A BACKGROUND PAPER FOR THE HOOVER EDUCATION SUCCESS INITIATIVE

Learning from the Past

SCHOOL ACCOUNTABILITY BEFORE ESSA

THOMAS S. DEE

Introduction

The increasing sophistication of information technology and the availability of rich data have contributed to the growing prevalence of data-driven performance monitoring across a diverse array of personal and institutional interactions. The complex and decentralized system of K–12 public schools in the United States is no exception. The broad availability of data on US public schools reflects an interest in supporting the mission of these important public institutions and in holding them accountable. Over the last half century, a variety of state and federal initiatives to promote accountability in education have relied on such data systems as well as encouraged their growth. The focal points of these varied accountability policies have included students (e.g., high-school exit exams), school districts, and, more recently, teachers. However, the dramatic and controversial expansion of *school*-focused accountability systems is arguably the signature development in education policy over the last three decades. In particular, nearly twenty years ago, the federal government brought test-based school accountability to a national scale through the No Child Left Behind (NCLB) Act. And the recent reauthorization of NCLB as the Every Student Succeeds Act (ESSA) provides a national framework for the latest developments in the complex evolution of state accountability systems focused on multiple dimensions of school performance.

In this paper, I provide an overview of the evolution and design of school performance accountability systems in the United States over the last fifty years. And I provide a brief survey of the literature on the effects of these reforms. I conclude with a discussion of the implications of this body of evidence for the state of school accountability systems as they now exist under ESSA. I note that under ESSA's flexibility, states have created strikingly diverse accountability regimes whose design features reproduce many of the important traits of prior state reform efforts. On the basis of the evidence on the impact of these prior reforms, I suggest that many of the state accountability systems currently being implemented under the increased autonomy provided by ESSA have worrisome features that are likely to attenuate their capacity for positive impact. For example, under ESSA, a number of states fail to articulate meaningful consequences for underperforming schools (i.e., even at the level of a clearly informative summative rating of a school's performance). Second, several current state accountability systems place comparatively little emphasis on the performance of key student subgroups (e.g., by race, ethnicity, and socioeconomic status). Third, recent evidence suggests that effective school reform relies critically on the

capacity of states and districts to articulate clear, evidence-based strategies and to provide energetic guidance that reaches schools and classrooms. However, ESSA appears to do little to support this sort of system capacity. For example, ESSA does require that targeted school reforms find support in explicit "tiers" of evidence. However, the striking breadth and flexibility of this evidentiary guidance may render this requirement meaningless. The empirical relevance of all these concerns will become more apparent in the near future as our experience with ESSA's flexibility unfolds and informs sensible adaptations of these varied accountability systems.

The Evolution of Education Accountability in the United States

State and federal efforts to promote accountability in public education followed at least two notable historical antecedents. One was the sharp growth during the twentieth century in the state and federal support for local public schools. In 1920, the share of public-school revenues from state and federal sources was 17 percent. However, a century later, state and federal sources instead contributed roughly 55 percent of school funding. Second, the Equality of Educational Opportunity study commissioned by Congress and released in 1966 dramatically reshaped how we understood those investments in public schools.[1] The "Coleman Report" provided important evidence on racial segregation, racial gaps in student achievement, and inequities in access to some school resources (e.g., textbooks, science laboratories, advanced curricula). More surprisingly, the Coleman Report also reported that the inequities in school resources did not appear to influence meaningfully the variation in school performance, noting "One implication stands out above all: That schools bring little influence to bear on a child's achievement that is independent of his background and general social context; and that this very lack of an independent effect means that the inequalities imposed on children by their home, neighborhood, and peer environment are carried along to become the inequalities with which they confront adult life at the end of school."[2] The eventual broad dissemination of this finding seeded concerns about the organizational performance of public schools and a long-lived controversy over conventional education policies that had focused on school inputs.

These concerns also informed the conceptual foundations for the existence and design of accountability systems. For example, a common motivation for school-focused accountability involves a distrust of conventional, input-based education policies and the parallel concern that parents and taxpayers cannot easily monitor the performance of their school districts. Such "information asymmetry" can allow for an equilibrium in which school leaders and teachers make self-interested decisions that do not necessarily contribute to school effectiveness. In theory, accountability systems facilitate oversight and create incentives that support school improvement. The growing prevalence of public-school choice (e.g., charters, intra-district choice) may have also contributed to the growing interest in performance data and school accountability. However, another more benign motivation for accountability systems is that schools may simply not be very well informed about the

outcomes that are uniquely salient to key education stakeholders. Similarly, in the absence of performance monitoring, districts and schools may have poor information on how well they are doing, particularly with regard to subgroups of students who have historically been served poorly by their schools.

The adoption and implementation of performance-based school accountability systems also raises several other important design considerations. In general, accountability systems have three necessary features. This "tripod" consists of the articulation of clear standards, the collection of aligned and valid performance measures, and productive consequences based on these measures.[3] These three policy choices are frequently the focus of both scrutiny and controversy. For example, the recent debate over the Common Core State Standards has underscored the challenge of promulgating ambitious and contemporary academic standards that are harmonized across states. However, the most prominent criticism of accountability systems is the view that linking consequences to performance measures may "corrupt the social process it is intended to monitor."[4] Opponents of test-based school accountability (e.g., teachers' unions) often point to "Campbell's Law" and argue that these polices should be abandoned because of their unintended consequences (e.g., narrowing curricula, the triaging of student support). However, it is worth pointing out that in his original article, Campbell did not frame this as a reason for abandoning performance measurement but rather as a problem to be solved: "We must develop ways of avoiding this problem if we are to move ahead. We should study the social processes through which corruption is being uncovered and try to design social systems that incorporate these features."[5] He also noted his belief that "the use of multiple indicators, all recognized as imperfect, will alleviate this problem." Interestingly, the widespread use of multiple indicators, each imperfect on its own, has become a hallmark of the evolution of school accountability I describe below.

Early accountability policies

In the wake of the Coleman Report, state-level policy innovations with a focus on standards, particularly at the high-school level, soon followed. Specifically, the "First Wave" of accountability began during the 1970s when several states introduced minimum competency exams that required students to demonstrate basic skills in order to graduate.[6] A broad motivation for these state policies was the growing perception that a high-school diploma, once a meaningful credential, no longer guaranteed student mastery of basic skills. This early form of student-facing accountability evolved into the exit-exam requirements currently active in several states. For the class of 2017, seventeen states required students to pass one or more exit exams in order to graduate.[7] The influential 1983 report *A Nation at Risk* also contributed to the enthusiasm for standards-based accountability. Drawing on both national and international comparisons, *A Nation at Risk* criticized the performance of US schools with strident rhetoric: "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."[8] The leading recommendation from this report was that states

require their high-school graduates to complete a "New Basics" curriculum that included four years of English; three years of mathematics, science, and social studies; and a half year of computer science. Nearly every state responded by increasing its course graduation requirements in core academic areas, though few implemented the ambitious "4/3/3/3" standard.[9]

The 1989 Education Summit between forty-nine state governors and President George H. W. Bush marked another watershed moment for standards-based reform.[10] The summit resulted in the articulation of national education goals and a National Education Goals Panel (NEGP) to oversee them. The 1989 summit was also notable in terms of the politics of federal education policy because it involved action by states and the president without congressional input. However, the federal government formally acknowledged national education goals and the NEGP's reporting responsibilities when President Clinton signed the Goals 2000: Educate America Act in 1994. Though Goals 2000 marked the articulation of ambitious standards and increased federal attention to schools, it also lacked explicit accountability mechanisms for meeting those goals. Similarly, the 1994 reauthorization of the Elementary and Secondary Education Act as the Improving America's Schools Act (IASA) brought federal encouragement to the development of state standards and state assessments as well as the alignment of these with curricula, instruction, and professional development. This period also saw the introduction of new state-level testing as part of the federal National Assessment of Educational Progress (NAEP) assessments.

State policy making responded to the enthusiasm for academic standards and assessments as well as to the federal encouragement. Beginning in the 1990s, most states began articulating their own academic standards, introducing statewide testing programs, and implementing different forms of test-based school accountability. On the eve of the No Child Left Behind (NCLB) Act, all fifty states tested their students and forty-seven states had academic standards in all four core subjects.[11] Most states also reported that these tests reflected student mastery of their ambitious academic standards. However, independent analyses suggested that some state tests actually covered only some of their academic standards and often did so with a focus on low-level skills and knowledge. By 2001, forty-five states had instituted accountability through requiring the publication of annual school "report cards." The most common basis for judging school performance at this time was student test scores. However, several states also relied on attendance rates and, for high schools, graduation rates. The use of site visits and interviews was much less common.

During this pre-NCLB era, far fewer states had adopted forms of accountability with any meaningful consequences beyond the public reporting of school performance data through report cards. In 2001, only seventeen 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 five states). Similarly, only a distinct

minority of states had clear authority for any sort of sanctions such as 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.

The No Child Left Behind (NCLB) era

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 United States 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 three through eight 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. However, NCLB also mandated the public rating of schools with respect to whether they were making "adequate yearly progress" (AYP) toward having all students at their state's "proficient" level by 2013–14. Schools that persistently failed to make AYP, even for a single subgroup of students, were also subject to a cascade of further sanctions that included allowing students to transfer to a better district school, offering supplemental services (e.g., free tutoring), and, after five years of failing to make AYP, restructuring (e.g., comprehensive school reform, closure, or reconstitution). However, as a practical matter, the implementation of these sanctions appears to have been bounded. The take-up of public-school choice among eligible students was low.[12] And schools required to take up restructuring often used the "any other major restructuring" option rather than the more prescriptive and politically contentious options defined in the law.[13]

Under "increasing pressure to grant states relief from NCLB's onerous accountability requirements," the Bush administration began offering states the opportunity to secure waivers from NCLB's key provisions.[14] Specifically, 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.[15] In 2010, the Obama administration's articulated a "Blueprint" for reauthorizing NCLB. However, this proposal was effectively "dead on arrival" in the wake of an increasingly rancorous political climate as well as other policy challenges (e.g., the Great Recession, the health care debate).[16] Instead, the Obama administration opted to make more extensive use of the waiver authority articulated in the NCLB Act. In 2011, the 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 seven states received these

federal waivers. Notably, this structured waiver process occurred within the context of other prominent federal efforts to promote particular policy innovations. Specifically, the US Department of Education used part of its 2009 stimulus spending to fund revamped School Improvement Grants (SIGs) and state competitions such as Race to the Top (RttT) and the Teacher Incentive Fund (TIF). Some stimulus-funded initiatives as well as the waiver process also encouraged states to adopt the Common Core State Standards and aligned testing. Adding further to the complex regulatory environment for schools and districts under NCLB and waivers, some states simultaneously maintained their own parallel and often incongruent school accountability systems.

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.[17] 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.[18] 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. The school-level indices developed by states featured the weighting of multiple factors in addition to achievement levels such as growth in student achievement, measures of college and career readiness, and of improvements in school climate.[19] However, relative to NCLB, these school performance indices were also based on fewer measures of subgroup performance or on more broadly defined subgroups. And, instead of a binary indicator for whether a school was achieving AYP, school ratings under waivers were based on multiple performance levels such as star ratings or letter grades.[20]

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. One group, "Priority Schools," consisted of the schools identified as persistently lowest performing in the state (i.e., 5 percent of the Title I schools in the state). The waiver process required states to implement one of several federally prescribed reforms in these schools (e.g., transformation, turnaround, restart, or closure). "Focus Schools," a second group, were defined as those with the lowest performance for specific subgroups or with the greatest within-school gaps in performance in the state (i.e., a minimum of 10 percent of a state's Title I schools). The federal regulations for Focus Schools were substantially less prescriptive than those for Priority Schools, asking only that states implement interventions that were "consistent with" federal turnaround principles or adopt any other "research-based" interventions to

meet the needs of students at the school. Notably, under these waivers, schools identified for improvement were no longer required to offer public school choice or supplemental services to their students.

The Every Student Succeeds Act (ESSA)

In late 2015, the era of NCLB and its waivers ended when President Obama signed the Every Student Succeeds Act (ESSA). The school-accountability requirements articulated in this long-anticipated reauthorization reflect a substantial degree of continuity with the design features of NCLB waivers as well as a delicate bipartisan compromise between those who wanted robust federal encouragement to promote school improvement and those who wanted to grant states more autonomy. In particular, ESSA requires states to articulate "challenging" academic standards in math, ELA, and science and the alignment of these standards with college and career readiness.[21] States are also required to continue fielding high-quality assessments aligned with these standards. ESSA also mandates the publication of annual report cards on school performance.[22]

However, ESSA requires that these school report cards reflect a breadth of information beyond the mandates of previous federal regulations, a clear indication of the broad dissatisfaction with NCLB's narrow focus on test-based proficiency in math and reading. Specifically, school report cards under ESSA must include several specified indicators. These include performance on math and English language arts (ELA) assessments relative to state goals but also a second academic indicator (i.e., graduation rates for high schools or a different test metric for elementary and middle schools), and a measure of the progress of English learners toward English proficiency. An additional indicator of school quality or student success is also required. Most states use a measure of chronic absenteeism. ESSA also allows measures of students' social-emotional learning (SEL) as an additional indicator. However, no state chose to use SEL measures, partly because of concerns that they are currently "unreliable and unusable for accountability purposes."[23] ESSA also requires that all of this school information be reported annually for both all students and multiple subgroups (i.e., major racial/ethnic groups, economically disadvantaged, children with disabilities, English learners, homeless students, students in foster care, and the children of active military members). Interestingly, ESSA further mandates reporting of per-pupil spending, a potentially interesting and politically controversial metric that has historically been unavailable at the school level.[24]

To comply with ESSA, states must define a system for the "annual meaningful differentiation of all public schools in the State" using performance data defined for both all students and major subgroups. States have latitude in how they map the underlying school indicators into an overall performance measure. According to data from the Education Commission of the States, forty-five states and the District of Columbia have defined accountability systems that apply summative ratings to schools.[25] Under NCLB, all states clearly labeled any school that failed to make progress toward the state's proficiency

standards as failing to make AYP. However, the character of the summative ratings developed by states under ESSA is more variegated. Thirteen states have introduced A through F ratings for schools. Another eleven states use descriptive ratings (e.g., Excellent, Good, Average, Needs Improvement) while four states (and Washington, DC) rate schools using stars (i.e., one to five stars). Twelve states rate schools using an index system (i.e., one to one hundred or one to ten). Four states (i.e., California, Idaho, North Dakota, and Oregon) have federally approved ESSA plans in which they provide *no* summative ratings for all public schools but rather focus on publishing a broad array of available indicators defined for both all students and subgroups. The exhaustive breadth of data and the manner in which the data are presented has raised some concerns. For example, on December 15, 2018, the editorial board of the *Los Angeles Times* argued that California's revised school dashboard, while an improvement in some ways, was "not helpful to parents or the public" because it can be "hard to parse" and "make schools look like they're doing a lot better than they are."[26]

Six other states rate schools by sorting them into "tiers of support" that are related to the targeted accountability that ESSA requires in all states. Specifically, like the Priority School and Focus School reforms required under NCLB waivers, ESSA requires states to "meaningfully differentiate" specific types of schools for targeted assistance and improvement.[27] However, ESSA requires the identification of three types of schools. Comprehensive Support and Improvement (CSI) schools must include the lowest-performing 5 percent of Title I schools in a state as well as high schools that fail to graduate at least two-thirds of their students. Targeted Support and Improvement (TSI) schools are those that have one or more persistently low-performing subgroups of students. ESSA gives states latitude in defining what constitutes persistent underperformance as long as it is based on all the accountability indicators. Additional Targeted Support and Improvement (ATSI) schools are those in which the performance of at least one subgroup of students on its own would lead to the school being identified as a CSI school.

Relative to previous federal policies, under ESSA, states have autonomy in choosing strategies for these targeted schools. ESSA only requires that the proposed interventions have evidence of effectiveness consistent with defined "tiers" of evidence (Section 8101[21]). Tiers 1 and 2 refer, respectively, to strategies supported by well-designed and implemented experimental and quasi-experimental studies. Tier 3 interventions find support in correlational studies with statistical controls while Tier 4 interventions need only a "rationale based on high-quality research findings or positive evaluation that such activity, strategy, or intervention is likely to improve student outcomes or other relevant outcomes." Schools that receive support from a federal School Improvement Grant (SIG) must choose an intervention supported by evidence in the first three tiers.

As newly designed accountability indicators have become available, states have begun identifying schools as CSI, TSI, or ATSI in accordance with their ESSA plans. However,

recent evidence indicates that states vary considerably in the share of schools they have identified in one of these three categories.[28] For example, Florida has identified 69 percent of its public schools and Rhode Island, 99 percent. In contrast, several states (e.g., Alabama, Connecticut, Georgia, Maryland, Massachusetts, Nebraska, Oklahoma, and Virginia) have identified 5 percent or less of their schools. Because states also allow for a planning period after being identified, the full implementation of school improvement activities is just now beginning or starting in the next year.[29]

The Effects of School Accountability

The extraordinary prominence and variation in accountability policies over the last four decades have motivated a large body of research studies that have sought to answer fundamental questions about their impact. Are such policies effective at measuring and reporting on school performance? Are such policies effective in improving school performance? In what ways might the design features of accountability policies have unintended as well as intended effects? For example, critics of accountability policies often emphasize the concern that the extrinsic incentives embedded in accountability systems may degrade intrinsic motivation.[30] The literature on student-facing incentives generally fails to find evidence in support of this concern.[31] In fact, there is some evidence that well-designed and targeted student incentives (e.g., rewarding behaviors students clearly control, leveraging loss aversion) can generate their intended effects.[32] However, there is also evidence that student accountability based on external standards has negative effects on student outcomes. Several studies find that restrictive high-school graduation requirements (e.g., exit exams and course graduation requirements) reduce the probability of completing high school, particularly for Black students.[33]

The more prominent concerns about the unintended consequences of accountability have focused on the school accountability systems brought to scale nationally by the federal government over the last twenty years. In particular, critics of NCLB and its state-level antecedents argue that test-based accountability narrows teaching to tested subjects and further to tested items within a subject. Consistent with this concern, states often had dramatic growth on high-stakes tests linked to accountability systems but without parallel gains on contemporaneous, low-stakes tests in the same subjects.[34] A related concern is that NCLB's focus on proficiency thresholds would lead to educational "triage" in which students well above or below the standard would receive little attention. The available empirical evidence does not provide empirical support consistent with this concern.[35] However, a review article by Figlio and Loeb discussed the evidence for other types of strategic responses to accountability incentives (e.g., teacher cheating, shaping the test-taking population).[36]

This research literature both underscores the importance of thoughtfully designed accountability systems and raises questions about whether accountability has been genuinely effective in improving school performance. Surveys of the studies that have examined the impact of accountability policies have privileged those with credible

quasi-experimental designs and a focus on low-stakes test outcomes.[37] In general, these studies found that accountability systems lead to meaningful, though not transformative, improvements in school performance. For example, Hanushek and Raymond examined the impact of pre-NCLB, state-level accountability systems on state-level growth in NAEP scores.[38] They found that "report card" accountability did not have statistically significant effects. However, the adoption of consequential accountability systems *did* lead to improved NAEP scores. Studies that examined the impact of differentiated incentives under NCLB found similar results. For example, Reback, Rockoff, and Schwartz leveraged the differences in the standards states set under NCLB to compare changes across schools that were subject to AYP pressure in one state but were not in others.[39] They found that AYP pressure increased student performance on low-stakes reading tests by 0.06 standard deviations with similarly positive and less precise gains in math and science. Like Hanushek and Raymond,[40] they did not find that these effects differed across subgroups. They also found no evidence that accountability pressure influences students' reported enjoyment of learning or test anxiety.

One methodological concern with studies based on accountability pressure is that the result impact estimates may be biased downward because they rely on comparisons among schools, all of which operate under the same accountability regime.[41] 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 with the changes in states where NCLB created entirely new experiences with test-based accountability.[42] They found that NCLB increased grade-four math scores on the NAEP by 0.23 standard deviations with smaller but positive effects on grade-eight math and grade-four reading performance. They reported negative but statistically insignificant effects on grade-eight reading but also noted that the nonlinear trends in this measure prior to NCLB may make this measure unsuitable for their quasi-experimental research design. Wong, Cook, and Steiner reported similar results using different design features.[43] Comparisons in the changes in low-stakes test performance across public and Catholic schools similarly suggest a positive impact of NCLB, though the contemporaneous change in Catholic school enrollment due to sexual-abuse scandals qualifies these findings.[44]

The National Research Council report on test-based accountability averaged over these heterogeneous estimates to conclude that NCLB increased test scores overall by 0.08 standard deviations.[45] That report also noted accurately that gains of this magnitude are "small compared to the improvements the nation hopes to achieve." However, Dee, Jacob, and Schwartz noted that these effects are not necessarily small from a cost-benefit perspective.[46] The present discounted value of the earnings gains implied by the math-score gains due to NCLB is large relative to the corresponding increases in expenditures in states where NCLB created new accountability systems (i.e., roughly $600 per pupil). Recent evidence on the potential long-run effects of NCLB is consistent with this inference.

Specifically, Harris, Liu, Barrett, and Li presented quasi-experimental evidence that NCLB contributed to the recent rise in high school graduation rates (and a decline in general equivalency degrees) and that these gains could not be easily explained by credit-recovery activities or the strategic manipulation of reporting.[47]

In sum, studies of pre-NCLB and NCLB-based accountability systems indicate that they generated meaningful, though not transformational, improvements in school performance. The available evidence on the impact of the targeted school reforms that took place subsequently under the aegis of NCLB waivers and stimulus-funded initiatives is more mixed. For example, a national study of the targeted school-turnaround reforms funded by federal School Improvement Grants (SIGs) suggested that these expensive reforms were largely ineffective.[48] However, this study appeared to have been poorly powered to detect meaningful effects.[49] Also, quasi-experimental studies that focused on the experiences of specific states found evidence that SIG-funded school reforms often but not always had positive effects.[50]

The Priority School reforms that occurred under NCLB waivers closely paralleled SIG-funded reforms in that they required schools to implement federal turnaround principles (e.g., staff and leadership change, increased learning time, embedded professional development, social and emotional supports) but lacked the additional financial support of SIGs. The available evidence on Priority Schools is more limited. Hemelt and Jacob examined Priority School reforms in Michigan and found that they had no impact on school staffing or performance.[51] Their evidence suggested that these null findings reflect weak implementation at the school and district levels coupled with limited capacity for state oversight. Similarly, Zimmer, Henry, and Kho reported weak effects of Priority School reforms in Tennessee except among schools placed under the management of a district "innovation zone" that had additional autonomy and resources.[52]

The policy initiatives that have the most relevance for predicting the impact of targeted school reforms under ESSA are the Focus School reforms introduced under NCLB waivers. Like the targeted reforms currently being introduced under ESSA (i.e., CSI, TSI, and ATSI), waiver-based Focus Schools gave states considerable flexibility in how they chose to improve these schools, asking only that their approach have evidentiary support. And, like the targeting of ESSA's TSI and ATSI reforms, states identified Focus Schools based on their contributions to achievement gaps. Most of the evidence on the impact of Focus School reforms is dispiriting. Quasi-experimental studies that examine Focus Schools in Rhode Island,[53] Louisiana,[54] and Michigan[55] all reported no evidence of school improvement coupled with evidence of weak implementation.

However, the exceptions to this pattern of null findings are particularly noteworthy. For example, in a study of the Focus School reforms introduced in Kentucky, Bonilla and Dee presented regression-discontinuity evidence indicating that these reforms

significantly increased the test performance of the targeted "gap group" students.[56] They also discussed evidence of the state and district activities (e.g., a distinctive and comprehensive school-planning process, high-quality teacher professional development) that mediated this impact. Zimmer, Henry, and Kho examined the school-turnaround initiatives implemented under the aegis of Tennessee's Race to the Top initiative.[57] They found that state takeovers of chronically underperforming schools were not effective. However, they also documented meaningful improvements in schools that were instead placed within semiautonomous units within their districts (i.e., "innovation zones"). They suggested that incentive pay and the selective retention and recruitment of high-performing teachers mediated these positive changes. Schueler, Goodman, and Deming also documented evidence of early improvements in a Massachusetts' school district that was taken over by the state and that implemented a variety of reforms (e.g., higher expectations, increased learning time, data-driven instruction, and changes in school leadership and staffing).[58] Overall, these results suggest that the impact of federal school-reform initiatives depends critically on the extent to which they catalyze meaningful and thoughtful implementation at the state, district, and school levels. The success of the few reform efforts that clearly engaged state, district, and school actors is consistent with the hypothesis that such purposeful and supportive contexts are necessary moderators of successful school-level reforms.

Assessing ESSA and Looking Ahead

The historical sweep of accountability reforms and research raises multiple questions about how schools can improve and advance productively in near future. In that context, the broad arc of federal and state policy changes that connect the current ESSA regime to prior accountability reforms merits special attention. In surveying the current policy environment, some have recently argued that "the education reform movement is dead."[59] However, to borrow from another context, reports of this death are greatly exaggerated. Instead, with respect to accountability, a more accurate assessment is that public schools now inhabit a diverse and complicated reform environment that features a devolution of authority to state and local actors[60] but that also has strong ties to the recent past. In particular, while accountability under ESSA has distinctive features, its flexibility has allowed states to produce diverse accountability systems that draw on different elements unique to each of the preceding policy regimes (i.e., pre-NCLB, NCLB, and waivers).

One prominent concern is how some state accountability plans under ESSA seek to support the educational potential of subgroups (e.g., Blacks, Hispanics, low-income, English learners, special education) that have been historically underserved by public schools. The emphatic attention that the NCLB era brought to the educational outcomes of these student subgroups was one of its most universally admired features. A coalition of civil rights groups recently asked to review their federally approved ESSA plans because of concerns that they "do not hold schools sufficiently accountable for their responsibility to all children, especially groups of children who have been shortchanged for too long."[61]

Specifically, critics have argued that "most states are shirking their responsibilities around two of the law's most important provisions for historically underserved groups of students."[62] Twelve states do not report student subgroups in their rating systems while thirty-eight other states appear out of compliance because they do not use data on all subgroups for rating all schools or because their reporting is at risk of obscuring subgroup performance. Similarly, only six states provide a "strong, distinct" definition of subgroup performance that merits school identification for Targeted Support and Improvement (TSI) under ESSA.[63] Because ESSA is likely to remain the federal framework for state accountability systems for the foreseeable future, engaging these issues at the state level will be an important domain of research and policy improvement.

In contrast, the most unique and widely appreciated feature of ESSA-based accountability is, arguably, the movement toward diverse measures of school performance that go beyond simple proficiency thresholds in math and ELA (e.g., growth measures, chronic absenteeism, high school graduation, college and career readiness). However, ESSA has also continued NCLB's traditional emphasis on universal testing and on reporting performance data for important subgroups. These richer data can be important in providing school leaders the information they need to understand the needs of the students they serve. These diverse data may also result in accountability incentives that attenuate, if not eliminate, the potential for unintended consequences associated with Campbell's Law.

However, to the extent that the benefits of school accountability are also to be found in reducing "moral hazard" in the behavior of school leadership and staff, these newly diverse accountability indicators will only be as effective as the consequences to which they are linked. On this point, there appear to be highly relevant differences across states. Some have developed school ratings that appear to carry consequential labeling not unlike those previously associated with AYP and some pre-NCLB state reforms (e.g., A–F and star ratings). However, other states have no summative ratings of schools or ratings that appear to lack consequentiality (e.g., tiers of support). Instead, these accountability systems resemble other state-level pre-NCLB accountability reforms in that they rely simply on publicizing school performance data (i.e., "report card" accountability) rather than on articulating a summative interpretation or clear consequences. The prior evidence that such information-only reforms weren't clearly effective[64] raises substantial doubt about the likely benefits of ESSA accountability in such states.

Similarly, the research on targeted school reforms sponsored by stimulus spending and conducted under NCLB waivers gives us relatively little reason to be optimistic about the likely success of the parallel efforts being undertaken under ESSA. More specifically, if this limited evidence is a reliable guide, we may expect ESSA's CSI, TSI, and ATSI reforms to improve student outcomes only if they happen to spark thoughtful and energetic initiatives at the state and local levels. This may well happen in a few contexts. However, we do not appear to have a particularly clear understanding of what factors contribute to whether

state and local actors respond to federal initiatives with genuine reform that is specific, authoritative, and stable rather than with weak implementation or cosmetic regulatory compliance. Furthermore, humility about the capacity of the federal government to drive targeted school improvements appears to be a policy lesson that is frequently taught but not well learned. For example, between 1998 and 2005, the US Department of Education disbursed $1.8 billion to low-performing schools participating in a Comprehensive School Reform (CSR) program. The available evidence suggests that CSR was not particularly successful in raising student achievement and that implementation of the federal CSR model was uneven in important ways.[65]

In other words, the available evidence from prior accountability systems suggests that the federal school-reform initiatives under ESSA have some capacity to drive increases in the performance of targeted schools. However, this power can be either enhanced or, more likely, attenuated by the federalist structure that connects national policy to school and classroom practices. In theory, the evidence standards defined under ESSA could compel states and districts to select uniquely promising initiatives to improve their targeted schools (e.g., studies singled out in the federal What Works Clearinghouse). However, the astonishing breadth of ESSA's "tiers of evidence" suggests that this is far from guaranteed. Under ESSA, states have the latitude to choose reform strategies that may find only evidentiary support from correlational studies (i.e., Tier 3) or even from a mere rationale (i.e., Tier 4). Furthermore, the enforcement of these capacious standards is, in the current policy environment, at best an open question. However, some accounts suggest that states are pursuing strategies that privilege the ambitious use of strategies supported by high-quality evidence.[66] And early characterizations of school improvement under ESSA suggest that the evidence standards are having a positive impact by allowing schools and districts to make "thoughtful decisions about school improvement strategies."[67] Our experiences with federally sponsored and targeted school improvement efforts (i.e., Comprehensive School Reform, School Improvement Grants, Priority and Focus School reforms) suggest this early optimism should be viewed provisionally. However, emerging evidence on the early implementation of school reform under the aegis of ESSA will soon provide more definitive evidence on this important question.

NOTES

1 James S. Coleman, *Equality of Educational Opportunity* (Washington, DC: National Center for Educational Statistics, 1966).

2 Coleman, *Equality of Educational Opportunity,* 325.

3 Chester E. Finn, *We Must Take Charge!: Our Schools and Our Future* (New York: Free Press, 1991), 147.

4 Donald T. Campbell, "Assessing the Impact of Planned Social Change," *Evaluation and Program Planning* 2, no. 1 (1979): 85.

5 Campbell, "Assessing the Impact," 86.

6 Thomas S. Dee, "The 'First Wave' of Accountability," in *No Child Left Behind? The Politics and Practice of Accountability,* ed. Paul Petersen and Martin West (Washington, DC: Brookings Institution Press, 2003).

7 Education Commission of the States, Response to Information Request, July 26, 2016, https://www.ecs.org/wp-content/uploads/Exit-Exam-Requirements-for-Class-of-2017_07.26.16.pdf.

8 National Commission on Excellence in Education, *A Nation at Risk: The Imperative for Educational Reform* (Washington, DC: US Department of Education, 1983).

9 Dee, "The 'First Wave' of Accountability."

10 Maris A. Vinovskis, *The Road to Charlottesville: The 1989 Education Summit* (Washington, DC: National Education Goals Panel, September 1999).

11 Education Week, 2001. "Quality Counts Annual Report, 2001," *Education Week* 20, No. 17 (January 11, 2001).

12 Lance D. Fusarelli, "Restricted Choices, Limited Options: Implementing Choice and Supplemental Educational Services in No Child Left Behind," *Educational Policy* 21, no. 1 (2007): 132–54, https://doi.org/10.1177/0895904806297579.

13 Rebecca W. DiBiase, *State Involvement in School Restructuring under No Child Left Behind in the 2004–2005 School Year,* ECS Policy Brief (Denver: Education Commission of the States, 2005); Caitlin Scott, *A Call to Restructure Restructuring: Lessons from the No Child Left Behind Act in Five States* (Washington, DC: Center on Education Policy, 2008); Frederick M. Hess and Chester E. Finn, eds., *No Remedy Left Behind: Lessons from a Half-Decade of NCLB* (Washington, DC: American Enterprise Institute Press, 2007).

14 Elizabeth Mann Levesque, "The Long-Term Impact of NCLB Waivers on ESEA Renewal," Brown Center Chalkboard, Brookings Institution, 2015, https://www.brookings.edu/blog/brown-center-chalkboard/2015/12/10the-long-term-impact-of-nclb-waivers-on-esea-renewal, accessed May 4, 2020.

15 David J. Hoff, "2007 NCLB Prospects Are Fading," *Education Week* 27, no. 11 (2007): 1–22.

16 Levesque, "Long-Term Impact of NCLB Waivers."

17 Wayne Riddle, *Major Accountability Themes of Second-Round State Applications for NCLB Waivers* (Washington, DC: Center on Education Policy, 2012).

18 Wayne Riddle, *What Impact Will NCLB Waivers Have on the Consistency, Complexity and Transparency of State Accountability Systems?* (Washington, DC: Center on Education Policy, 2012).

19 Riddle, *What Impact Will NCLB Waivers Have.*

20 Andrew Ujifusa, "Waiver Plans Push School Grading Systems," *Education Week* 31, no. 30 (2012): 21–23.

21 Diane S. Rentner, Maria Ferguson, and Nancy Kober, *A Stronger Future of Evidence-Based School Improvement in ESSA* (Washington, DC: Center on Education Policy, December 2019).

22 Alyson Klein, "The Every Student Succeeds Act: An ESSA Overview," *Education Week* (March 15, 2016); "ESSA Plans: Takeaways from the First Batch of Approvals," *Education Week* (September 20, 2017).

23 Evie Blad, "States Skip Out on Social-Emotional Measures for ESSA," *Education Week* 37, no. 8 (October 11, 2017): 1, 9.

24 Daarel Burnette, "A Slow Build for Reporting on ESSA Data," *Education Week* 38, no. 32 (May 8, 2019): 1, 16.

25 Education Commission of the States, "50 State Comparison: States' Accountability," 2018, https://www.ecs.org/50-state-comparison-states-school-accountability-systems.

26 Los Angeles Times Editorial Board, "California's New School Rating Tool Is Better, but Still Flawed," *Los Angeles Times*, December 15, 2018, https://www.latimes.com/opinion/editorials/la-ed-california-school-dashboard-20181215-story.html.

27 Rentner, Ferguson, and Kober, *Evidence-Based School Improvement in ESSA.*

28 Center on Education Policy, *Number of Low-Performing Schools by State in Three Categories (CSI, TSI, and ATSI), School Year 2018–19* (Washington, DC: Center on Education Policy, 2019).

29 Council of Chief State School Officers, *ESSA Implementation Timeline: A Guide to Key State and Local Processes* (Washington, DC: Council of Chief State School Officers, 2018).

30 National Research Council, *Incentives and Test-Based Accountability in Education,* ed. Michael Hout and Stewart W. Elliott (Washington, DC: National Academies Press, 2011).

31 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.

32 Bradley M. Allan and Roland G. Fryer, *The Power and Pitfalls of Education Incentives,* Hamilton Project (Washington, DC: Brookings Institution, 2011); Thomas S. Dee and Brian Jacob, "The Impact of No Child Left Behind on Student Achievement," *Journal of Policy Analysis and Management* 30, no. 3 (2011): 418–46.

33 Steven W. Hemelt and Dave E. Marcotte, "High School Exit Exams and Dropout in an Era of Increased Accountability," *Journal of Policy Analysis and Management* 32, no. 2 (2013): 323–49; Thomas S. Dee and Brian A. Jacob, "Do High School Exit Exams Influence Educational Attainment or Labor Market Performance?," in *Standards-Based Reform and the Poverty Gap: Lessons for No Child Left Behind,* ed. Adam Gamoran (Washington, DC: Brookings Institution, 2007), 154; Dee, "The 'First Wave' of Accountability."

34 Stephen Klein, Laura Hamilton, Daniel McCaffrey, and Brian Stecher, "What Do Test Scores in Texas Tell Us?," *Educational Policy Analysis Archives* 8, no. 49, http://epaa.asu.edu/ojs/article/view/440/563.

35 Dale Ballou and Matthew G. Springer, "Has NCLB Encouraged Educational Triage? Accountability and the Distribution of Achievement Games," *Education Finance and Policy* 12, no. 1 (2017): 77–106.

36 David Figlio and Susanna Loeb, "School Accountability," in *Handbook of the Economics of Education,* vol. 3, ed. Eric A. Hanushek, Stephen Machin, and Ludger Woessmann (Amsterdam: North Holland,), 383–417.

37 National Research Council, *Incentives and Test-Based Accountability in Education;* Figlio and Loeb, "School Accountability."

38 Eric A. Hanushek and Margaret Raymond, "Does School Accountability Lead to Improved School Performance?," *Journal of Policy Analysis and Management* 24, no. 2 (2005): 297–329.

39 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 (2014): 207–41.

40 Hanushek and Raymond, "Does School Accountability Lead to Improved School Performance?"

41 National Research Council, *Incentives and Test-Based Accountability in Education,* 55.

42 Dee and Jacob, "The Impact of No Child Left Behind on Student Achievement."

43 Manyee Wong, Thomas D. Cook, and Peter M. Steiner, "Adding Design Elements to Improve Time Series Designs: No Child Left Behind as an Example of Causal Pattern-Matching," *Journal of Research on Educational Effectiveness* 8, no. 2 (2015): 245–79.

44 Dee and Jacob, "The Impact of No Child Left Behind on Student on Achievement."

45 National Research Council, *Incentives and Test-Based Accountability in Education,* 44.

46 Thomas S. Dee, Brian Jacob, and Nathaniel L. Schwartz, "The Effects of NCLB on School Resources and Practices," *Educational Evaluation and Policy Analysis* 35, no. 2 (2013): 252–79.

47 Douglas N. Harris, Lihan Liu, Nathan Barrett, and Ruoxi Li, "Is the Rise in High School Graduation Rates Real? High-Stakes Accountability and Strategic Behavior," EdWorking Paper No. 20-210, Annenberg Institute at Brown University, March 2020.

48 Lisa Dragoset, Jaime Thomas, Mariesa Herrmann, John Deke, Susanne James-Burdumy, Cheryl Graczewski, Andrea Boyle, Rachel Upton, Courtney Tanenbaum, and Jessica Giffin, *School Improvement Grants: Implementation and Effectiveness: Executive Summary* (NCEE 2017-4012) (Washington, DC: US Department of Education, Institute of Education Sciences, 2017).

49 Alan Ginsburg and Marshall S. Smith, *Revisiting SIG* (Washington, DC: FutureEd, Georgetown University, 2018).

50 Thomas S. Dee, *School Turnarounds: Evidence from the 2009 Stimulus,* NBER Working Paper No. 17990 (Cambridge, MA: National Bureau of Economic Research, 2012); Daniel Player and Veronica Katz, "Assessing School Turnaround: Evidence from Ohio," *Elementary School Journal* 116, no. 4 (2016): 675–98; Min Sun, Emily K. Penner, and Susanna Loeb, "Resource- and Approach-Driven Multidimensional Change: Three-Year Effects of School Improvement Grants," *American Educational Research Journal* 54, no. 4 (2017): 607–43; Jennifer A. Heissel and Helen F. Ladd, "School Turnaround in North Carolina: A Regression Discontinuity Analysis," *Economics of Education Review* 62 (2018): 302–20; Deven Carlson and Stéphane Lavertu, "School Improvement Grants in Ohio: Effects on Student Achievement and School Administration," *Educational Evaluation and Policy Analysis* 40, no. 3 (2018): 287–315.

51 Steven W. Hemelt and Brian A. Jacob, "How Does an Accountability Program That Targets Achievement Gaps Affect Student Performance?," *Education Finance and Policy* 15, no. 1 (2020): 45–74.

52 Ron Zimmer, Gary T. Henry, and Adam Kho, "The Effects of School Turnaround in Tennessee's Achievement School District and Innovation Zones," *Educational Evaluation and Policy Analysis* 39, no. 4 (2017): 670–96.

53 Shaun M. Dougherty and Jennie M. Weiner, "The Rhode to Turnaround: The Impact of Waivers to No Child Left Behind on School Performance," *Educational Policy* 33, no. 4 (2017): 555–86.

54 Thomas S. Dee and Elise Dizon-Ross, "School Performance, Accountability and Waiver Reforms: Evidence from Louisiana," *Educational Evaluation and Policy Analysis* 41, no. 3 (2019): 316–49.

55 Hemelt and Jacob, "How Does an Accountability Program That Targets Achievement Gaps Affect Student Performance?"

56 Sade Bonilla and Thomas S. Dee, "The Effects of School Reform under NCLB Waivers: Evidence from Focus Schools in Kentucky," *Education Finance and Policy* 15, no. 1 (2020): 75–103.

57 Zimmer, Henry, and Kho, "The Effects of School Turnaround."

58 Beth E. Schueler, Joshua S. Goodman, and David J. Deming, "Can States Take Over and Turn Around School Districts?: Evidence from Lawrence, Massachusetts," *Educational Evaluation and Policy Analysis* 39, no. 2 (2017): 311–32.

59 Annie Murphy Paul, "Diane Ravitch Declares the Education Reform Movement Is Dead," *New York Times,* January 21, 2020, https://nytimes.com/2020/01/21/books/review/slaying-goliath-diane-ravitch.html.

60 Patrick McGuinn, "From No Child Left Behind to the Every Student Succeeds Act: Federalism and the Education Legacy of the Obama Administration," *Publius: The Journal of Federalism* 46, no. 3 (Summer 2016): 392–415.

61 Alyson Klein, "States Are Failing to 'Put Students' Civil Rights First' in ESSA Plans, Advocates Say," http://blogs.edweek.org/edweek/campaign-k-12/2018/06/states_essa_civil_rights.html.

62 Alliance for Excellent Education, 2018. "Too Many States Minimize Student Subgroup Performance in ESSA Accountability Systems," December 14, 2018, https://all4ed.org/reports-factsheets/too-many-states-minimize-student-subgroup-performance-in-essa-accountability-systems.

63 Alliance for Excellent Education, 2018. "Too Many States Minimize Student Subgroup Performance in ESSA Accountability Systems."

64 Hanushek and Raymond, "Does School Accountability Lead to Improved School Performance?"

65 Laura Desimone, "How Can Comprehensive School Reform Models Be Successfully Implemented?," *Review of Educational Research* 72, no. 3 (2002): 433–79; Bethany Gross, T. Kevin Booker, and Dan Goldhaber, "Boosting Student Achievement: The Effect of Comprehensive School Reform on Student Achievement," *Educational Evaluation and Policy Analysis* 31, no. 2 (2009): 111–26; US Department of Education, *Evaluation of the Comprehensive School Reform Program Implementation and Outcomes: Fifth-Year Report* (Washington, DC: Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service, 2010).

66 Anne Hyslop, "How the Every Student Succeeds Act Empowers States to Find Innovative Uses for Federal Funds," *The 74,* April 17, 2017, https://www.the74million.org/article/hyslop-how-the-every-student-succeeds-act-empowers-states-to-find-innovative-uses-for-federal-funds.

67 Rentner, Ferguson, and Kober, *Evidence-Based School Improvement in ESSA.*

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

About the Author

THOMAS S. DEE

Thomas S. Dee, PhD, is the Barnett Family Professor at Stanford University's Graduate School of Education (GSE), a research associate at the National Bureau of Economic Research (NBER), a senior fellow at the Stanford Institute for Economic Policy Research (SIEPR) and the faculty director of the John W. Gardner Center for Youth and Their Communities.

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