A BACKGROUND PAPER FOR THE HOOVER EDUCATION SUCCESS INITIATIVE

The Future of School Accountability

DAVID STEINER AND ALANNA BJORKLUND-YOUNG

> Try to think of an education policy that 1) has been shown, in dozens of studies across multiple decades, to positively affect student outcomes, 2) has the overwhelming support of parents and voters, 3) reinforces many other policies and facilitates quality research, and 4) has been used widely at the district, state, and national levels for decades or more. You might be thinking that such a policy doesn't exist, and if it did, we'd surely want to keep it around. But the truth is precisely the opposite. Such a policy does exist—it's called school accountability—yet the powers that be seem increasingly ready to throw it out and leave education to the whims of the all-but-unregulated free market.
>
> —Morgan S. Polikoff, "Is Test-Based Accountability Dead?"[1]

> The preponderance of evidence suggests positive effects of the accountability movement in the United States during the 1990s and early 2000s on student achievement, especially in math.
>
> —Susanna Loeb and David Figlio, "School Accountability"[2]

> With test-based accountability, distant officials have imposed their preferences on the rest of us. In addition, studies such as the ongoing research of David Grissmer and colleagues indicate that long-term achievement in math and reading depends on a broader education that includes the type of general knowledge conveyed by history, science, art, and music.
>
> —Jay P. Greene, "Futile Accountability Systems Should Be Abandoned"[3]

What Is School Accountability?

From 30,000 feet, the definition of school accountability in American K–12 education is relatively clear: it is "the process of evaluating school performance on the basis of student performance measures."[4] But descend a bit and agreement comes to an end.

There is disagreement over the *purpose* of evaluating school performance. For some, the value is primarily in collecting good measures of student learning to track learning outcomes. For others, the goal is to collect measures of student learning in order to act effectively on that information. Within this second group, there is wide variation of opinion: should the focus be on the very lowest performing schools, or on all schools where most students are not at proficiency levels, or on the performance of all students, including those who are especially gifted? Moreover, there is debate about what measures of student learning should be used—and whether they should include broader indicators of student well-being.

The purpose of accountability also depends on the *level,* or perspective, from which the accountability system is viewed. From a federal standpoint, a clear picture of state-level performance—including the learning outcomes for minority and underprivileged students—is crucial. Districts focus on school-level accountability and on developing systems that hold principals and teachers accountable for student learning. Parents are interested in information about their child's classroom, as well as information about school culture that goes well beyond those of traditional accountability systems. At each of these levels, the goals, measures of interest, and any associated actions (that is, consequences or rewards) are different.

There are also debates over which kinds of *systems* best monitor and incentivize behaviors that lead to these end goals. Scholarly reviews have identified at least four common accountability systems:[5]

- **Outcome-based accountability** in which the mechanism of accountability is students' skills and knowledge at each stage of their education (e.g., through measures of student learning)

- **Rules-based accountability** in which the mechanism of accountability is a set of rules that define actions that should or should not be taken

- **Market-based accountability** in which the marketplace, typically through consumers, provides the accountability mechanism

- **Professional accountability** in which professional standards, such as entrance exams and certification requirements, provide the accountability mechanism

Some researchers have argued that the education system should embrace several of these systems of accountability.[6] In practice, however, most have heavily focused on the outcome-based accountability that standardized tests provide.

While we remain mindful of these different perspectives on what school accountability is for and how it should function, our focus here is on a federal model that incentivizes better academic outcomes for all K–12 students while particularly accelerating the learning of those furthest behind.

Our starting point is the current federal accountability system as required by ESSA (Every Student Succeeds Act).

ESSA has three main innovations from its predecessor, No Child Left Behind (NCLB). First, ESSA was designed to give flexibility to the states by allowing states to define their own goals for student learning—as long as, in the long term, subgroups that have been struggling make the most progress. Second, ESSA continues to obligate states, in concert with the relevant school districts, to develop plans for schools that are weak overall, or

which have specific subgroups of students who are struggling, although the content of those plans is now left largely to state discretion. Finally, ESSA broadens the allowable categories of metrics used for accountability, including school quality factors such as college-readiness.

For a quick refresher on ESSA vs. NCLB provisions, see "The Difference Between the Every Student Succeeds Act and No Child Left Behind."[7] For a fuller account of the ESSA law provisions, see "Every Student Succeeds Act (ESSA) Overview."[8] The Center for American Progress has a useful summary of the basic ESSA provisions including the following concise description:[9]

> States must annually and publicly report on how well all of their public school students are performing on the following measures, as well as set goals for the first, second, and fourth indicators:
>
> 1. Academic achievement in reading and math for third grade through eighth grade and once in high school [and in science, once in elementary, middle, and high school]
> 2. High school graduation rate
> 3. Growth or another academic indicator for elementary and middle schools
> 4. English language proficiency for English learners only
> 5. At least one measure of school quality or student success
>
> Additionally, states must collect and report on a new, more varied set of data than the five indicators listed above. These new data provide insights into levels of student engagement and the availability of resources that support broader student learning. These data include access to advanced coursework, exclusionary discipline rates, chronic absenteeism, professional qualifications of educators, per-pupil expenditures, and postsecondary enrollment rates.

Optimal Accountability Model under ESSA Rules

ESSA permits variability between state accountability systems. What components comprise the strongest possible model? (We turn later to other ideal elements that would require a change in federal law.)

Transparent information. School-, district-, and state-level accountability measures should be made public and intuitively accessible online so the most important information is easy to grasp. We find that clearly explained, simple summative ratings (e.g., A–F grades) help ensure that all stakeholders, especially parents, understand both how the school is performing and when urgent improvement is needed.[10]

In addition to easily understood summative ratings, systems should also provide more detailed information to interested readers. This includes explanations of all calculations, disaggregated information, and measures of local interest not included in the formal

accountability system, such as a description of each school's curriculum and a description of students' exposure to each academic and nonacademic subject.

National comparisons. Whenever possible, school scores should be placed in context—ideally with other states, or at least in comparison to other schools and districts within the state. In addition, each state should clearly indicate on the top-facing page of its accountability system the percentage of students overall, and by subgroup, who achieved proficiency on the most recent National Assessment of Educational Progress (NAEP). This also means that we believe that *NAEP should release state-level data in all key subjects at all three NAEP grade levels*.[11]

Clear, honest measures. Measures included in accountability systems send clear and strong incentives that can improve students' outcomes. In an ideal world, educators would know that stronger performance on a data metric would *cause* better student outcomes. The accountability measures must also be valid and reliable, differentiate between schools, and be primarily within the schools' control (i.e., not primarily within the purview of families or other out-of-school factors).

Indicators that are hard to manipulate or game. Previous accountability systems have been weakened by gameable elements, including discretionary proficiency standards that left students in many states "meeting" proficiency levels but being far from on track for college or career opportunities. Currently, graduation standards vary tremendously from state to state and include safety valves in the form of "alternative routes" to graduation—which often lack external standards.[12] Accountability systems should minimize this kind of behavior through careful selection of measures that are hard to game, a requirement that states clearly define their measures, and meticulous tracking and oversight.

For example, in the case of graduation rates, since ESSA requires states to report them, we believe that each state needs to indicate very clearly what its standards are and what percentage of its students "graduate" through alternative routes. All alternative routes should also be clearly defined. In all cases, the state should indicate the failure rates (i.e., those getting F grades in graduation-required courses and those failing to achieve graduation-required assessment scores) and particularly the failure rates among students relying on alternative routes.

In the case of grade point average (GPA), schools and districts should indicate five years of average GPA juxtaposed with their high school assessment scores so as to enable stakeholders to see if GPA is being inflated. (This could be the case, for example, if GPAs rise over that period but state scores are flat.)

We recommend that states report the calibration between the required cut scores for proficiency on high school state tests with public university GPAs in that state. For example, in Maryland, the state board recently was able to equate cut scores on the state's

English language arts (ELA) and math assessments with the GPAs students with those scores achieved in college.

Transparent treatment of subgroups. A high-quality accountability system clearly and transparently communicates the performance of all subgroups, especially those that have been traditionally underserved. States should avoid the use of "super-groups" in which such subgroups are aggregated into a single measure; doing so hides important differentiating information. States should go beyond ESSA's requirements and should include in their accountability measures the performance of gifted students (as will be the case in Maryland).[13]

Reporting proficiency and growth. *Students' performance—both in terms of absolute achievement and in terms of growth—should be evaluated by criterion-based assessments.* Including growth ensures that there is an incentive for schools and teachers to place extra focus on the students with the greatest needs without ignoring those students who are performing at grade level or above. *The question is what the relative weighting should be between achievement and growth.* Some policy analysts argue that "because [achievement] measures are strongly correlated with student demographics and prior achievement, we believe they should count for at *most* a quarter of schools' ratings."[14] We reject that view: while it is true that schools serving disadvantaged children will have more work to do to raise absolute performance levels, a well-designed achievement metric that gives increasing weight to progressively higher achievement (from below basic to above basic, for example) will reward schools whose students move upward through performance levels. More important still, parents' and students' greatest interests are not captured by normative measures: they should expect clear signals as to the absolute performance of students in a school. The workplace rarely rewards any candidate for some relative growth. Therefore, we believe that equal weight should be assigned to achievement and growth measures.

Criterion-referenced growth measures. States have rightly added measures of students' growth in learning to their accountability systems—but there is some question about which metric of growth to use. There are two primary debates: 1) whether the growth measure should be norm-referenced or criterion-referenced and 2) the proper way in which to calculate the growth measure. We will focus only on the first, more fundamental question—it must also be determined before the second.[15]

Criterion-based growth measures compare students' progress toward a predetermined level of performance (e.g., proficiency). These systems might, for example, report the percentage of students who moved from one level of performance to the next.

In contrast, norm-based systems compare students' progress to other students who have similar characteristics (e.g., race, socioeconomic status, attendance at similar schools). This definition of growth highlights relative growth but is silent on how students compare to grade-level expectations.

A criterion-based growth calculates the percentage of students who move from one level to the next (e.g., below basic to basic, basic to proficient, and proficient to advanced). Such gains could be reported for all students and by student subgroup. Such a measure incentivizes teachers to focus on many students, and therefore weakens gaming that appeared when only one "cut-point" was prioritized under NCLB (i.e., meeting proficiency). While the cut scores that states establish to indicate "basic," "proficient" and "advanced" may not all be (yet) strongly correlated with external measures such as college or career readiness, they do allow intrastate comparisons with all other public schools. On the other hand, only certain growth is measured. For example, a two-point increase in a test score may matter only if it elevates the student onto the next level; if not, the model ignores the same growth. In addition, criterion-referenced measures are criticized because they do not sufficiently disentangle the relationship between student performance and poverty. For example, Cory Koedel, an economist from the University of Missouri, argues, "As an accountability metric, growth-to-proficiency is a terrible idea for the same reason that achievement-level metrics are a bad idea—it is just about poverty."[16]

These arguments highlight the benefit of norm-referenced measures: such measures compare students who are similarly advantaged or disadvantaged, and therefore break the relationship between measures of growth and student characteristics. In this way, these growth measures provide context for students' growth. However, if all schools produce very little growth, the least dreadful of the such underperforming would look strong and be rewarded in the accountability system, thus giving a falsely rosy picture to parents and school personnel alike.

On balance, we favor criterion-based measures and recommend that states use a criterion-based growth measure. *Whichever method is used—criterion or normative based—the state report card should make clear to all readers what exactly the growth measure is indicating, and what it isn't.*

Attendance. States should *include student attendance rates in the accountability measures.* Attendance and chronic absenteeism are the most common nonacademic indicators currently used by states. The research basis for such a measure is compelling. Quasi-experimental research suggests that less instructional time causes lower student achievement.[17] However, absences are more detrimental than simply adding additional days to the school calendar.[18] Indeed, students' absences also cause decreased student achievement.[19] Research also suggests that student absences are particularly detrimental to low-income students and English-language learners, especially in reading. Further, chronic absenteeism is detrimental not only to the absent students, but to their classmates as well.[20]

In addition, absenteeism—and chronic absenteeism in particular (often defined as missing more than 10 percent of school days)—is a significant problem, varying with socioeconomic status and age. Research suggests that poorer students are significantly more likely to be chronically absent.[21] According to the US Department of Education's 2013–14 Civil Rights Data Collection, which includes data from every public school in the country, more than

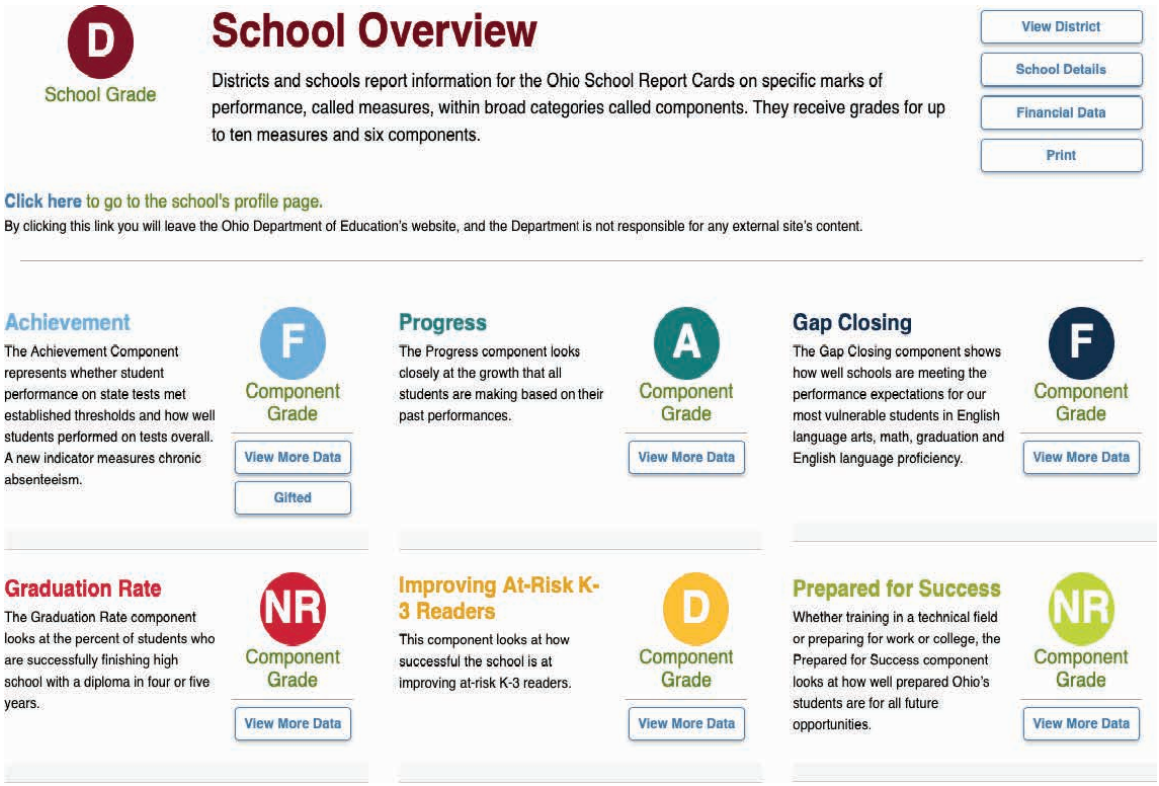
six million students missed fifteen or more days of school in 2013–14. This amounts to 14 percent of the student population.

Given both the negative influence of absenteeism on students and the prevalence of the problem, an accountability system that includes measures of student attendance makes sense, especially chronic absenteeism.

A Case Study: Ohio

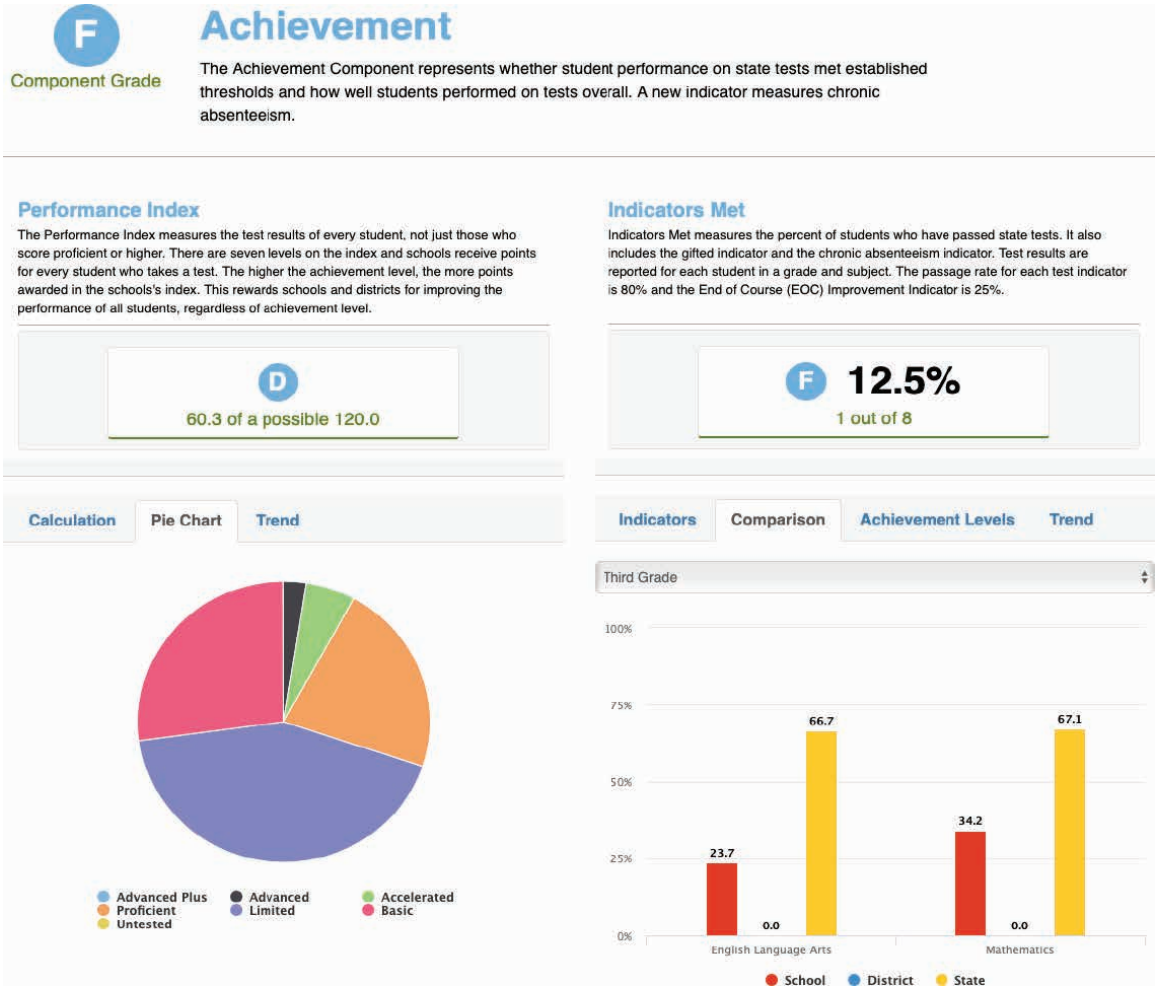
With the preceding recommendations in mind, it is worth taking a strong example of a state ESSA plan and suggesting how it could be stronger still. The example here is Ohio. The presentation, as indicated below, is visually very clear; levels of further analysis are available and do not immediately become opaque. The inclusion of gap closing and the calling out of "at-risk K-3 readers" are strong elements. The state uses a performance index or scale scores in place of proficiency rates when measuring achievement. We would like that index to be tethered to multiple levels of performance; we would prefer the growth metric to be criterion based; and we would like to see the NAEP proficiency rates posted along with clear indications of any alternative graduation pathways and who used them. But as a whole, as seen in figures 1 and 2, the system is a strong example of what can be done under the ESSA provisions.

Figure 1. Snapshot of a school's top-level report card under the Ohio Accountability System



Source: 2018–2019 Report Card Resources." Ohio Department of Education (2020). http://education.ohio.gov/Topics/Data/Report-Card-Resources.

Figure 2. Snapshot of unpacking the Achievement Level indicator in the same school



Source: 2018–2019 Report Card Resources." Ohio Department of Education (2020). http://education.ohio.gov/Topics/Data/Report-Card-Resources.

Optional Measures under ESSA Rules

An effective accountability system of today should also explore, with caution, innovation measures that may be included fully in the future.

Recent national research demonstrates that 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 their independence, success, and happiness.[22] Parents also welcome information on these broader school missions, including social-emotional learning, career readiness, and long-term student attainment.[23]

The reliability of data on such nontraditional metrics varies. For instance, it is possible in some circumstances to follow students' progress in the years after high school graduation.

Schools might then be able to report on the percentage of graduates who attend college; persist into their second year of college; or are employed in career positions with tracks to a family-sustainable income.[24] In such circumstances, it makes sense for states to include measures that are within a couple of years of matriculation within a formal accountability system.

Additional measures are worth exploring for reporting purposes, despite not being ready for inclusion in formal accountability systems. Public reporting of such metrics would be helpful to families, schools, and system leaders, as long as it is made very clear that they do not meet the same standards of validity and reliability. To be included in formal accountability systems, the following requirements must be met: *1) the measure(s) included must be clearly defined, measurable, and related to other well-established student outcomes; and 2) the measures must be valid (and reliable) for their intended use.*

Some measures are closer to meeting these requirements than others. For example, "growth mindset" is clearly defined, and the research on its importance and relationship with other well-understood student outcomes is growing.[25] However, we are not yet aware of any measures that have moved beyond self-reported student surveys, which are not currently valid for use in a high-stakes environment.[26]

The following elements, then, lend themselves to inclusion on school-level report cards but not in formal accountability systems.

Social and emotional well-being

There is common agreement that social intelligence and emotional intelligence are made up of multiple skills, which are distinctly different from academic skills.

Different lines of research, however, focus on different sets of skills. One line of research focuses on a set of five interrelated skills, often called social and emotional learning (SEL): self-awareness, self-management, social awareness, relationship skills, and responsible decision making.[27] Other research focuses more broadly on all skills that are different from academic skills. This broader class of skills, noncognitive skills or "personality traits," is understood as openness to experience, conscientiousness, extraversion, agreeableness, and emotional stability.[28]

School culture and climate

A strong school culture is associated with numerous positive outcomes for students and teachers. While there is no single definition of the term, there is agreement that it is multidimensional and includes the following constructs:[29]

- **Academic emphasis.** This is the quality of the teaching and learning within the school, including the use of strong curriculum and an expectation of strong academic outcomes for all children.

- **Community.** This encompasses measures of the interpersonal relationships within the school, trust, and a shared understanding and practice of mission across all stakeholders.

- **Safety.** This includes both physical and emotional safety within the school, including disciplinary practices and procedures and a positive, fully inclusive racial ethos.

- **Institutions.** This includes the organization and structure of the school, including the availability of resources within the school.

A strong school climate is predictive of numerous positive student outcomes, including improved indicators of academic achievement, decreased behavioral problems, and stronger student psychological and social outcomes.[30] Furthermore, research has found that from kindergarten through twelfth grade, students in schools with a stronger school climate are more likely to achieve higher test scores.[31]

For teachers, school culture influences morale, retention, and performance. Research shows that when teachers perceive that their schools have a stronger school culture, they are more likely to have high job satisfaction and self-efficacy.[32] A better school culture is also associated with decreased teacher turnover.[33] And a supportive school culture is associated with increased teacher effectiveness.[34]

School culture is typically measured through surveys, which often query students, educators, and parents. Even where measures of school climate are valid and reliable, however, they were not specifically developed for high-stakes accountability systems.[35] We are particularly concerned by the possibility of gaming on surveys (e.g., through leading directions, selective sampling, etc.), and therefore believe that these measures should be reported but not employed as a high-stakes component in an accountability system.[36] In addition, while a strong school culture is correlated with numerous positive outcomes, this research literature has not established a causal relationship from strong school climate to these positive outcomes.[37]

Civic engagement

Civic engagement is defined as "working to make a difference in the civic life of our communities and developing the combination of knowledge, skills, values and motivation to make that difference . . . through both political and non-political processes."[38] This definition requires that individuals *act* to "make a difference." *Actions,* therefore, such as voting, volunteering, participation in organizations, and working toward community improvement, could all be used as measures of civic engagement.[39] Such measures are ripe for gaming; signing up all students in the school for an organization or hosting required volunteer activities, while not necessarily bad for students, would not be indicators that

students are civically engaged. A better measure of civic engagement is educational attainment, as it is less easily gameable and rigorous research shows that increased educational attainment causes increased civic engagement in the United States.[40]

A public-facing portrait

Parents often know more about television content than about what is going on in their children's schools. The following information, once added to more formal accountability measures, could be of great value to parents in making choices about their children's schools, where to live, whom to vote for, and whether to support extra funding to supplement the school budget:

- The core values and belief systems of a school, and how these are reflected in the school's practices[41]
- A description of the curriculum and instructional materials in key subjects and an explanation of why they were chosen, together with any evidence of efficacy
- A map, by semester, of the knowledge domains students are expected to master in each subject
- A list, by semester, of the core texts (or other content) for ELA and social studies, the requisite math concepts, and scientific domains
- A clear account of class time devoted to the arts (including art, music, drama, poetry, film, and other media)
- The time, by semester, given to teachers for collaborative planning work
- Transparency across all forms of tracking: when and by what means students are placed in different academic tracks, the difference in academic outcomes by track, and the percentage of students who move up from a lower track to a higher one from one grade to the next

Even if one or more of these measures reach the bar for inclusion within the formal accountability system, they should not be overweighted in school evaluations. No school should be able to receive a strong accountability score if it is not preparing students academically for college and careers. In order to send a clear message about schools' primary academic purpose, we recommend that nonacademic measures not account for more than a quarter or a third of the total accountability score.

Beyond ESSA: Model 3.0

The above sections set out the ideal, and the potential, measures within ESSA's framework. However, ESSA does not enable the accountability levers that other, higher-performing systems—such as Singapore, South Korea, the Netherlands, and Alberta, Canada—take for granted.[42]

Specifically, ESSA's insistence on testing annually from grades three through eight and once anytime during high school, in our judgment, pushes unnecessary signals into the system and removes a key opportunity for high school students to focus on meaningful work in eleventh and twelfth grades. Moreover, although ESSA allows other testing, the fact that it doesn't require that testing removes the key incentive for states to embrace a stronger accountability system.

In general, we should not divide children's learning into a hierarchy of subjects, some of which count and others of which do not. In a number of countries, graduation examinations in economics and art, for example, have equal weight with math and ELA. We should also shift from assessments that assess generic skills (such as finding the main idea in a text a student has never seen before) to those that assess the content that students have studied. Strong research "suggests 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."[43]

With this in mind, states could bring the following changes to their accountability systems:

Test students at the entrance to kindergarten for readiness to learn.

Rationale: A test such as the Kindergarten Readiness Assessment (KRA) gives teachers invaluable information about the children they are about to teach, allowing early differentiation and skills support—and intensive support for children with the greatest needs.[44] This test would both provide diagnostic information and establish a baseline measure of students' learning.

Reassess students' ELA and math skills at the end of second grade.

This second test serves as a measure of academic quality in the early grades. In addition, such measures ensure that students have the basic math and ELA skills required for success in later grades. Low scores serve as a signal that immediate additional support is needed.

No tests are currently mandated for early elementary grades. However, research shows that early math and literacy skills are critical for later skills.[45] Early interventions are also often more effective (and cost-effective) than later interventions.[46] This suggests that measuring early skills critical for later student success and incentivizing effective early interventions

would improve the education system and student outcomes. A main challenge with assessing younger students would be to balance age-appropriate assessments with the time and resources required to administer them consistently.

Require capstone exams at the end of elementary (fifth grade); middle school (eighth grade); and at tenth grade in high school.

These assessments should be subject-based and, wherever possible, integrated with high-quality curricula. Our model is the Pilot Assessment in Louisiana in ELA: students complete an assessment after each major unit of study, based on the texts and domain knowledge in that unit. These interim assessments are then summed and averaged to create an annual assessment result at the end of the grade (the final result includes the score on an end-of-year essay).[47]

Fifth-grade tests would be in ELA, math, social studies (with a strong focus on history and civics), and science. Eighth-grade and tenth-grade tests should also include the arts and foreign languages. Tenth-grade tests could be made available in other subjects (philosophy and economics, for example).

These tests would expand the content areas currently assessed and provide a more complete measure of the school's impact on students' longer-term learning. Thus, they would shift educators' incentives toward long-term gains, place equal emphasis on all academic subjects taught, and encourage the entire school to work collectively to develop students' academic skills and knowledge.

One consequence of our proposed system is that moving toward less-frequent tests and multisubject tests makes it unlikely that the tests can be used as a measure of teacher quality. Measuring growth at the school level within school systems with high student mobility is another potential challenge. While we acknowledge these potential shortcomings, we believe that decreasing the testing burden, ensuring that tests provide more actionable information to educators, and placing greater emphasis on more subjects balance out these potential challenges. There is no one perfect system; all proposed systems will produce trade-offs. Our approach is based on the strongest international practices.

Testing at tenth grade (a recent recommendation of Maryland's Commission on Innovation and Excellence in Education)[48] will enable students to choose a substantive pre-college or career and technical (CTE) course of study for their last two years of high school—or utilize that additional high school time to reach the tenth-grade academic level.

The introduction of three distinct pathways in the final two years of high school requires improvement in CTE programs.[49] Specifically, CTE requirements should be articulated by state or national industry standards. Success in CTE proficiency should be added to school accountability metrics and should be measured by student acquisition of a state or national

industry-specific certification or licensure upon completion of a CTE Program of Study.[50] We believe that the International Baccalaureate Career-related Programme (IBCP) serves as a model of a high-quality CTE program; it prepares students for specific career pathways but does not sacrifice academic rigor in the process.[51]

One benefit from the capstone exams we suggest is that students will take the assessments more seriously. Students understand when assessments are important (e.g., SAT, ACT, and Advanced Placement for college admission) and when they are not (a state-mandated test for school accountability that has no consequences for students). A major shortcoming of accountability in the United States is that for many students who do not aspire to entry into a competitive college or career, getting a D in course work will do just fine.

We support states moving to use high school assessments that signal students' readiness for different paths and that colleges can actually use for admissions purposes. The issue with using ACT/SAT is that these tests were designed not as subject-matter tests but as aptitude tests. Recently, there is some evidence that the test designers are slowly moving the content of these assessments toward greater content testing, which would in our view make these more acceptable for school accountability use.

We believe that whatever high school assessment is used for accountability purposes should be linked to state college admissions—and, wherever possible, in a tiered way so that higher performance is rewarded with access to the stronger institutions of higher learning.

What about Consequences?

Accountability systems are meaningless unless they lead stakeholders to take effective action that improves academic outcomes for all students, and especially for those in historically low-performing schools.

Preparing for effective interventions depends on collecting accurate and comprehensive data that include the measures outlined above. School inspection models, which most OECD (Organisation for Economic Co-operation and Development) member countries already employ, can effectively provide this support to schools. Such systems rely upon highly trained observers to conduct in-depth site visits. The visits produce granular information and reports used to craft school improvement plans, against which progress can be monitored and consequences, such as school closure, be considered.

Note that the inspectorate system provides multiple functions. At the school level, inspectors provide both diagnostic information and recommendations on school improvement, such as helping develop school improvement plans, prioritizing reforms, and identifying research-based interventions and supports. The state level is provided

with detailed qualitative information about schools as well as recommendations for consequences within the accountability system. Assuming that the inspectorate's reports are made available to the public, they also provide a further level of accountability. Parents will simply be better informed about the school and can thus advocate more effectively for change—or move their children to another school if that is possible.

An Inspectorate

On-site inspections are widely used in Europe to evaluate schools. While the impacts of such systems on student outcomes are not fully understood, research suggests that such systems can have positive impacts on student learning, particularly when there are perceived stakes attached to systems and inspectorates are seen as high-quality.[52] We believe that such inspections could serve as a means to both improve schools and provide consequences within the accountability system.

Ofsted, the inspectorate model developed in England, provides one example.[53] Ofsted employs experienced educational leaders to conduct site visits.[54] During site visits, inspectors review accountability data, as well as current conditions within the school, through classroom observations, interviews, and parent and student satisfaction surveys.

A similar inspectorate could serve to review a school's progress, develop a plan to help the school improve, and determine the timing of the school's next review. For example, the inspectorate might decide that the school needed another review in five, three, or one year. One-year reviews would signal that the school needs to improve urgently and would set very specific goals for the visit the following year. Failure to meet such goals could result in leadership turnover, staff changes, or school closure. In contrast, a five-year review would signal that the school is serving students well and would give the school goals to work toward yet further improvement.

One element of an inspectorate model already exists under ESSA: School Improvement Plans (SIPs).[55] In their current form, SIPS are for schools requiring *comprehensive support* (the 5 percent lowest-performing Title 1 schools) and must be approved by the State Education Agency (SEA) and the Local Education Agency (LEA). School Improvement Plans are also required for schools requiring *targeted support* (where subgroups of students are showing very low academic outcomes) to be developed with, approved, and monitored by their LEA.

Having reviewed many examples of SIPs from across the country, we know that there is an enormous variation in the data they contain. There is *some* evidence of a positive correlation between high-quality SIPs and student learning. High-quality SIPs themselves depend on accurate school needs assessment tools which must offer schools and districts accurate interpretations of school-level data from which the SIP will be generated. Colorado has at least tried to summarize what could be in a needs assessment,[56] while New Jersey

shows the extent of data which can be considered in creating the tool.[57] New York offers a self-diagnosis tool that contains useful elements.[58] A more helpful model, however, would include a Likert scale to create a weighting of the level of performance in each domain.

We recommend that a key outcome of the inspectorate role in accountability is to generate a draft needs assessment that is then shared with school stakeholders (teachers, parents, and district) resulting in a School Improvement Plan to be approved by the state.

The core of this school-based, adult-behavior diagnostic is instruction: what are the teachers teaching and how effectively are they teaching it—and *why*? An effective needs assessment tool will surface what is known and not known about the enacted curriculum by teacher, subject, and grade level; the degree of alignment horizontally across grades; and vertical alignment (from one grade to the next) in subject matter. What is known about teacher effectiveness? For example, have the principal's observation results been correlated with student growth data? What has changed in the last year in terms of instruction: content in subject or grade level, teaching or leadership personnel, key mandates, professional development, and parental outreach and involvement? And what has *not* changed in these domains?

This tight coupling of well-trained inspectors, clear needs assessments, and calibrated consequences not only leads to better academic outcomes, but it provides parents with a rich picture of their children's schools. But even if the United States' school systems established a robust inspectorate and clear needs assessments, there would remain a fundamental systems challenge: we would be trying to create effective changes within a fundamentally incoherent educational system.

Tying It All Together: Accountability within a Coherent System

Accountability systems are not the silver bullet of education reform. They are one piece of the education system. Like other interventions, their potential cannot be evaluated except in reference to that system. For example, countries that use strong public examinations that have material consequences for individual students show stronger TIMSS (Trends in International Mathematics and Science Study) results than those that do not. But countries that add school-level autonomy over teacher salaries show still higher rates of student learning.[59]

There are in fact two key questions—not one—to be asked about the likely impact of any proposed educational intervention. The first is how strong the evidentiary base for that intervention is within our existing system. The second is whether there is any evidence for the impact of that same intervention within a restructured system—drawing, for example, on research from the use of that intervention abroad or from novel school structures within the United States.

Broadly speaking, what one often finds is that the efficacy of many educational interventions is modest in our current system when compared to their impact elsewhere. A high-quality curriculum that is regarded as an essential part of Singapore's success, for example, may make little difference in American schools. But why?

One answer is that the American public education system is extremely fragmented. Schools of Education, professional development providers, curriculum publishers, assessment companies—all potentially providing critical elements of our system—work in considerable isolation from one another and are subject to different motivations. In short, we simply don't have a coherent education system.

By contrast, top-performing countries design for educational coherence. The elements listed above are expected to mesh and mutually reinforce one another. To cite but one source here, the empirical findings cited in the PELP Coherence Framework from the Harvard Education Leadership Project outline how system coherence acts as a positive multiplier on the impact of specific educational interventions, magnifying their positive effects.[60]

Without coherence, a system looking for improvement will tend to gravitate toward an intervention that is "system-proof." For example, there is strong evidence in the United States for the efficacy of tutoring programs and of teacher-proof curriculum.[61] But would either of these initiatives have the same relative impact—compared, for example, to greatly increased planning time for teachers—if our system of teacher preparation, assessments, and school timetabling were comparable to those in top-performing systems? To take another example, we discussed the trade-offs between annual testing, which can create granular data to be used for teacher evaluation, and testing in multiple subjects only at key grade levels. Once again, if our system of teacher recruitment, compensation, preparation, and professional development matched the top systems, would we be worrying so much about trying to sort teachers into bands of performance?

In the end, top-performing countries create lighter-touch accountability systems that nevertheless provide more effective signals to students, teachers, and parents alike, because the instructional core (highly effective teachers using high-quality instructional materials) is integrated with the rest of the educational system (selective recruitment of teachers into a respected profession, substantial school planning time for teachers, high expectations, and rigorous, content-rich assessments). In a high-performing education system, each key element is held accountable for its contribution to the good of education as a whole.

Naturally, we can't in this paper assume wholesale change in the structure of education in the United States. What we have done is to outline the most efficacious accountability plan we can devise within the constraints of an otherwise slow-to-change system, with the hope that effective accountability, once implemented, will accelerate that systemic transformation.

NOTES

1 Morgan S. Polikoff, Jay P. Greene, and Kevin Huffman, "Is Test-Based Accountability Dead?" *Education Next* 17, no. 3 (Summer 2017), https://www.educationnext.org/is-test-based-accountability-dead-forum-polikoff-greene-huffman.

2 Susanna Loeb and David Figlio, "School Accountability," in *Handbook of the Economics of Education,* ed. Eric A. Hanushek, Stephen Manchin, and Ludger Woessman, vol. 3 (San Diego, CA: North-Holland, 2011), 383–423, https://cepa.stanford.edu/content/school-accountability.

3 Jay P. Greene, "Futile Accountability Systems Should Be Abandoned," *Education Next* 17, no. 3 (Summer 2017), https://www.educationnext.org/futile-accountability-systems-should-be-abandoned-forum-greene.

4 Loeb and Figlio, "School Accountability."

5 Brian P. Gill, Jennifer S. Lerner, and Paul Meosky, "Reimagining Accountability in K–12 Education," *Behavioral Science & Policy* 2, no. 1 (2016): 57–70; Susanna Loeb and Erika Byun, "Testing, Accountability, and School Improvement," *Annals of the American Academy of Political and Social Science* 683, no. 1 (2019): 94–109.

6 Gill, Lerner, and Meosky, "Reimagining Accountability."

7 "The Difference Between the Every Student Succeeds Act and No Child Left Behind," Understood, https://www.understood.org/en/school-learning/your-childs-rights/basics-about-childs-rights/the-difference-between-the-every-student-succeeds-act-and-no-child-left-behind.

8 "Every Student Succeeds Act (ESSA) Overview," National Association of Secondary School Principals, accessed May 20, 2020, https://www.nassp.org/policy-advocacy-center/resources/essa-toolkit/essa-fact-sheets/every-student-succeeds-act-essa-overview.

9 Laura Jimenez and Scott Sargrad, "A New Vision for School Accountability," *Center for American Progress,* March 3, 2017, https://www.americanprogress.org/issues/education-K–12/reports/2017/03/03/427156/a-new-vision-for-school-accountability.

10 "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." See 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.

11 We are indebted to Chester "Checker" Finn Jr. for this recommendation.

12 For an example, see the use of the Bridge Plan for Validation in Maryland, Maryland State Department of Education, December 4, 2017, http://www.marylandpublicschools.org/stateboard/Documents/12042017/BridgePlanAcademicValidation.pdf.

13 For current state plans, see "Defining Gifted and Talented Student Group," Maryland State Department of Education, October 23, 2018, https://www.google.com/search?client=safari&rls=en&q=Maryland+will+count+gifted+student+perfomance+in+accountability&ie=UTF-8&oe=UTF-8.

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

15 Resolving the way in which to calculate growth is also dependent upon a number of data requirements. For more information, see Katherine E. Castellano and Andrew D. Ho, "A Practitioner's Guide to Growth Models," Council of Chief State School Officers, 2013, https://scholar.harvard.edu/files/andrewho/files/a_pracitioners_guide_to_growth_models.pdf; and Mark Ehlert, Cory Koedel, Eric Parsons, and Michael Podgursky, "Choosing the Right Growth Measure," *Education Next* 14, no. 2 (Spring 2014), https://www.educationnext.org/choosing-the-right-growth-measure.

16 Matt Barnum, "Growth Plus Proficiency? Why States Are Turning to a Hybrid Strategy for Judging Schools (And Why Some Experts Say They Shouldn't)," *Chalkbeat* (blog), June 22, 2017, https://chalkbeat.org/posts/us/2017/06/22/growth-plus-proficiency-why-states-are-turning-to-a-hybrid-strategy-for-judging-schools-and-why-some-experts-say-they-shouldnt.

17 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; Dave E. Marcotte and Steven W. Hemelt, "Unscheduled School Closings and Student Performance," *Education Finance and Policy* 3, no. 3 (Summer 2008): 316–38.

18 Esteban Aucejo and Teresa Foy Romano, "Assessing the Effect of School Days and Absences on Test Score Performance," *Economics of Education Review* 55 (2016): 70–87.

19 Michael A. Gottfried, "The Detrimental Effects of Missing School: Evidence from Urban Siblings," *American Journal of Education* 117, no. 2 (2011): 147–82; Joshua Goodman, "Flaking Out: Student Absences and Snow Days as Disruptions of Instructional Time," National Bureau of Economic Research, 2014; Seth Gershenson, Alison Jacknowitz, and Andrew Brannegan, "Are Student Absences Worth the Worry in US Primary Schools?" *Education Finance and Policy* 12, no. 2 (Spring 2017): 137–65; Jing Liu, Monica Lee, and Seth Gershenson, "The Short- and Long-Run Impacts of Secondary School Absences," IZA Discussion Paper No. 12613, 2019.

20 Gershenson, Jacknowitz, and Brannegan, "Are Student Absences Worth the Worry?"; Michael A. Gottfried, "Chronic Absenteeism in the Classroom Context: Effects on Achievement," *Urban Education* 54, no. 1 (January 2019): 3–34.

21 For example, see Robert Balfanz and Vaughan Byrnes, "The Importance of Being in School: A Report on Absenteeism in the Nation's Public Schools," *Education Digest* 78, no. 2 (October 2012): 4–9; Mariajosé Romero and Young-Sun Lee, "The Influence of Maternal and Family Risk on Chronic Absenteeism in Early Schooling," National Center for Children in Poverty, 2008; Mariajosé Romero and Young-Sun Lee, "A National Portrait of Chronic Absenteeism in the Early Grades," National Center for Children in Poverty, 2007.

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

23 Michael Q. McShane and Paul DiPerna, "Do-over or Double Down? Working Toward a New K–12 Education Accountability Ecosystem," EdChoice, October 2018, https://www.edchoice.org/wp-content/uploads/2018/10/2018-10-Do-Over-or-Double-Down-by-Michael-Q.-McShane-and-Paul-DiPerna.pdf; "Parents 2018: Going Beyond Good Grades," Learning Heroes, December 2018.

24 We are open—in theory—to the idea of holding primary schools and middle schools accountable for the performance of students in the next level of their schooling. But the frequent and constant changes in schools annually—from staffing to instructional initiatives, for example—present problems in the meaning and usefulness of such measures and we are not convinced that such challenges can be overcome.

25 Susana Claro and Susanna Loeb, "Students with Growth Mindset Learn More in School: Evidence from California's CORE School Districts," Policy Analysis for California Education, October 2019; David S. Yeager, Paul Hanselman, Gregory M. Walton, et al., "A National Experiment Reveals Where a Growth Mindset Improves Achievement," *Nature* 573, no. 7774 (2019): 364–69.

26 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; Joshua Cox, Brandon Foster, and David Bamat, "A Review of Instruments for Measuring Social and Emotional Learning Skills among Secondary School Students," Regional Educational Laboratory Northeast & Islands, October 2019.

27 Joseph A. Durlak, Roger P. Weissberg, Allison B. Dymnicki, Rebecca D. Taylor, and Kriston B. Schellinger, "The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal

Interventions," *Child Development* 82, no. 1 (2011): 405–32. Note that these skills are often referred to as skills that promote "social and emotional learning."

28 James J. Heckman and Tim Kautz, "Hard Evidence on Soft Skills," *Labour Economics* 19, no. 4 (2012): 451–64.

29 Amrit Thapa, Jonathan Cohen, Shawn Guffey, and Ann Higgins-D'Alessandro, "A Review of School Climate Research," *Review of Educational Research* 83, no. 3 (September 2013): 357–85; Ming-Te Wang and Jessica L. Degol, "School Climate: A Review of the Construct, Measurement, and Impact on Student Outcomes," *Educational Psychology Review* 28, no. 2 (June 2016): 315–52, https://doi.org/10.1007/s10648-015-9319-1.

30 Thapa et al., "School Climate Research"; Wang and Degol, "School Climate."

31 Angus J. MacNeil, Doris L. Prater, and Steve Busch, "The Effects of School Culture and Climate on Student Achievement," *International Journal of Leadership in Education* 12, no. 1 (January 1, 2009): 73–84, https://doi.org/10.1080/13603120701576241.

32 Jill M. Aldridge and Barry J. Fraser, "Teachers' Views of Their School Climate and Its Relationship with Teacher Self-Efficacy and Job Satisfaction," *Learning Environments Research* 19, no. 2 (July 1, 2016): 291–307, https://doi.org/10.1007/s10984-015-9198-x; Olli-Pekka Malinen and Hannu Savolainen, "The Effect of Perceived School Climate and Teacher Efficacy in Behavior Management on Job Satisfaction and Burnout: A Longitudinal Study," *Teaching and Teacher Education* 60 (November 1, 2016): 144–52, https://doi.org/10.1016/j.tate.2016.08.012.

33 Matthew A. Kraft, William H. Marinell, and Darrick Shen-Wei Yee, "School Organizational Contexts, Teacher Turnover, and Student Achievement: Evidence from Panel Data," *American Educational Research Journal* 53, no. 5 (October 1, 2016): 1411–49, https://doi.org/10.3102/0002831216667478.

34 Matthew A. Kraft and John P. Papay, "Can Professional Environments in Schools Promote Teacher Development? Explaining Heterogeneity in Returns to Teaching Experience," *Educational Evaluation and Policy Analysis* 36, no. 4 (December 1, 2014): 476–500, https://doi.org/10.3102/0162373713519496.

35 See "School Climate Survey Compendium," National Center on Safe Supportive Learning Environments, accessed May 20, 2020, https://safesupportivelearning.ed.gov/topic-research/school-climate-measurement/school-climate-survey-compendium.

36 This is particularly important, as research shows that teachers, students, and parents all have differing views of school climate indicators. See Christine M. Ramsey, Adam P. Spira, Jeanine M. Parisi, and George W. Rebok, "School Climate: Perceptual Differences between Students, Parents, and School Staff," *School Effectiveness and School Improvement* 27, no. 4 (October 1, 2016): 629–41, https://doi.org/10.1080/09243453.2016.1199436.

37 See, for example, Wang and Degol, "School Climate"; Ruth Berkowitz, Hadass Moore, Ron Avi Astor, and Rami Benbenishty, "A Research Synthesis of the Associations Between Socioeconomic Background, Inequality, School Climate, and Academic Achievement," *Review of Educational Research* 87, no. 2 (April 2017): 425–69.

38 Thomas Ehrlich, *Civic Responsibility and Higher Education* (Westport, CT: Greenwood Publishing Group, 2000), vi.

39 Note that this is a standard way of measuring civic engagement. See, for example, Adam Voight and Judith Torney-Purta, "A Typology of Youth Civic Engagement in Urban Middle Schools," *Applied Developmental Science* 17, no. 4 (October 1, 2013): 198–212, https://doi.org/10.1080/10888691.2013.836041.

40 Kevin Milligan, Enrico Moretti, and Philip Oreopoulos, "Does Education Improve Citizenship? Evidence from the United States and the United Kingdom," *Journal of Public Economics* 88, no. 9 (2004): 1667–95; Rachel Milstein Sondheimer and Donald P. Green, "Using Experiments to Estimate the Effects of Education on Voter Turnout," *American Journal of Political Science* 54, no. 1 (2010): 174–89; Thomas S. Dee, "Are There Civic Returns to Education?" *Journal of Public Economics* 88, no. 9 (2004): 1697–1720.

41 England already requires the reporting of such information from all of its publicly funded schools. "What Maintained Schools Must Publish Online," Department for Education, United Kingdom (2018), cited in Ashley Berner, "How School Culture Drives Civic Knowledge and Shapes the Next Generation of Citizens," *The 74*

Million, April 18, 2017, https://www.the74million.org/article/berner-how-school-culture-drives-civic-knowledge-and-shapes-the-next-generation-of-citizens.

42 "Singapore: Learning Systems," National Center on Education and the Economy, accessed May 20, 2020, http://ncee.org/what-we-do/center-on-international-education-benchmarking/top-performing-countries/singapore-overview-2/singapore-learning-systems; "South Korea: Learning Systems," National Center on Education and the Economy, accessed May 20, 2020, http://ncee.org/what-we-do/center-on-international-education-benchmarking/top-performing-countries/south-korea-overview/south-korea-instructional-systems; "K to 12 Provincial Assessment," Government of Alberta, accessed May 20, 2020, https://www.alberta.ca/K–12-provincial-assessment.aspx.

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

44 "Kindergarten Readiness Assessment (KRA)," Maryland State Department of Education, accessed May 20, 2020, http://www.marylandpublicschools.org/about/Pages/DAAIT/Assessment/KRA/index.aspx.

45 Greg J. Duncan, Amy Claessens, Aletha C. Huston, et al., "School Readiness and Later Achievement," *Developmental Psychology* 43, no. 6 (2007): 1428; Amy Claessens, Greg Duncan, and Mimi Engel, "Kindergarten Skills and Fifth-Grade Achievement: Evidence from the ECLS-K," *Economics of Education Review* 28, no. 4 (August 2009): 415–27.

46 See, for example: Christopher Schatschneider, David J. Francis, Coleen D. Carlson, Jack M. Fletcher, and Barbara R. Foorman, "Kindergarten Prediction of Reading Skills: A Longitudinal Comparative Analysis," *Journal of Educational Psychology* 96, no. 2 (2004): 265; James J. Heckman, "Schools, Skills, and Synapses," *Economic Inquiry* 46, no. 3 (2008): 289–324; Meree Reynolds, Kevin Wheldall, and Alison Madelaine, "What Recent Reviews Tell Us about the Efficacy of Reading Interventions for Struggling Readers in the Early Years of Schooling," *International Journal of Disability, Development and Education* 58, no. 3 (2011): 257–86.

47 "Louisiana's Innovative Assessment Pilot National Update," Louisiana Department of Education, August 15, 2019, https://www.louisianabelieves.com/docs/default-source/assessment/national-update-on-la%27s-innovative-assessment-pilot.pdf?sfvrsn=9e849d1f_4.

48 "Maryland Commission on Innovation & Excellence in Education: Interim Report," Maryland Commission on Innovation and Excellence in Education, January 2019, http://dls.maryland.gov/pubs/prod/NoPblTabMtg/CmsnInnovEduc/2019-Interim-Report-of-the-Commission.pdf.

49 We note that while such changes are currently possible under ESSA, such a change is particularly important given the proposed changes in tenth grade.

50 Marisa Castellano, Kirsten E. Sundell, Laura T. Overman, George B. Richardson, and James R. Stone III, "Rigorous Tests of Student Outcomes in CTE Programs of Study: Final Report," National Research Center for Career and Technical Education, April 2014.

51 "Career-Related Programme," International Baccalaureate, accessed May 21, 2020, http://www.ibo.org/programmes/career-related-programme.

52 Current research shows mixed results that—not surprisingly—depend on the specifics of inspection regimes. For example, recent research by Annika B. Bergbauer, Eric A. Hanushek, and Ludger Woessman ("Testing," National Bureau of Economic Research, NBER working paper 24836, November 2019) finds no significant effect on student outcomes when accountability systems introduce observations by external observers (i.e., classes observed "by inspectors or other persons external to the school"). A systematic review of research about inspectorate systems finds mixed results of the impact of inspectorate systems on student outcomes. See Sarah I. Hofer, Doris Holzberger, and Kristina Reiss, "Evaluating School Inspection Effectiveness: A Systematic Research

Synthesis on 30 years of International Research," *Studies in Educational Evaluation* 65 (June 2020), https://www.sciencedirect.com/science/article/abs/pii/S0191491X20301127. However, this review finds more positive impacts of inspectorate systems under certain conditions, such as perceived high quality of the inspectorate system and perceived accountability pressure from the inspectorate.

53 The following section draws extensively on a paper on the topic written by our colleague, Ashley Berner. For the full paper, see "Would School Inspections Work in the United States?" Johns Hopkins School of Education, Institute for Education Policy, September 14, 2017, https://edpolicy.education.jhu.edu/would-school-inspections-work-in-the-united-states.

54 Craig D. Jerald, "On Her Majesty's School Inspection Service," Education Sector, 2012, https://edpolicyinca.org/sites/default/files/UKInspections-RELEASED.pdf.

55 "Every Student Succeeds Act—Accountability, State Plans, and Data Reporting: Summary of Final Regulations," US Department of Education, November 28, 2016, https://www2.ed.gov/policy/elsec/leg/essa/essafactsheet1127.pdf.

56 "Needs Assessment," Colorado Department of Education, accessed May 21, 2020, http://www.cde.state.co.us/sites/default/files/documents/fedprograms/dl/consapp_na_tia-na.pdf.

57 Alder Avenue Middle School evaluation, State of New Jersey, 2015, https://www.nj.gov/education/pr/1415/01/011310038.pdf.

58 "Diagnostic Tool for School and District Effectiveness (DTSDE) Process and Resources," New York State Education Department, http://www.nysed.gov/accountability/dtsde.

59 West and Woessman, "School Choice, Autonomy, and Accountability."

60 "Case Studies & Research Notes," Public Education Leadership Project at Harvard University, accessed May 21, 2020, https://pelp.fas.harvard.edu/case-studies-notes; "Coherence Framework," Public Education Leadership Project at Harvard University, accessed May 21, 2020, https://pelp.fas.harvard.edu/coherence-framework.

61 Robert Slavin, "New Findings on Tutoring: Four Shockers," *Robert Slavin's Blog,* April 5, 2018, https://robertslavinsblog.wordpress.com/2018/04/05/new-findings-on-tutoring-four-shockers; "Project Follow Through," National Institute for Direct Instruction, accessed May 21, 2020, https://www.nifdi.org/what-is-di/project-follow-through.

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

About the Authors

DAVID STEINER

David Steiner is executive director of the Johns Hopkins Institute for Education Policy and professor of education at Johns Hopkins University. He currently serves as a member of the Maryland State Board of Education. He previously served as commissioner of education for New York State. He holds BA and MA degrees from Oxford University and a PhD from Harvard University.

ALANNA BJORKLUND-YOUNG

Alanna Bjorklund-Young is a senior research and policy analyst at the Johns Hopkins Institute for Education Policy and an assistant professor of education at Johns Hopkins University. Her research primarily focuses on teacher quality, student achievement, and curriculum. She previously taught in New York and Mexico City. She holds an MA and PhD in economics from Johns Hopkins University.

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