portions.)

Other indicators that share the virtue of apparent objectivity, such as graduation and attendance rates, don't have much to do with actual learning, while other indicators that focus on learning—e.g., grades, portfolios, projects, teacher comments—are inherently subjective, noncomparable from one classroom or school or district to the next, and more susceptible to being manipulated to produce a positive result. Because of the individualized attention they require, they're also generally costlier.

It need scarcely be added that academic learning isn't the only valuable school outcome. We also seek the development of character, self-discipline, and tolerance; of social, emotional, and physical well-being; of a host of "twenty-first-century" or "soft" skills such as creativity and the ability to work well with others; and of long-term outcomes such as career success and good citizenship. Many efforts are now underway to develop metrics and instruments for such things—and some are far enough along to appear on school websites and dashboards where they can benefit parents and educators, not to mention real estate agents—but we have a long way to go before many metrics of this sort are stable, valid, comparable, and reliable enough to play a legitimate role in consequential school-accountability calculations.

The Evidence to Date

Many scholars have sought to answer fundamental questions about the impact of accountability regimes.[24] How effective are they at measuring and reporting on school performance? At boosting school performance? What about unintended effects? Do their consequences undermine students' and teachers' own motivation? The literature on student-centered incentives finds scant evidence that this is a problem. Indeed, there's some evidence that carefully formulated student incentives yield stronger performance.[25] And there are mixed messages as to whether linking high school graduation to passing exit exams and end-of-course exams (EOCs) depresses graduation rates.

There is persuasive evidence that state-developed accountability systems of the kind that emerged in the 1990s boosted achievement.

As noted earlier, there is persuasive evidence that state-developed accountability systems of the kind that emerged in the 1990s boosted achievement, although without visibly narrowing the black-white achievement gap.[26]

As for the federally driven school-accountability regimes of recent decades, numerous efforts to appraise their effectiveness have yielded little consensus.

Critics contend that NCLB and kindred test-based accountability regimens led to overemphasis on "teaching to the test" and narrowed the curriculum to tested subjects, even to tested portions of a subject. There are signs, for example, that states showing solid growth on high-stakes tests failed to show similar gains on low-stakes tests in the same subjects.[27] A further concern is that focusing, as NCLB did, on getting students over a single "proficiency" bar leads to neglect of those far above or below that bar. The available evidence suggests that low achievers made greater academic gains than high achievers during the NCLB era, particularly during its first decade, though we find no systematic evidence that high achievers were neglected by their schools.[28] On the other hand—cue Campbell's Law—in a few places the proficiency push did contribute to undesirable practices, such as cheating on exams, manipulating scores, and influencing which students would participate in key assessments.[29]

At the same time, we see solid evidence of modest achievement gains—most prominently in math and in the earlier grades—by some student groups, particularly poor and minority youngsters, during the early years of NCLB. Twelfth-grade NAEP scores, however, have remained stubbornly flat for decades.

We see solid evidence of modest achievement gains—most prominently in math and in the earlier grades—by some student groups, particularly poor and minority youngsters, during the early years of NCLB.

The National Research Council stirred the pot with a forceful "panel" statement in 2011 that the test-based programs its team was able to evaluate showed scant effect on achievement and may have depressed high school graduation rates. Other careful scholars, however, have challenged that analysis.[30]

Further studies of student performance under accountability regimes with palpable consequences have generally found some—but not huge—gains in the tested subjects.[31] Dee makes this observation:

> One methodological concern with studies based on accountability pressure is that the resulting impact estimates may be biased downward because they rely on comparisons

> among schools, all of which operate under the same accountability regime. Motivated in part by this concern, Dee and Jacob instead examined the impact of NCLB on low-stakes test scores by comparing changes across states that already had NCLB-like accountability to the changes in states where NCLB created entirely new experiences with test-based accountability. They found that NCLB increased grade-4 math scores on the NAEP by 0.23 standard deviations with smaller but positive effects on grade-8 math and grade 4 reading performance. . . .
>
> The National Academy report on test-based accountability stated . . . that gains of this magnitude are "small compared to the improvements the nation hopes to achieve." However, Dee, Jacob, and Schwartz note that these effects are not necessarily small from a cost-benefit perspective. . . . *In sum, studies of pre-NCLB and NCLB-based accountability systems indicate that they generated meaningful, though not transformational, improvements in school performance.*[32] [emphasis added]

A 2020 analysis by Bellwether Education Partners reached a similar conclusion:

> Standards-based accountability policies have contributed to measurable improvements in student performance. . . . NCLB-era accountability policies produced meaningful improvements in student achievement, particularly for traditionally underserved student groups. . . . While these gains are substantial, [however,] the impact of standards-based accountability has not fully lived up to its initial promise. The improvement . . . in math has not been matched with similar gains in reading. And student achievement has stalled over the past decade, with growing gaps between high- and low-performing students.[33]

The US evidence, it should be noted, is consistent with international evidence on school accountability, which shows that countries with testing programs that allow for external comparisons have students who do better on international achievement tests. And when such measures also incorporate consequences, whether for individual students (as in end-of-course and matriculation exams) or for entire schools, results improve further.[34]

PART III: WILL EVERY STUDENT SUCCEED?

Long before it was finally replaced in 2015, NCLB was roundly criticized, sidestepped, waived, and sometimes mocked. In addition to the "backward" structure noted earlier, it labeled too many schools as unsatisfactory. That's because its single-minded focus on achievement rather than student gains tended to make low-income schools look bad, and it placed otherwise-satisfactory schools on the hit list if they fared poorly with one or two pupil subgroups. Its 2014 target for universal proficiency was naively unrealistic, even as its remedies for troubled schools were both overly rigid and vulnerable to manipulation.[35] And its authors' decision that every state should set its own proficiency standard led to wildly discrepant expectations for students and schools.[36]

On the plus side, NCLB's reporting and transparency requirements shone a much-needed spotlight on subgroup performance, which could no longer be masked under school- and district-wide averages.

On the plus side, NCLB's reporting and transparency requirements shone a much-needed spotlight on subgroup performance, which could no longer be masked under school- and district-wide averages.

NCLB was also getting long in the legislative tooth. Historically, Lyndon Johnson's Elementary and Secondary Education Act (ESEA), first enacted in 1965, had been updated by Congress every half dozen or so years, but at this point in US history the legislative branch was often overcome by dysfunction. With schools, districts, and states all wanting out from under a too-heavy accountability hand from Washington, having already harvested the relatively modest achievement gains that NCLB would produce, and with waivers becoming more the norm than the exception, the time had plainly come for an overhaul. Yet others—especially in minority and civil rights communities—wanted no easing of the federal push for transparency and attention to subgroup achievement gaps.

The policy glacier finally cracked in 2014 and 2015, thanks primarily to a rare display of bipartisan compromise on the part of the chairman and ranking minority member of the Senate Committee on Health, Education, Labor and Pensions (HELP), namely Republican Lamar Alexander and Democrat Patty Murray, who painstakingly managed a major makeover of key portions of NCLB that didn't fully satisfy any constituency but—classic compromise—could be tolerated by all. HESI contributing authors Paul Manna and Arnold Shober describe the compromise thus:

> In trying to craft a federal presence that was neither too hot nor too cold, lawmakers writing ESSA embraced a theory of action built around two main ideas.
>
> First, they assumed NCLB had laudable goals but also had employed a heavy-handed approach that should be pulled back so states could exercise broader authority when implementing ESSA. The federal government's admirable ambitions led it to overreach with NCLB, imposing new requirements and procedures on states and local school districts. . . . A better approach, which ESSA attempted to embody, would scale back federal prescriptions for school accountability but not abandon them altogether.
>
> Second, ESSA's authors assumed state governments were capable innovators and with the regulatory relief ESSA provided would faithfully act to ensure all students succeeded. A major line of state advocacy during the NCLB era, which encouraged the Obama administration to experiment aggressively with NCLB waivers, was that states needed greater flexibility from federal constraints to promote educational excellence and equity for all, goals that several states had themselves set in the 1980s. The authors of ESSA

agreed. As a result, "accountability" remained a controlling idea in the new law and it kept some of NCLB's elements while on others it permitted states to shift in new directions.[37]

ESSA's Big Gamble

It's useful, in explaining the changes wrought by ESSA, to view accountability in two frames, the first focused on standards and measures of school (and student) performance, the second on consequences. As Manna and Shober write:

> The biggest shifts in accountability came in the second category of activities, in particular, the consequences that flow from test results and other performance measures. The most striking difference was that ESSA significantly lessened the cascade of NCLB remedies that states were supposed to enforce on schools and districts that failed to make adequate yearly progress (AYP). The law no longer requires that 100 percent of students meet proficiency goals by a certain date. Also gone are NCLB's requirements that schools missing AYP must offer their students public school choice and supplemental educational services (e.g., tutoring), and the additional corrective actions and potentially major restructuring to improve school performance that NCLB had demanded for schools that struggled in several consecutive years. ESSA's new approach focused on a smaller number of schools—dubbed schools requiring Comprehensive Support and Improvement (CSI), Targeted Support and Improvement (TSI), and Additional Targeted Support and Improvement (ATSI)—and gave states greater flexibility in designing various sanctions or supports to address their needs. Although these new categories came with demands of their own, the states had their primary wish granted, namely, much room to maneuver on how to identify and react to schools that struggle to meet performance goals.[38]

ESSA proffered a virtual invitation to states to think innovatively and creatively about school accountability and—to the consternation of some state officials seeking clear federal guidance—that opportunity widened when Congress allowed the Trump administration to withdraw detailed guidance that the Obama team had issued regarding the required elements of state plans.[39] Many, though by no means all, states made use of that freedom to craft accountability plans that suited their own priorities and politics. Indeed, the hallmark of ESSA so far is how differently it's being interpreted and applied, depending on where one lives. It is clear, Manna and Shober write, "that the states have sought to squeeze as much variation out of ESSA as possible." But is that good for kids, for academic achievement, and for the country? Depends on your perspective:

> From one perspective, states adopting relatively diverse approaches, some even backing away from what some might consider a higher bar for accountability, is no problem at all and might even be considered a success given that one interpretation of ESSA's mandate was to simply unleash the states to try different strategies. That perspective sees ESSA, and in particular its Title I provisions, mainly as a funding stream designed to help

top-off state and local education budgets. A different perspective sees ESSA as a more robust federal education program with particular goals that embraces a particular vision for how states should go about accomplishing them. Those endorsing that point of view would be disappointed at the significant variation across state approaches to accountability, and in some cases state resistance to federal preferences.

Manna and Shober reviewed several independent analyses of the strength of each state's accountability plan and sought correlations between those plans' strength and other variables, such as states' prior levels of achievement, demographics, and education-governance structures. A handful of states ranked high and several ranked very low, with most—naturally—strewn across the middle of the distribution. But correlations were few and mostly weak, which is to say it's nearly impossible, on the basis of observable state attributes, to predict how solid an ESSA accountability plan a state would develop.

ESSA proffered a virtual invitation to states to think innovatively and creatively about school accountability.

Perhaps most promising is that states whose fourth graders had made respectable achievement gains (gauged by NAEP) during the early NCLB years were somewhat more likely to develop strong plans under ESSA, suggesting that they desired to keep a good thing going. (The eighth-grade link was weaker.) There was also some evidence that states with appointed rather than elected state boards and education commissioners were likely to produce stronger plans, suggesting—no real surprise here—that in this realm being answerable to a reformist governor is better for accountability than answering directly to a more passive or complacent electorate.

Mostly, though, the vigor and rigor of a state's accountability plan turn out to be idiosyncratic and situation specific, the result of different priorities and the political dynamics at work in individual states. "The politics of education," write Manna and Shober, "is not nationalized to the extent that other issues, such as health care and gun control, are. Episodic and contextual politics seem to dominate at the state level." And because the US Department of Education was broadly tolerant of whatever approach a state submitted, sometimes raising questions and pushing back on specifics but ultimately signing off without much of a fight, states were—and remain—in the driver's seat. So long as they don't cross the road's centerline too often, they can drive as fast or slow as they like, and with few emission controls or MPG standards.

The Manna-Shober conclusion is fairly glum: "It appears that the political emphasis on academic performance has receded" and "it may be that accountability is now more of a millstone to politicians than a motivator of schools. . . . After thirty years of beating the accountability drum, there may simply be too little political gain for contemporary politicians to continue playing off that same sheet of music. Instead, state governments have

to fight rear-guard actions justifying the length of accountability measures and preventing parents from nixing their children's test participation."

It's also possible, however, that while outcomes-driven accountability has become a problem for politicians in many places, it still functions as an important diagnostic and management tool for skillful, achievement-minded education leaders and as an effective if unpopular motivator for those who work in the schools. Where those features are prized, ESSA can help, not least by supplying state and local leaders with political cover, as in "Like it or not, we have to do these things or risk our Title I funding."

While outcomes-driven accountability has become a problem for politicians in many places, it still functions as an important diagnostic and management tool for skillful, achievement-minded education leaders and as an effective if unpopular motivator for those who work in the schools.

Making the Best of ESSA

The testing hiatus and "data hole" resulting from school closures and federal waivers in spring 2020 will make calculation of year-to-year achievement growth difficult or impossible in the short run.[40] Schools may have to substitute two-year growth (i.e., spring 2019 to spring 2021) as best it can be gauged. The unevenness of such data and fragility of calculations may lead some states to suspend their summative school ratings for a year—a real disservice to parents. We urge careful improvisation in the near term and a resumption of familiar calculations, comparisons, and ratings as soon as possible.

While ESSA remains the law of the land, we've learned some things about the requisites for a thorough and responsible accountability plan that fulfills ESSA's requirements while operating within its limits. The following eight elements characterize such a plan, stipulating—of course—that unless these are well implemented and consequences follow for problem schools (whether via the ESSA plan or otherwise), we cannot count on shortcomings being rectified and achievement gains being made.

- *Transparent information.* School accountability measures should be made public in a way in which the most important information is easy for readers to grasp. Clearly explained summative ratings for individual schools (e.g., A–F grades or one to five stars), although controversial in some quarters, help ensure that all stakeholders, especially parents, understand both how a school is performing and when improvement is needed.[41] Easy accessibility via state, district, and school websites should be obvious, but parents would benefit if annual report cards on their children's schools could also come home via snail mail or backpacks.

- *Comparisons.* School scores should be placed in context, ideally with other states, certainly in comparison with other schools and districts within the state. Beyond

ESSA's requirements, states seeking to compare their results internationally can participate directly in the "PISA" assessments conducted by the Organisation for Economic Co-operation and Development (OECD), as Massachusetts and North Carolina did in 2015. Also highly desirable for accountability-minded states would be comparable external data on achievement at the end of high school, which argues for the federal government to ensure that twelfth-grade results from the National Assessment get regularly reported for individual states and participating districts, something neither required nor forbidden by current law.

- *Measures amenable to gains.* Measures included in accountability systems should have intrinsic meaning for those within a school (as well as parents and other constituents). They should be primarily within the school's control, that is, not heavily affected by factors such as socioeconomics.[42] Because the purpose is to improve student outcomes, it should be evident to those in the school that stronger performance on a data metric signifies or is associated with better student outcomes. Achievement and growth measures are most obvious, but so are elements such as attendance.

- *Indicators that are hard to manipulate.* Aside from actual fudging of test results, today's biggest challenge is graduation rates, which ESSA requires to be included in high school accountability metrics. Graduation standards vary widely by state and typically include "alternate routes" to diplomas that often lack external standards.[43] It's important that states indicate clearly what their standards are for every route, and what percentage of graduating students came through which route.

- *Transparent treatment of subgroups.* A high-quality accountability system clearly displays the performance of all pupil subgroups, especially those that have historically been underserved. States should avoid the use of "supergroups" that aggregate subgroups into a single measure. States should also go beyond ESSA's subgroup requirements and include performance and gains made by other important groups, such as gifted students (as Ohio has been doing and Maryland will do) and bottom-quartile students (as Florida has been doing).[44]

- *Report both proficiency and growth.* Pupil performance should be reported in terms of both absolute achievement and learning gains. Including growth encourages teachers to pay attention to students at every level of achievement, not just those nearing the proficiency bar. For teachers and parents alike, the growth of individual pupils must be tracked and (confidentially) reported.

- *Carefully chosen growth measures, whether criterion- or norm-referenced.* There's no "right way" to calculate achievement growth (also known as value-added), as every known method has both pluses and minuses.[45]

- *Attendance.* States should include student attendance rates in their accountability measures. Attendance and chronic absenteeism are the most common nonacademic indicators currently used by states, and the research basis for tracking them is compelling—particularly monitoring chronic pupil absenteeism. Markedly less instructional time clearly leads to lower achievement, and student absences are particularly detrimental to low-income and English-language-learner students.[46]

Criterion-based measures compare students' progress toward predetermined levels of performance. These systems might, for example, report the percentage of students who moved from basic to proficient or proficient to advanced. In contrast, norm-based systems compare students' progress to that of other students with similar characteristics. This highlights relative growth but is silent on how students compare to standards or expectations.

Norm-referenced measures have some advantages, as a rising average on the norm also signifies that something significant is happening in the school, the district, or the state. Yet if all schools produce little growth, the least dreadful among them would look strong, thus giving a falsely rosy—Lake Wobegon—picture to parents and school personnel alike.

A criterion-based growth measure calculates the percentage of students who move from one level to the next, gains that can also be reported by subgroup. So long as multiple benchmarks are employed, such a measure incentivizes teachers to focus on many students, and therefore discourages the kind of gaming—focusing on not-quite-proficient kids—that appeared when only one cut-point was prioritized under NCLB. The risk, still, is that states may set too-low cut-points, that "proficient" may be misleading, and that nothing is closely correlated with external indices such as college or career readiness.

Because accountability under ESSA (as under NCLB) is closely tied to achievement aligned with a state's academic standards, we tend to favor measures of growth (as well as achievement) that are anchored to those standards. Hence, criterion-based assessments and growth calculations are generally to be preferred. Whichever method is used, however, a school's report card should make clear to all readers what exactly the growth measure indicates and what it doesn't.

State leaders understand that accountability metrics differ considerably between elementary and middle schools on the one hand and high schools on the other. In the former case, essentially every young person is enrolled in school; academic standards are typically set by grade level and assessments are annual. High school is quite different, as students take different courses, academic standards are often aligned with courses rather than grade levels, end-of-course exams are widely used, graduation rates (though vulnerable to inflation and manipulation) are relevant, true readiness for college/career is an appropriate criterion,

and long-term tracking data—wherever possible—can reasonably be linked back to school effectiveness.[47]

How, Where, and What to Report

School report cards and dashboards serve three purposes. First, akin to sophisticated mirrors and microscopes, they equip the school's own team to scrutinize its overall performance, subgroup performance, and sundry elements that may feed into strategies for improvement. Second, they enable supervising authorities—districts, charter networks, states, and more—to monitor, evaluate, and diagnose a school's performance for purposes of recognition, improvement, and possible intervention. Third, they supply parents and other school constituents with information that may help them decide whether this is the right school for their child or family home, while also informing interested taxpayers, local merchants, journalists, and law enforcement agencies.

> For school-level reporting and accountability, the achievement growth of subgroups must be carefully analyzed and reported. Because states have discretion under ESSA to determine the relative weights their accountability systems assign to proficiency and growth, it's important to consider what mix is best. Some analysts argue that "because [achievement] measures are strongly correlated with student demographics and prior achievement, . . . they should count for at most a quarter of schools' ratings."[48] That's a well-meaning position, as pure achievement tends to privilege schools with middle-class pupils versus schools with many disadvantaged youngsters, while increased weight for growth will reward all schools whose pupils make gains. Yet parents and students also deserve clear signals as to the absolute performance of students in a school. The "real world" rarely rewards relative growth. Hence we recommend equal weighting for proficiency and growth.

Parents increasingly expect schools, in addition to imparting academic knowledge and skills, to help their children build character and life and career skills that will lead to independence, success, and happiness.[49] Parents welcome information on broader school missions such as social-emotional learning, career readiness, and long-term attainment.[50] Parents and educators alike care about the school's "climate," including discipline.

The quality of data on such metrics varies.[51] Incidents of serious indiscipline can readily be tabulated, for example, but it's far harder to gauge a school's "climate" other than through surveys with questionable reliability. But it's sometimes possible to follow students' progress after high school. Schools might then be able to report on the percentage of graduates who attend and persist in college and/or are employed in solid jobs.[52] In such circumstances, it makes sense for states to include such measures within their accountability systems. But

this approach obviously hinges on access to reliable longitudinal data for large numbers of young people after they exit the K–12 system.

ESSA contains elaborate requirements for information that must be reported annually on state- and district-level report cards, and the district reports must include information on individual schools.[53] In practice, this means that ESSA requires a report card to be made public for each public school in the United States, and these typically appear on websites operated by state education departments. Sometimes they're also accessible through district and school websites, as well as those of charter-school networks.[54] Sadly, however, they're often hard for parents to find and understand.

Some states augment their school report cards with more elaborate dashboards containing additional information. When created and displayed with readers in mind, both have merit. The report card should be concise and easy to understand, consisting primarily or exclusively of data that the state actually uses for evaluating, rating, and—if necessary—taking action on to improve a school. A parent looking at the report card should be able swiftly to grasp how well the school is or isn't doing—an argument for simple, summative ratings such as letter grades or stars.

A school's dashboard, when provided, is more sophisticated and contains more information, not all of which is necessarily valid and reliable for accountability purposes but which enables interested parties within or beyond the school to delve into what it tries to do and, on multiple dimensions, how well it is doing those things. This is where, for example, one could find the school's curriculum, its pedagogical philosophy, staffing arrangements, budget, and indicators such as climate surveys, indicators of social and emotional well-being, and citizenship clues, in addition to more detailed data.[55]

A good example of school report cards can be found in Ohio, while dashboards may go further, as is shown vividly in California.[56] Today, however, Ohio has only report cards, no dashboard, while California has the reverse. An optimal state system would provide both, focusing on simplicity and transparency in report cards and completeness in dashboards.

Consequences?

As discussed above—and emphasized by Manna and Shober—ESSA is far more laid-back than NCLB when it comes to consequences for low-performing schools. It does require states, working through districts, to ensure that plans get developed to address shortcomings in the lowest-performing 5 percent of their schools as well as others where subgroups of students are faltering. But it doesn't prescribe specific remedies or interventions, and states have applied these nebulous provisions in dramatically different ways. Although there's *some* evidence of a positive correlation between high-quality school-improvement plans

and student learning, the plans themselves depend on accurate needs-assessment tools that furnish schools and districts with accurate interpretations of school-level data. (See discussion in Part IV of how school "inspectors" may help with this.) And even the best plans are only that. If not joined to actions, including potentially unpopular actions, their impact may be nil. This is why if, over time, ESSA yields few achievement gains or gap narrowing in the states that prefer to shun such actions, it will likely be due to its lack of specificity with respect to consequences, which in turn may be due to its reluctance to upset educators and local leaders who don't want them.

ESSA is far more laid-back than NCLB when it comes to consequences for low-performing schools.

PART IV: LOOKING BEYOND ESSA: ACCOUNTABILITY 3.0

We cannot know what lies ahead. As of mid-2020, in the thick of COVID-19 shutdowns, shut-ins, and shaky reopenings, every state had obtained a waiver from ESSA accountability testing for the 2019–20 school year and most canceled their own end-of-course exams and eased their usual graduation requirements. The College Board, for the first time in more than half a century, moved Advanced Placement exams online, and it and ACT deferred college-admissions testing to autumn. More colleges—most prominently, the influential University of California system—are moving away from admissions testing altogether. Yet no one is certain how much the 2020–21 year will be further disrupted—and formal education changed in major ways—by a continuation or recurrence of the plague that surrounds us. At this writing, it's not even clear whether the National Assessment will be able to honor its statutory obligation to assess fourth- and eighth-grade math and reading during the 2020–21 school year.

On top of this, the pushback against testing in particular and accountability in general seems to be intensifying. In a thoughtful recent essay, Lynn Olson and Craig Jerald wrote the following:

> Pressure to reduce testing has come from many, often confounding sources: teachers' unions and their progressive allies opposed to test-based consequences for schools and teachers; conservatives opposed to what they consider an inappropriate federal role in testing; suburban parents who have rallied against tests they believe overly stress their children and narrow instruction; and educators who support testing but don't believe current regimes are sufficiently helpful given how much teaching time they consume. . . . The pushback against testing in recent years has led to a substantial retreat on testing among state policymakers. A new national analysis by FutureEd has found that between 2014 and 2019, lawmakers in 36 states passed legislation to respond to the testing backlash, including reducing testing in a variety of ways, a direction also taken by many state boards of education and state education agencies.[57]

We can't be sure how much of this is objection to testing per se and how much is shooting at the test messenger because many educators—and those they've persuaded—dislike the accountability systems for which testing now furnishes key metrics.

The "testing burden" on schools and schoolchildren is commonly exaggerated by foes of testing. A Brookings Institution analysis estimated that the average annual cost per pupil of mandated federal and state testing was about thirty-four dollars in 2012, when NCLB was still the law of the land. At the time, per-pupil spending in US public schools averaged about ten thousand dollars, meaning that government-required assessments consumed about 0.0034 percent of school budgets.[58] As for the time consumed by testing, much has been said and written about how excessive this is, but it varies widely by district and school, usually because of tests not being required by federal or state law.[59] A Hunt Institute study reported in 2014 that "across 12 urban districts, the average amount of time students spend on state and district tests equals 1.7 percent of the school year in third and seventh grades."[60] In many places, there's no government testing at all in some grades. Nowhere is every core subject tested every year.

The "testing burden" on schools and schoolchildren is commonly exaggerated by foes of testing.

Regardless of the facts, however, the future of high-stakes testing and test-based accountability is in some jeopardy, which means that, if results-driven accountability is to remain a force in toning up American K–12 education, there will be a huge appetite for different kinds of tests, for alternatives to testing, and for a broader conception of school information and transparency. There may be an equal appetite for the consequences leg of the accountability tripod to be replaced by a different design, as well as heightened interest in the fourth leg—capacity, opportunity, expertise, resources—even as that avenue grows more challenging under straitened budgets.

Our view, buttressed by long experience and solid research, is that tomorrow's accountability systems for K–12 schools should be geared to students' true readiness for what follows and to schools' effectiveness in readying them; that objective evidence should be employed to gauge both individual readiness and school effectiveness; and that there should indeed be consequences when this does not happen.

Tomorrow's accountability systems for K–12 schools should be geared to students' true readiness for what follows and to schools' effectiveness in readying them.

To that end, states should deploy sound measures of readiness in core subjects, calibrated to levels of mastery required for success in entry-level college courses without remediation and aligned with an array of industry certifications for employment. Mapping backward from those desired outcomes, a state's academic standards during the primary and secondary

grades should cumulate to college/career readiness while also incorporating subjects necessary for competent citizenship, such as history and civics.

In retrospect, we should have done it that way since the dawn of standards-based reform and results-based accountability. Had we done so over three-plus decades, we would almost surely find many more high school graduates truly "ready" than we're seeing today. But better late than never.

States should deploy sound measures of readiness in core subjects, calibrated to levels of mastery required for success in entry-level college courses without remediation and aligned with an array of industry certifications for employment.

ESSA is not immortal, and the following recommendations for assessment and school accountability assume that it can and should be adapted to changing needs and circumstances. We remain focused on school-level accountability and on academic achievement of the kinds necessary to prepare young Americans for what follows their K–12 schooling. We cannot here address every accountability-related issue that may concern state and local officials, such as whether and how to include social and emotional learning, school safety, and online learning.

Assessment

ESSA's insistence on testing annually from grades three through eight and once during high school has both over- and undertested while yielding too little information. We know next to nothing about learning or school effectiveness during the crucial early grades and almost nothing (save from NAEP) about pupil achievement in subjects other than reading and math, fundamental as those are. At the high school level, states have a variety of graduation requirements (still consisting mainly of course credits), and some include end-of-course testing in a few subjects. Yet aside from those, plus skimpy twelfth-grade NAEP data and less-than-universal results from college-admissions tests and Advanced Placement exams, we have precious little information about actual learning. Meanwhile, ESSA's reliance on graduation rates as a primary indicator of high school performance creates a potentially perverse incentive—Campbell's Law run amok—to push ill-prepared youngsters through to diplomas untethered to indicators of true readiness for what follows.[61]

With all this in mind, when the time comes for Accountability 3.0, and with profound thanks to HESI contributing authors David Steiner and Alanna Bjorklund-Young, who supplied key elements of what follows, we recommend the following, based on solid evidence that "a well-designed system of external exit examinations should be curriculum-based, define achievement relative to an external standard, measure across the full range of student performance, signal multiple levels of accomplishment, and cover the vast majority of students in a given school system."[62]

Kindergarten readiness Teachers in kindergarten and the early elementary grades need baseline information about the preparedness of their young charges to succeed in school and beyond. Such information also helps education systems gauge how much progress children make during early elementary school while generating data by which to rate prekindergarten providers.

Elementary schools should only be accountable for ensuring that such assessments are properly administered, not for their results, as schools cannot properly be held responsible for what children may or may not have learned before they arrive in kindergarten. But without such baseline data it's impossible to gauge the progress that those children make during the early grades.

Teachers in kindergarten and the early elementary grades need baseline information about the preparedness of their young charges to succeed in school and beyond.

Kindergarten-readiness assessments are not conventional paper-and-pencil tests but are more like checklists that teachers use at the beginning of the year to rate children's status on several key dimensions. Maryland has developed an exemplary version of such an assessment, which it describes this way:

> The assessment is a developmentally appropriate observational and assessment tool consisting of 50 items. The KRA measures the knowledge, skills, and behaviors that children should be able to demonstrate at the start of kindergarten. All items can be administered by the teacher using a test administration manual, but 17 of the items are also available through an App which allows students to select responses using a tablet or computer device.[63]

Good assessments of this kind probe several domains of school readiness, including but not limited to a child's preparedness to undertake reading and arithmetic. Maryland's plan, for example, also appraises the child's "social foundations," physical well-being, and motor development, all of which are important for success in school and far beyond, thereby better equipping kindergarten teachers to meet the needs of individual youngsters.

ELA and math prowess Continue to test students' ELA and math prowess—and progress—on a regular basis, preferably beginning in second grade.[64] States wishing to reduce the testing burden may prefer to assess in alternative grades during elementary and middle school, but annual testing yields more-precise growth calculations. The combination of kindergarten-readiness assessment and second (or third)-grade math and ELA tests will deliver far more information than we have today about what's happening in the earliest grades, when acquisition of basic reading competence in particular underpins everything that comes after.

Assessments across the grades should be linked to curricula aligned with quality standards, which should cumulate by the end of high school to externally benchmarked readiness for college and/or career. However, a student's progress from one stage to the next should be based on evidence of mastery of those standards, not the passage of time or a teacher-conferred grade that lacks external confirmation. This argues for loosening the shackles that have traditionally linked students' grade levels to their chronological ages and instead allowing them to proceed from one unit of a subject to the next whenever they can demonstrate mastery. This, in turn, argues for visualizing the cumulative standards (and accompanying assessments) as units rather than one-year-at-a-time expectations.

End-of-course exams States should consider adding capstone ("end-of-course," or EOC) exams in other core subjects during middle and high school, aligning these with high-quality curricula in those subjects, including career and technical subjects as well as the traditional academic core. Yes, this adds to the total testing burden, but that may be a price worth paying, as such exams signal to teachers what must be learned during the year and help them align curricula to facilitate instruction in those things. When passing such exams "counts" for promotion and graduation, students tend to study for them in a purposeful way, much as now happens in the Advanced Placement (AP) and International Baccalaureate (IB) programs. Yet—again, as we see in those programs—many teachers will welcome the challenge of preparing their pupils to succeed on those exams. As is often noted in the AP realm, this arrangement places teachers and their students on the same team, even as it serves as an external check on teacher-conferred course grades. Moreover, end-of-course exam results furnish valuable information to parents, college admissions offices, future employers, and school officials seeking to determine which teachers are effective and which students are truly on track.

A number of states already deploy EOCs for key high school courses such as algebra and civics, although—regrettably—that number has been shrinking in the face of antitesting pressures. Many states and districts also make AP and/or IB available to interested students.

A broader array of EOCs that start in middle school will expand the content areas that are monitored and that count, will help ensure that more students are adequately prepared in core skills and contents before reaching high school, and will provide a fuller picture of a school's impact on student learning than reliance on reading and math assessments alone. This will encourage entire schools to work together to develop students' academic skills and knowledge and furnish them with a balanced education.

Enabling students to proceed through their studies at their own pace, based on mastery, means that "end of course" will not necessarily mean "end of year" for a student. We understand that such shifts will require complex organizational changes in most schools, compared to which the assessment adaptations may be relatively straightforward. In that vein, we note that—as Louisiana is now working to develop for English-language arts—it should be

possible to construct and deploy a series of end-of-unit assessments that are both formative from the instructors' perspective and summative for students, assessments that cumulate until an entire course or course sequence has been mastered, whenever that may be.

True Readiness

With EOCs in place for key high school courses, requiring that they be passed at a satisfactory level becomes an excellent way of ensuring that a diploma attests to actual accomplishment. Ideally, the diploma—earned on the basis of mastery, not course credits—will be aligned with college/career readiness. States may, however, determine that that's unrealistic and opt instead to designate one score as establishing graduation credit while a higher score, validated against actual college expectations, qualifies for "distinction" or "college ready" on the diploma. This means some form of a two-tiered diploma system may be needed for a period of time.

If EOCs are employed in Career and Technical Education (CTE) as well as academic subjects, schools, districts, and states can approach the challenge of documenting students' career readiness as well as their college readiness, mindful that often these will not be the same students. A comprehensive CTE program goes far beyond the K–12 system, of course, and includes apprenticeships, postsecondary study, industry certification, and more.

While some students will struggle to attain readiness—and education systems should be organized to give them the time and help they require—others will demonstrate their readiness before the end of twelve traditional grades, and we encourage arrangements that enable those students to continue moving forward. That argues for providing several different pathways that may be pursued during the remaining years of high school, which may include dual-credit, early-college, Advanced Placement, and various CTE options.

For accountability purposes at the high school level, however, states are advised to weigh both a school's success in getting all students to the passing level on the EOCs (and thence to graduation) and its success in getting as many as possible to the college/career readiness threshold and beyond. When unified data systems make it possible to track students' progress beyond high school—into the adult worlds of college and career—it's desirable to incorporate valid gauges of that progress into high school accountability calculations.

When unified data systems make it possible to track students' progress beyond high school—into the adult worlds of college and career—it's desirable to incorporate valid gauges of that progress into high school accountability calculations.

Analysis and Reporting

Achievement and growth should remain the core criteria by which school performance is evaluated. We recommend continuing with ESSA-style data gathering on student learning, based primarily on external exams developed by states and aligned with their academic standards. Results should be disaggregated by student group and reported at the school, district, and state levels, as should growth whenever that can be calculated. Data on individual students should be provided to parents and teachers, and we encourage states and districts to incorporate achievement and growth data in teacher evaluations whenever possible. Schools should continue to be assigned easily understood ratings based on their performance, and information about them should be made available, as under ESSA, on both report cards and dashboards.

As for Consequences . . .

Accountability systems aren't worth the bother unless they lead to actions that boost academic outcomes for schools and students—all schools and students, of course, but especially low performers and those with egregious achievement gaps.

Accountability systems aren't worth the bother unless they lead to actions that boost academic outcomes for schools and students.

Going forward, two approaches deserve consideration. Both depend on quality data and transparency about student and school performance. One strives to improve upon the familiar thrust of federal policy in recent decades, which is to intervene in low-performing schools with an eye to making them better. The other stops trying to "fix" troubled schools from the outside and instead relies on informed choices by parents to place their children in better schools.[65]

Neither approach is foolproof, neither has a perfect track record, and it's possible to refine, vary, and amalgamate them in ways that may be mutually reinforcing—but first let's recap what a reasonably pure form of each might consist of.

"Intervention" assumes that authoritative outsiders can dictate changes in weak schools that lead to greater effectiveness and ultimately to higher pupil achievement. Those changes may range from low-key (as in "develop an improvement plan") to forceful, such as by replacing staff, imposing a different curriculum, outsourcing the school's operation, converting it to charter status, or even closing it entirely (and possibly opening a fresh-start school in the same facility). Such changes may be automatic, triggered by an NCLB-style formula, or they may be discretionary and customized. They may be imposed directly by the state or made via the district or charter network within which the school is located. Actually making

such changes, however—particularly the forceful kind—requires congenial regulatory and political conditions, for it must be assumed that schools and those close to them will usually resist externally imposed change.

By contrast, "informed choice" relies on market forces to yield stronger school achievement by giving families opportunities to extricate their children from weak schools and send them instead to stronger ones. Along the way, these dynamics may also induce changes in weak schools—driven by the need to retain market share—and encourage strong schools to grow or replicate to accommodate more students. There's some evidence that all of these things actually happen. But there are preconditions here, too: success hinges on policies that enable an array of choices to be made, on sufficient population density to warrant the existence of multiple schools (augmented with virtual options), on the existence or creation of sufficient capacity in strong schools, on well-informed parents, on nondiscriminatory school-entry procedures, on adaptable financing formulae, and on default provisions to ensure that no child is left without an acceptable school to attend.

Beyond their common reliance on clear and transparent information about school and student performance, these two approaches to accountability are quite different and it's likely that, if Uncle Sam were to get out of the way, some states would opt for one and some for the other. As our experience with ESSA has shown, the political cultures and recent histories of American states differ greatly in all these ways. Some would also favor variations, combinations, or hybrids. Some may forgo results-driven accountability altogether, which in our view would do long-term harm to students, schools, and states alike.

Will Consequences Work?

Neither approach is a sure thing. The evidence to date on externally mandated school improvements, turnarounds, and takeovers is far from encouraging.[66] Put simply, the gentler versions, though politically more acceptable, rarely lead to significant changes, while the sterner versions are likelier to yield better schools but are so fiercely resisted by numerous stakeholders as to be rare indeed. As the NCLB experience demonstrated, when faced with a menu of possible interventions, state and district practices tend to gravitate to the flabbiest and least intrusive—hence least effective—options, expensive as even these may be. What typically happens is that avoidance of short-term pain for educators and politicians prevails over the needs of students, families, and communities.

As for educational transformation via parental choice, in addition to the political pushback that it usually engenders, experience has revealed marketplace frailties on both the supply and demand sides.

When faced with a menu of possible interventions, state and district practices tend to gravitate to the flabbiest and least intrusive—hence least effective—options, expensive as even these may be.

The opportunity to create new schools—private, charter, district, etc.—has yielded some remarkably fine offerings but has also beckoned both profit-seeking hucksters and earnest, well-meaning individuals and groups that turn out to be ill prepared for the complexities of running an effective school. Hence the need for some quality control and oversight—but such measures inevitably constrain supply. Thus arises the difficult trade-off that states have faced as they've allowed charter schools to open over the past two decades: Lots of schools with mixed quality, or fewer but better?

On the demand side, parents sometimes turn out to be unfussy regarding the educational effectiveness of the schools in which they place their children. That's due sometimes to inexperience making such choices, sometimes to satisfaction with the familiar, sometimes to settling—understandably—for a school that's safe, convenient, and welcoming without focusing hard on academic value-add or measurable learning outcomes. Parents must weigh many factors and may not be sophisticated "school shoppers," especially when the available information on school performance is complex, sometimes misleading, and frequently difficult even to locate.

More positively, there's evidence from a longitudinal study of parents choosing private schools in the District of Columbia with the help of vouchers that, as they gain experience with making choices, most become more discerning and demanding consumers.[67]

More negatively, a perceptive essay by Naomi Schaefer Riley reminds us that a subset of children—such as those in foster care—do not have well-functioning parents or other adults in their lives who can manage a complex system of school choices. (Hence the need for "default" arrangements.)[68]

In Combination

Implicit in the discussion above is a form of interdependence that enables intervention-style and marketplace-driven consequences to reinforce each other.

When interventions don't yield better outcomes, it's folly—educationally pointless and, we think, immoral—to force kids to remain in failing schools if better ones can be made available to them.

When interventions don't yield better outcomes, it's folly—educationally pointless and, we think, immoral—to force kids to remain in failing schools if better ones can be made available to them. That calls for choices and for the market-style mechanisms that enable families to exercise choice.

But because the marketplace itself doesn't necessarily yield quality schools, some steps toward quality control need to be taken by the state, and these begin with standards, testing, and transparent reporting of the performance of every school as well as regulatory interventions to ensure that shoddy schools don't remain open and every child has an acceptable school to attend.

Hence serious accountability points toward a blending of the two forms of consequence—but states may not want to stop there. Particularly in view of mounting resistance to test-based accountability of the traditional kind—the "gotcha" kind—policy makers may also want to attend to issues of school capacity, which begin with the competence, expertise, and organizational structures needed to foster excellence and achievement. Any discussion of capacity also points to equity considerations. Some schools and communities are far more generously resourced than others, and many schools are charged with educating traditionally underserved pupil populations, youngsters whose achievement prospects hinge in no small part on the extent and quality and efficacy of services provided to them.

That fourth leg takes many forms—personnel, leadership, curriculum, pedagogy, data systems, governance, budget, ancillary services, and much more—that go far beyond the scope of this paper and engage a host of "effective schools" and school-improvement issues. We offer one example here, emphasizing that such strategies do not typically count as part of accountability per se, but they may make it fairer and more palatable—and may also lead to better-functioning schools for children whose futures depend on them.

Consider Inspecting Troubled Schools

An "inspectorate" system, done properly, begins with a team of well-trained individuals who spend enough time in a faltering school to collect accurate and comprehensive data on what is and isn't working well among the school's many moving parts. They can be thought of as expert diagnosticians who combine professional judgment, good analytic skills, keen eyes, and a reasonable grasp of reputable "what works" literature.

On-site inspections are widely used overseas to evaluate schools, generally in countries where families already have the right to choose among them. One of the stronger models is England's Ofsted, which deploys experienced education leaders to conduct site visits, during which they review accountability data as well as current conditions in the school via classroom observations, interviews, and parent and student surveys.[69]

The visit leads to a school-improvement plan, which serves as both blueprint for change and baseline against which the school's progress can be monitored, with reinspections conducted, and additional data collected, on a set schedule to gauge how the repair effort is going. But a well-functioning inspectorate system is not just a simple exercise in site visits, number crunching, planning, and revisiting. If inspectors find that the renewal effort is going

We urge state and federal officials not to forsake results-based accountability or to shun high-quality assessments of student learning as an indispensable source of essential information as to where and how well those results are being achieved. To forgo them would cause K–12 education in America to fly blind.

badly and seems unlikely to get better, they can recommend additional consequences. Those might take the form of technical assistance for the school, an externally mandated restructuring of some kind, and/or the provision of better alternatives for the children attending that school. (Though parents may already have the right to move their children, that's no guarantee that accessible, quality alternatives exist and have room for more students.)

In that way, the inspection sequence adds a human face—and professional judgments—to what otherwise appears to be dry, data-driven, and test-score-heavy accountability, thus affording troubled schools help in understanding what is and isn't working and developing plans to rectify the situation.

Although evidence is mixed as to the achievement-boosting efficacy of school inspections in and of themselves, a tight coupling of well-trained inspectors, solid data, clear needs assessments, professionally informed improvement plans, and calibrated consequences may both lead to better academic outcomes and provide parents with a richer picture of their children's schools than is supplied by cold data on a school report card.

CONCLUSION

We urge state and federal officials not to forsake results-based accountability or to shun high-quality assessments of student learning as an indispensable source of essential information as to where and how well those results are being achieved. To forgo them would cause K–12 education in America to fly blind, like a plane in the fog without instruments, unable to determine where it is in relation to its desired destination, and would return us to a pre-Coleman era of school inputs, promises, and processes. That would sentence America to remain "at risk" due to the weak achievement of its students and the inadequate performance of its schools. It would be as if the National Commission on Excellence in Education had never warned us, almost four decades ago, that "if an unfriendly foreign power had attempted to impose on America the mediocre educational performance that exists today, we might well have viewed it as an act of war. As it stands, we have allowed this to happen to ourselves."[70]

Yet we "ourselves" can change that destiny if we steel ourselves not to tolerate failure.

ACKNOWLEDGMENTS

This paper draws heavily from three reviews of the relevant research literature: Thomas S. Dee, "Learning from the Past: School Accountability before ESSA"; Paul Manna and Arnold Shober, "Answering the Call? Explaining How States Have (or Have Not) Taken Up the ESSA Accountability Challenge"; and David Steiner and Alanna Bjorklund-Young, "The Future of School Accountability," all Hoover Institution, 2020, https://www.hoover.org/hesi-initiative. Those authors, as well as Nora Gordon, Michael McShane, and Andrew Smarick, plus members of the executive committee commented on the papers at a video symposium held on April 2, 2020. Many of the findings and recommendations in this paper are taken directly from those research papers and I thank the authors for their permission to do so, while I retain full responsibility for the final product. Thanks also to Pedro Enamorado for invaluable assistance in the preparation and completion of this paper, a draft of which was discussed with the executive committee and the HESI Practitioner Council at a virtual symposium held on July 23, 2020. My presentation that day was facilitated by the technical expertise of Marianne Lombardo. And I'm deeply indebted, for help at every stage, to HESI executive director and Hoover distinguished policy fellow Christopher Ruszkowski.

NOTES

1 "Roadmap for Next-Generation State Accountability Systems," Council of Chief State School Officers, 2020, https://ccsso.org/resource-library/roadmap-next-generation-state-accountability-systems.

2 "Measuring Growth in 2021: What State Leaders Need to Know," Data Quality Campaign, Alliance for Excellent Education, and Collaborative for Student Success, 2020, https://dataqualitycampaign.org/resource/measuring-growth-in-2021; and "Civil Rights, Education and Immigration Organizations Urge Assessments to Support Student Success," Education Trust, September 3, 2020, https://edtrust.org/press-release/civil-rights-education-and-immigration-organizations-urge-assessments-to-support-student-success.

3 Katie Pearce, "An Enduring Legacy: The Study That Changed the Conversation on Education Equality," Johns Hopkins University, September 30, 2016, https://hub.jhu.edu/2016/09/30/coleman-report-at-50-conference; and Matt Barnum, "50 Years Later, What America Still Hasn't Learned from the Coleman Report," *The 74*, October 16, 2016, https://www.the74million.org/article/50-years-later-what-america-still-hasnt-learned-from-the-coleman-report.

4 In 1981, Diane Ravitch and I founded the Educational Excellence Network (precursor of today's Thomas B. Fordham Institute), a loose network of individuals who saw too little excellence coming out of American schools. That year I also published, in *Life*, "A Call for Quality Education," asserting that "Literacy is in decline in one state after another."

5 Christine Patterson, "Measuring the Lasting Impact of a Nation at Risk," Walton Family Foundation, April 2018, https://www.waltonfamilyfoundation.org/stories/k-12-education/measuring-the-lasting-impact-of-a-nation-at-risk.

6 Julie A. Miller, "Educators Hail Nomination of Alexander as Secretary," *Education Week*, January 9, 1991, https://www.edweek.org/ew/articles/1991/01/09/10090030.h10.html.

7 Effective Schools Research Base, Lake Forest College, https://foresternet.lakeforest.edu/academic-resources/library/archives/effective-schools-research-base; and Marshall S. Smith and Jennifer O'Day, "Systemic School Reform," *Journal of Education Policy* 5, no. 5 (April 2008): 233–67, https://www.researchgate.net/publication/233218215_Systemic_school_reform.

8 Chester E. Finn, Jr., "Still Swimming in Lake Wobegon," Thomas B. Fordham Institute, July 20, 2005, https://fordhaminstitute.org/national/commentary/still-swimming-lake-wobegon.

9 Alyson Klein, "Historic Summit Fueled Push for K–12 Standards," *Education Week*, September 23, 2014, https://www.edweek.org/ew/articles/2014/09/24/05summit.h34.html.

10 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://www.nber.org/papers/w10591.

11 America 2000: An Education Strategy (Washington, DC: US Department of Education, May 1991), https://files.eric.ed.gov/fulltext/ED327985.pdf.

12 "Summary of the Improving America's Schools Act," *Education Week*, November 9, 1994, https://www.edweek.org/ew/articles/1994/11/09/10asacht.h14.html.

13 Marshall S. Smith, Brett W. Scoll, and Jeffrey Link, "Research-Based School Reform: The Clinton Administration's Agenda," in *Improving America's Schools: The Role of Incentives* (Washington, DC: National Research Council, National Academies Press, 1996), https://www.nap.edu/read/5143/chapter/3.

14 Thomas S. Dee, "Learning from the Past: School Accountability before ESSA," Hoover Institution, 2020, https://www.hoover.org/research/learning-past-school-accountability-essa.

15 Tiffany Danitz, "Texas, North Carolina Seen as Models of Education Reform," Pew Charitable Trusts, November 2, 1999, https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/1999/11/02/texas

-north-carolina-seen-as-models-of-education-reform; and David W. Grissmer, Ann Flanagan, Jennifer H. Kawata, and Stephanie Williamson, "Improving Student Achievement: What State NAEP Test Scores Tell Us," RAND Corporation, 2000, https://www.rand.org/pubs/monograph_reports/MR924.html.

16 Dee, "Learning from the Past."

17 Dee, "Learning from the Past."

18 "Putting the Pieces Together: Systemic School Reform," Consortium for Policy Research in Education, 1991, https://eric.ed.gov/?id=ED343215; Joseph O. Oluwole and Preston C. Green III, "State Takeovers of School Districts: Race and the Equal Protection Clause," *Indiana Law Review* 42, no. 2 (January 2009), http://journals.iupui.edu/index.php/inlawrev/article/view/3995; Patrick O'Donnell, "State Would Take Over Schools after Six Years of Failing to Improve, under New Plan," *The Plain Dealer*, September 10, 2019, https://www.cleveland.com/news/2019/09/state-would-take-over-schools-after-six-years-of-failing-to-improve-under-new-plan.html; and Beth E. Schueler, "School District Turnaround: Learning from Leadership in Lawrence, Massachusetts," Harvard Kennedy School, January 2018, https://www.hks.harvard.edu/sites/default/files/centers/rappaport/files/thirdway%20v5.pdf.

19 Alex Spurrier, Chad Aldeman, Jennifer O'Neal Schiess, and Andrew J. Rotherham, "The Impact of Standards-Based Accountability," Bellwether Education Partners, Summer 2020, https://bellwethereducation.org/sites/default/files/Bellwether_Accountability-Impact_Final.pdf.

20 David A. Stuit, "Are Bad Schools Immortal?," Thomas B. Fordham Institute, 2010, https://fordhaminstitute.org/national/research/are-bad-schools-immortal.

21 A thoughtful, candid, research-based treatment of such improvement efforts can be found in Anthony S. Bryk's *Improvement in Action: Advancing Quality in America's Schools* (Cambridge, MA: Harvard Education Press, 2020).

22 Matthew M. Chingos, Katharine M. Lindquist, and Grover J. "Russ" Whitehurst, "Evaluating Teachers with Classroom Observations: Lessons Learned in Four Districts," Brookings Institution, May 2014, https://www.brookings.edu/research/evaluating-teachers-with-classroom-observations-lessons-learned-in-four-districts.

23 Donald T. Campbell, "Assessing the Impact of Planned Social Change," *Evaluation and Program Planning* 2, no. 1 (1979): 67–90, https://doi.org/10.1016/0149-7189(79)90048-X.

24 This section borrows extensively from the excellent paper by Dee, "Learning from the Past."

25 Adam Tyner and Michael J. Petrilli, "The Case for Holding Students Accountable: How Extrinsic Motivation Gets Kids to Work Harder and Learn More," *Education Next* 18, no. 3 (2018): 26–33, https://www.educationnext.org/case-for-holding-students-accountable-how-extrinsic-motivation-gets-kids-work-harder-learn-more/.

26 Hanushek and Raymond, "School Accountability"; and Annika B. Bergbauer, Eric A. Hanushek, and Ludger Woessmann, "Testing with Accountability Improves Student Achievement," *Vox*, September 2018, http://hanushek.stanford.edu/publications/testing-accountability-improves-student-achievement.

27 Stephen P. Klein, Laura S. Hamilton, Daniel F. McCaffrey, and Brian M. Stecher, "What Do Test Scores in Texas Tell Us?," *Education Policy Analysis Archives* 8, no. 49 (2000), https://epaa.asu.edu/ojs/article/view/440.

28 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://www.mitpressjournals.org/doi/10.1162/EDFP_a_00189.

29 David Figlio and Susanna Loeb, "School Accountability," in *Handbook of the Economics of Education*, ed. Eric A. Hanushek, Stephen Machin, and Ludger Woessmann (Amsterdam: North-Holland, 2011) 3, 383–421, https://cepa.stanford.edu/sites/default/files/Accountability_Handbook.pdf.

30 Michael Hout and Stuart W. Elliott, eds., *Incentives and Test-Based Accountability in Education* (Washington, DC: National Research Council, National Academies Press, 2011), https://www.nap.edu/catalog/12521/incentives-and-test-based-accountability-in-education; Eric A. Hanushek, "The NRC Judges Test-Based Accountability,"

Education Next, June 3, 2011, https://www.educationnext.org/the-nrc-judges-test-based-accountability; and Eric A. Hanushek, "Grinding the Antitesting Ax," *Education Next* 12, no. 2 (Spring 2012), https://www.educationnext.org/grinding-the-antitesting-ax.

31 Figlio and Loeb, "School Accountability"; and Randall Reback, Jonah Rockoff, and Heather L. Schwartz, "Under Pressure: Job Security, Resource Allocation, and Productivity in Schools under No Child Left Behind," *American Economic Journal: Economic Policy* 6, no. 3 (August 2014): 207–41, https://www.aeaweb.org/articles?id=10.1257/pol.6.3.207.

32 Dee, "Learning from the Past."

33 Spurrier et al., "Standards-Based Accountability," 2.

34 Bergbauer, Hanushek, and Woessmann, "Testing with Accountability"; Eric A. Hanushek and Ludger Woessmann, "The Economics of International Differences in Educational Achievement," in *Handbook of the Economics of Education*, ed. Eric A. Hanushek, Stephen Machin, and Ludger Woessmann (Amsterdam: North-Holland, 2011) 3, 89–200, https://www.nber.org/papers/w15949.

35 The statute set forth a cascade of intensifying plans and actions to rectify a faltering school's failure but in the end retreated from forceful action to fundamentally alter the school. Here are the final-stage options for school overhaul:

> (i) Reopening the school as a public charter school.
>
> (ii) Replacing all or most of the school staff (which may include the principal) who are relevant to the failure to make adequate yearly progress.
>
> (iii) Entering into a contract with an entity, such as a private management company, with a demonstrated record of effectiveness, to operate the public school.
>
> (iv) Turning the operation of the school over to the State educational agency, if permitted under State law and agreed to by the State.
>
> (v) Any other major restructuring of the school's governance arrangement that makes fundamental reforms, such as significant changes in the school's staffing and governance, to improve student academic achievement in the school and that has substantial promise of enabling the school to make adequate yearly progress as defined in the State plan.

It was, of course, the "any other" option that generally got chosen and implemented.

36 John Cronin, Michael Dahlin, Yun Xiang, and Donna McCahon, "The Accountability Illusion," Thomas B. Fordham Institute, 2009, https://fordhaminstitute.org/national/research/accountability-illusion.

37 Paul Manna and Arnold Shober, "Answering the Call? Explaining How States Have (or Have Not) Taken Up the ESSA Accountability Challenge," Hoover Institution, 2020, https://www.hoover.org/research/answering-call-explaining-how-states-have-or-have-not-taken-essa-accountability-challenge.

38 This and following quotations from Manna and Shober are in "Answering the Call?."

39 "ESSA Accountability Design Competition," Thomas B. Fordham Institute, February 2, 2020, https://fordhaminstitute.org/events/essa-accountability-design-competition; and Dana Goldstein, "Obama Education Rules Are Swept Aside by Congress," *New York Times*, March 9, 2017, https://www.nytimes.com/2017/03/09/us/every-student-succeeds-act-essa-congress.html.

40 This section draws heavily on the work of HESI contributors David Steiner and Alanna Bjorklund-Young in "The Future of School Accountability," Hoover Institution, 2020, https://www.hoover.org/research/future-school-accountability-0.

41 "Researching Florida's A–F system found schools facing accountability pressure changed their instructional practices in meaningful ways, which explained some of the test score gains": Cecilia Elena Rouse, Jane Hannaway,

Dan Goldhaber, and David Figlio, "Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure," *American Economic Journal: Economic Policy* 5, no. 2 (May 2013): 251–81, https://www.nber.org/papers/w13681. It also turns out that publicly available letter grades make a difference in other realms, such as the health inspection of restaurants: https://nyti.ms/2xKxb4P.

42 Out-of-school influences such as parents, race, and socioeconomic circumstances obviously continue to have powerful effects on student achievement but somewhat less on competent gauges of achievement *growth.*

43 For an example, see Maryland's use of the Bridge Project: Karen B. Salmon, "Maryland Bridge Plan for Academic Validation," Maryland State Board of Education, February 25, 2020, http://marylandpublicschools.org/stateboard/Documents/2020/0225/TabJ_BridgePlan.pdf.

44 Karen B. Salmon, "Every Student Succeeds Act (ESSA) Implementation Update," Maryland State Board of Education, September 2018, https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjQzIa11pHsAhUQoHIEHXVxBPMQFjAAegQIBhAB&url=http%3A%2F%2Fmarylandpublicschools.org%2Fstateboard%2FDocuments%2F09252018%2FTabI-ESSAImplementationUpdate.pdf&usg=AOvVaw0vAjDpsoxZrnPXFAroe7_u; and "2019 School Grades Overview," Florida Department of Education, accessed September 21, 2020, http://www.fldoe.org/core/fileparse.php/18534/urlt/SchoolGradesOverview19.pdf.

45 Henry Braun, Naomi Chudowsky, and Judith Koenig, eds., *Getting Value out of Value-Added: Report of a Workshop* (Washington, DC: National Research Council, National Academies Press, 2010), https://www.nap.edu/catalog/12820/getting-value-out-of-value-added-report-of-a-workshop; and Chris Balow, "7 Methods for Measuring Student Growth," *Illuminate Education*, December 14, 2017, https://www.illuminateed.com/blog/2017/12/7-methods-measuring-student-growth.

46 Jörn-Steffen Pischke, "The Impact of Length of the School Year on Student Performance and Earnings: Evidence from the German Short School Years," *Economic Journal* 117, no. 523 (October 2007): 1216–42, https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0297.2007.02080.x; and Dave E. Marcotte and Steven W. Hemelt, "Unscheduled School Closings and Student Performance," *Education Finance and Policy* 3, no. 3 (2008): 316–38, https://eric.ed.gov/?id=EJ851520.

47 Indiana's panel on school accountability has set forth an exemplary version of accountability for Hoosier high schools. See Carley Lanich, "State Panel Recommends Changes for How High Schools Receive State A-F Accountability Grades," *Northwest Indiana Times*, November 4, 2019, https://www.nwitimes.com/news/education/state-panel-recommends-changes-for-how-high-schools-receive-state-a-f-accountability-grades/article_4f3d1321-d2ed-58de-8826-41008f89349b.html.

48 David Griffith and Michael J. Petrilli, "How States Should Redesign Their Accountability Systems under ESSA," Thomas B. Fordham Institute, November 10, 2016, https://fordhaminstitute.org/national/commentary/how-states-should-redesign-their-accountability-systems-under-essa.

49 "Millennial Parents and Education," Walton Family Foundation, Echelon Insights, 2018, https://8ce82b94a8c4fdc3ea6d-b1d233e3bc3cb10858bea65ff05e18f2.ssl.cf2.rackcdn.com/1f/b9/c41cff2942ff9e9fba4f54b06794/millennial-parents-12.2018%20FINAL.pdf.

50 "Parents 2018: Going beyond Good Grades," Learning Heroes, December 2018, https://r50gh2ss1ic2mww8s3uvjvq1-wpengine.netdna-ssl.com/wp-content/uploads/2018/12/2018_Research_Report-final_WEB.pdf.

51 Angela L. Duckworth and David Scott Yeager, "Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes," *Educational Researcher* 44, no. 4 (2015): 237–51, https://www.aera.net/Newsroom/Recent-AERA-Research/Measurement-Matters-Assessing-Personal-Qualities-Other-Than-Cognitive-Ability-for-Educational-Purposes; and Joshua Cox, Brandon Foster, and David Bamat, "A Review of Instruments for Measuring Social and Emotional Learning Skills among Secondary School Students. REL 2020-010," US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands, 2019, https://eric.ed.gov/?id=ED599014.

52 One can imagine an accountability regimen that judges primary and middle schools according to the performance of their graduates in the next level of schooling, but frequent changes in school operations—from staffing to instructional initiatives, for example—as well as student mobility create major hurdles in developing and deploying such measures.

53 "State Report Card Requirements," Penn Hill Group, Council of Chief State School Officers, January 26, 2016, https://ccsso.org/resource-library/state-report-card-requirements; "ED Releases Draft ESSA Report Card Guidance," Brustein & Manasevit PLLC, ESEA Network, April 1, 2019, https://www.eseanetwork.org/news-and-resources/blogs/legislation/ed-releases-draft-essa-report-card-guidance; "A Parent Guide to State and Local Report Cards," US Department of Education, November 2018, https://www2.ed.gov/policy/elsec/leg/essa/parent-guide-state-local-report-cards.pdf; and "The Every Student Succeeds Act: What's in It? What Does It Mean for Equity?," Education Trust, January 2016, https://edtrust.org/wp-content/uploads/2014/09/What-is-in-ESSA-Public-Reporting.pdf.

54 See, for example, KIPP, Regional Report Card, https://www.kipp.org/results/regional.

55 Sometimes, as in New York City, a school's dashboard data may be too dense for anyone but an expert to mine (https://tools.nycenet.edu/dashboard). In such situations, a well-crafted school website, though possibly not useful for apples-to-apples comparisons with other schools, may be more informative for parents, teachers, and others seeking information about what actually happens in that school. Also helpful are third-party websites that depict some school information in uniform fashion, such as those maintained by GreatSchools.org.

56 "Guide to 2019 Ohio School Report Cards," Ohio Department of Education, September 2019, http://education.ohio.gov/getattachment/Topics/Data/Report-Card-Resources/Report-Card-Guide.pdf.aspx; and "California School Dashboard," California Department of Education, 2017, https://www.caschooldashboard.org.

57 Lynn Olson and Craig Jerald, "Statewide Standardized Assessments Were in Peril Even before the Coronavirus. Now They're Really in Trouble," *Education Next*, April 21, 2020, https://www.educationnext.org/statewide-standardized-assessments-were-in-peril-before-coronavirus-bipartisan-backlash. See also the authors' longer study at Lynn Olson and Craig Jerald, "The Big Test: The Future of Statewide Standardized Assessments," *FutureEd*, April 2020, https://www.future-ed.org/the-big-test-the-future-of-statewide-standardized-testing/. Also relevant is Margaret E. Raymond and Eric A. Hanushek, "Shopping for Evidence against School Accountability," in *Developments in School Finance: 2003*, ed. William J. Fowler Jr. (Washington, DC: National Center for Education Statistics, 2003), 119–30, http://hanushek.stanford.edu/publications/shopping-evidence-against-school-accountability.

58 Matthew M. Chingos, "Testing Costs a Drop in the Bucket," Brookings Institution, February 2, 2015, https://www.brookings.edu/blog/up-front/2015/02/02/testing-costs-a-drop-in-the-bucket.

59 Valerie Strauss, "Confirmed: Standardized Testing Has Taken Over Our Schools. But Who's to Blame?," *Washington Post*, October 24, 2015, https://www.washingtonpost.com/news/answer-sheet/wp/2015/10/24/confirmed-standardized-testing-has-taken-over-our-schools-but-whos-to-blame.

60 Mark Teoh, Celine Coggins, Christine Guan, and Tamara Hiler, "The Student and the Stopwatch: How Much Time Do American Students Spend on Testing?," *Teach Plus*, February 2014, https://teachplus.org/news-events/publications/student-and-stopwatch-how-much-time-do-american-students-spend-testing.

61 For an extended discussion, see this companion HESI paper: Margaret E. Raymond, "The Diploma Dilemma," Hoover Institution, February 2020, https://www.hoover.org/research/diploma-dilemma.

62 Martin R. West and Ludger Woessmann, "How School Choice, Autonomy, and Accountability Impact Student Achievement: International Evidence," in *Balancing Freedom*, *Autonomy and Accountability in Education*, ed. Charles L. Glenn and Jan De Groof (Tilburg, Netherlands: Wolf Legal Publishers, 2012), 4, 275–98, https://edpolicy.education.jhu.edu/wp-content/uploads/2019/08/v4-international-evidence.pdf.

63 Kindergarten Readiness Assessment (KRA), Maryland State Department of Education, http://www.marylandpublicschools.org/about/Pages/DAAIT/Assessment/KRA/index.aspx, accessed September 21, 2020.

64 Federal law currently requires that the testing begin in grade three, and state testing systems are accustomed to that. A number of states also have third-grade "reading guarantee" policies. We believe, however, that the end of second grade is preferable to commence the assessment of early reading.

65 Paul Peterson, "Toward Equitable School Choice," Hoover Institution (forthcoming).

66 Andy Smarick, "The Turnaround Fallacy," *Education Next* 10, no. 1 (Winter 2010), https://www.educationnext.org/the-turnaround-fallacy; and Stuit, "Are Bad Schools Immortal?."

67 Thomas Stewart and Patrick J. Wolf, *The School Choice Journey: School Vouchers and the Empowerment of Urban Families* (New York: Palgrave Macmillan US, 2014).

68 Naomi Schaefer Riley, "In Defense of 'Paternalism': Why Choice Schools Need to Get Better at Grappling with the Family Lives of Their Toughest-Case Students," *The 74*, February 23, 2020, https://www.the74million.org/article/in-defense-of-paternalism-why-choice-schools-need-to-get-better-at-grappling-with-the-family-lives-of-their-toughest-case-students-2.

69 This section draws extensively on Ashley Berner, "Would School Inspections Work in the United States?," Johns Hopkins University, Institute for Education Policy, September 2017, https://edpolicy.education.jhu.edu/would-school-inspections-work-in-the-united-states; and Craig D. Jerald, "On Her Majesty's School Inspection Service," Education Sector Reports, 2012, https://stuartfoundation.org/wp-content/uploads/2012/11/UKInspections-RELEASED.pdf.

70 "A Nation at Risk," April 1983, https://www2.ed.gov/pubs/NatAtRisk/risk.html.

26 25 24 23 22 21 20 7 6 5 4 3 2 1

About the Author

CHESTER E. FINN, JR.

Chester E. Finn, Jr., is a senior fellow at the Hoover Institution and a distinguished senior fellow and president emeritus at the Thomas B. Fordham Institute. He previously served as assistant secretary for research and improvement, and counselor to the cabinet secretary at the US Department of Education; and as legislative director for Senator Daniel P. Moynihan.

Hoover Education Success Initiative

With passage in 2015 of the Every Student Succeeds Act (ESSA), states are again in charge of American education policy. To support them in this undertaking, the Hoover Education Success Initiative (HESI), launched in 2019, seeks to provide state education leaders with policy recommendations that are based upon sound research and analysis.

HESI hosts workshops and policy symposia on high-impact areas related to the improvement and reinvention of the US education system. The findings and recommendations in each area are outlined in concise topical papers.

The leadership team at HESI engages with its Practitioner Council, composed of national policy leaders, and with interested state government leaders. HESI's ultimate goal is to contribute to the ongoing transformation of the nation's education landscape and to improve outcomes for our nation's children.

For more information about the Hoover Education Success Initiative, visit us online at hoover.org/hesi.

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

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

HOOVER INSTITUTION