Defining Strong State Accountability Systems

We're not the first to delineate a framework for sound state accountability systems. Our thinking was informed by the good work of several other organizations, including the Council of Chief State School Officers (CCSSO) and the George W. Bush Institute. CCSSO's *Roadmap for Next-Generation State Accountability Systems* lays out nine principles—some of which we adopt verbatim—relative to standards, student outcomes, annual determination of school and district performance, timely data, and interventions for failing schools, among other areas.[6] Similarly, in *Advancing Accountability*, the Bush Institute delineates ten school-level accountability principles for states, including provisions pertaining to performance and content standards, assessments, reporting, intervention, and choice alternatives.[7] The primary difference between those excellent reports and this one is the focus on *who* is to be held accountable. Prior reports primarily target schools and districts, while we broaden the focus to include individuals: students, teachers, principals, and other adults. After all, smart accountability should start with the individual. Not that collective accountability has no value; rather, it should begin with responsible individuals who also "have skin in the game." (See sidebar, *Putting Our Cards on the Table*, for more of our assumptions.)

Thus, we posit that strong state accountability systems have six essential elements:

1. Adoption of demanding, clear, and specific standards in all core content areas, and rigorous assessment of those standards;

2. Reporting of accessible and actionable data to all stakeholders, including summative outcome data and other formative data to drive continuous improvement;

---

[6] See Council of Chief State School Officers, *Roadmap for Next-Generation State Accountability Systems* (Washington, D.C.: Council of Chief State School Officers, 2011), http://www.ccsso.org/Resources/Publications/Roadmap_for_Next-Generation_State_Accountability_Principles.html.

[7] See The George W. Bush Institute, *Advancing Accountability* (Dallas, TX: The George W. Bush Institute, February 2012), http://www.bushcenter.com/blog/wp-conten/uploads/2012/02/Advancing_AccountabilityFeb-2012.pdf.

3. Annual determinations and designations for each school and district that meaningfully differentiate their performance;

4. A system of rewards and consequences to drive improvement at the school and district levels;

5. A system of rewards and consequences to drive improvement at the individual student level; and

6. A system of rewards and consequences to drive improvement at the individual teacher and administrator level.

Below, we briefly describe each; for detailed explanation and scoring guidelines, see Appendix A.

1) Adoption of demanding, clear, and specific standards in all core content areas, and rigorous assessment of those standards

Sound content standards are an essential *starting point* for improved student outcomes. States should, at a minimum, have them in place for English language arts (including writing), mathematics, U.S. History, and science, and should assess student mastery of subjects at regular intervals. Further, states should hold all students to high standards by setting rigorous cut scores for each test.[8]

2) Reporting of accessible and actionable data to all stakeholders, including summative outcome data and other formative data to drive continuous improvement

Robust and timely diagnostic and outcome data on school, student, and adult performance enable educators, policymakers, parents, students, and the broader public to take appropriate action on the decisions that concern them (e.g., teacher hiring, school choice, classroom interventions, selection of turnaround model, just to name a few). Though we appreciate the need for a wide variety of data to assess sundry outcomes at all levels, we target here the data of most concern: student performance.

States should release annual aggregate student outcome data in an accessible manner (meaning clear and user-friendly). In short, states should publish results for every school that includes a snapshot of how many students are reaching various achievement levels (such as basic, proficient, advanced, etc.), how many students are making expected progress over time, and whether or not the school

[8] Note that proper alignment of standards and tests is critical—including a matchup of rigor—but we're unable to evaluate that alignment.

Putting Our Cards on the Table

The metric that we outline here is premised on certain assumptions and faces some limitations. First, the principles we discuss are meant to help states define the *parameters* for strong accountability systems. We don't pretend that there is one best way to design these systems. Even if there *were* one best approach today, the constantly changing landscape requires that states persist in experimenting with and tweaking their systems. States continue to scale the steep learning curve erected since the establishment of No Child Left Behind: Timid sanctions and unimplemented interventions leave many questioning how best to address failing schools; local stakeholders remain befuddled in their attempts to navigate dual federal and state accountability systems (though the waivers should help on this front); unexplained student growth or stagnation begs additional examination; and so on. Thus, many states are experimenting with innovative methods to hold schools, adults, and students accountable for outcomes; to gather and report data more usefully; to differentiate among dismal, failing, average, good, and great schools; and to incentivize and crack down on particular behaviors. Our principles are therefore more holistic than prescriptive, leaving ample room for experimentation and innovation. For example, we are agnostic as to how states choose to incentivize effective leadership—but they should recognize the importance of such policies.

Second, measuring the *quality* of state policies is easier said than done, and we do not attempt to do so here. We are well aware, for instance, that simply deploying a high school graduation test for the purpose of student-level accountability does not mean that the test is any good. It may tell us nothing about whether or not a passing grade indicates a student is ready for college or career. But to answer the "quality question" would require more time and hands than our resources allow, as metrics by which to evaluate the usefulness, rigor, and application of particular policies simply do not exist. For instance, we favor states that require individual students to pass "gateway" assessments, perhaps in elementary reading, before being promoted to the next grade. But to evaluate the application of that policy we'd also need to know how many students failed the gateway exam in a given year and were subsequently retained as a result. Few states release these data in timely fashion, if at all. So states can have a policy on the books but not enforce it—or not measure its enforcement—and we do not account for such lax implementation in our metric.

Third, we concentrate on provisions relative to accountability and incentives rather than those intended to support failing schools and students or weak teachers and principals. Our focus is on ensuring robust data, making results transparent, rewarding good work, and penalizing chronically low performance. We recognize that providing tailored support to organizations and individuals in need is critical to a strong accountability system, but assessing the adequacy of these measures is difficult: Many states offer similar supports of the same name and it is nearly impossible to draw meaningful distinctions among them without in-depth analysis.

Finally, a clarification: While we recognize that the term "intervention" can take one of two forms—supports (e.g., state-operated regional service centers or technical assistance in improvement planning) or consequences (e.g., mandatory replacement of staff or contracting with outside entities to run the school)—we address in these pages only the latter.

itself is improving from year to year. (Of course, states should also require that schools provide reports to parents on their own child's performance and progress.) In addition, the state should collect and report data on multiple measures of student performance whenever possible. These might include attendance, graduation rates, on-track indicators, postsecondary matriculation/retention or remediation rates, course completion (transcript data), attainment of industry certification, and performance on SAT, ACT, and AP/IB exams. State databases should also permit school-level comparisons.

States should also require that the data above be disaggregated by subgroups within schools of race/ethnicity, gender, free and reduced-price lunch status, special-needs status, English language learner status, and so on. States might also choose to substitute one or more of these groups with a low-performing subgroup or high-performing subgroup (i.e., a "super subgroup").

3) Annual determinations and designations for each school and district that meaningfully differentiate their performance

Stakeholders need to know how schools and districts compare with one another. States need to characterize and differentiate among schools and districts based on student achievement in a valid and reliable manner.

These designations should not obscure the truth (i.e., via inflated performance or categories that permit high percentages of schools to fall disproportionally into the top tiers). They should be user-friendly—we prefer the A-to-F grading system. And the state should integrate state and federal accountability systems so that designations do not conflict.[9]

School and district designations should be derived, at least in part, from measures of individual student growth (either normative growth or growth to standard). Those designations should also be informed by the performance of specific student groups (e.g., race/ethnicity, income, English language learner status, and special-needs status) or "super subgroups."[10]

4) A system of rewards and consequences to drive improvement at the school and district levels

Strong state accountability systems engage the multiple entities involved in educating children, including districts, schools, educators, and students, and incorporate both carrots and sticks to spur positive action and behavior. At the

[9] As mentioned, the federal ESEA waivers are encouraging this integration.
[10] States might also want to use some form of human judgment when rating schools. The "school inspectorate" idea, borrowed from the United Kingdom, is particularly intriguing. Such inspectors could potentially adjust a school's letter grade up or down depending on what they observe on the ground.

district and school level, incentives may include "honor roll" or similar "blue ribbon/distinguished" status, financial rewards, increased autonomy in operations/spending, and other regulatory relief (e.g., automatic renewal of district accreditation) in exchange for meeting heightened district accountability provisions and/or as a reward for performance or growth.

State accountability systems should also require targeted interventions into low-performing schools.[11] These may include replacement of staff, charter conversion, state takeover, contracts with outside entities to operate schools, and automatic school closure after consecutive years of failure. (Interventions may also be catalyzed via "parent trigger" provisions.) Such consequences need to be *real* (or have "teeth"), so the state's system should avoid introducing loopholes by which low-performing schools or districts can evade tough sanctions; water down or delay sanctions; and/or restrict sanctions to a subset of the lowest-performing schools.[12]

5) A system of rewards and consequences to drive improvement at the individual student level

In the last decade, we've spent much time thinking about how to hold districts, schools, and—less often—educators accountable. But students are often not held to account for their own learning. Strong state accountability systems put some of the responsibility for successful student outcomes back where it belongs—on the shoulders of individual students, and, implicitly, their parents, guardians, and others who influence their actions and attitudes.

Consequently, states should have on the books multiple methods by which individual students are held to account for their performance. These may include, but are not limited to: requirements that students pass cumulative high-school exit exams in the core subject areas in order to receive diplomas; pass end-of-course exams to get credit for courses or pass end-of-course exams that are integrated into final course grades; and pass "gateway" assessments in one or more key grades and subjects in order to be promoted to the next grade (e.g., pass a reading test at the end of third grade). States may also have a "no pass, no play" policy that requires students to pass all of their courses in order to participate in extracurricular activities, or a "no pass, no driver's license" policy. (We're open to other ideas here, too.)

[11] As discussed on page 7, interventions comprise both supports and sanctions. We've targeted the latter here.
[12] As mentioned, we're unable to assess quality in this metric as thoroughly as we'd like. Still, we've attempted here to illustrate what we think good state policy does and does not do when it comes to meaningful interventions for low-performing schools.

Finally, states might also reward individual students via guaranteed entry into a state college and/or guaranteed college financial aid (or "priority status") for eligible students with high GPAs or high SAT or ACT scores.

6) A system of rewards and consequences to drive improvement at the individual teacher and administrator level

States should require annual evaluations of all teachers. Further, teacher evaluations for all core subject areas should include measures of student performance and/or growth on state assessments—and those results should inform decisions related to continuous employment.

The state should also have in place policies that incentivize individual teacher effectiveness and productivity, such as merit pay and/or other individual performance-based bonuses (including for teaching in high-needs schools), extended contracts, career ladders (e.g., master teachers), personnel decisions based on performance (e.g., no "last in, first out" policies), and so on.

Adult accountability does not end with teachers. The state should also mandate annual evaluations of all principals and these should be based, in part, on their schools' performance; results should inform employment status. Finally, the state should offer incentives and rewards for effective principals, including bonuses, extended contracts, enhanced autonomy, and so on. States might also offer incentives for superintendents and/or school board members based on their schools' performance.

• • •

If lawmakers can put these six elements in place, they'll be well on their way to bolstering the likelihood of success for students, educators and other adults toiling away in the state's schools. Let's turn now to the findings in some of these states.

Summary of Pilot Results

As indicated, our work builds upon the prior research of other organizations. But it is also largely informed by in-depth analysis of the accountability systems in seven states: California, Colorado, Florida, Indiana, Massachusetts, Ohio, and Texas. (See sidebar, *Selection of States*, for why we chose this set.)

Each of the seven detailed profiles can be found in Appendix B. We conducted these analyses before the first batch of states applied for ESEA waivers, so some of the information for four of our states (Colorado, Florida, Indiana, and

Selection of States

We chose states for this pilot study based in part on laudable properties of their state accountability systems, but more broadly on their strong history of education reform. We reasoned that aggressive reform states might also be prime candidates for strong accountability systems. Here's a sampling of the elements that informed our selection.

California and **Massachusetts** have long been recognized for their strong academic standards,[13] though the Golden State cannot boast the impressive performance of the Bay State, which routinely scores at the top of the heap domestically and near the top internationally.[14] Massachusetts is well-known and respected for setting rigorous passing standards on state assessments and for requiring that students meet those standards to earn a diploma.

Florida and **Indiana** have been recognized for their high-quality state assessments.[15] The Sunshine State has also posted impressive gains for student subgroups, particularly its Hispanic students.[16] A perennial leader in school innovation since Governor Jeb Bush launched reforms in 1999, Florida passed a package of reforms in spring 2011 that abolished tenure for newly hired teachers, established a performance-based pay system, provided additional pay for high-need subject areas and at-risk schools, and further expanded charter schools and digital learning. Likewise, the Hoosier State reformed collective bargaining in 2011, expanded charter campuses, removed caps on virtual schools, and passed an expansive voucher program for low-income students.

In 2003, **Ohio** was one of the first states to mandate that value-added models be used to measure student performance. The Buckeye State also boasts two voucher programs and a robust school choice climate. In 2006, the state instituted an automatic closure ("death penalty") provision for poorly performing charter schools, considered to be the toughest closure law in the nation (a similar policy does not exist for traditional schools).

Texas was one of the first states to shine a light on the performance of all student groups—a concept that then-Governor George W. Bush sought to enshrine in No Child Left Behind when he became president. The Lone Star State has posted impressive gains for minority students

---

[13] See the Fordham Institute's recent reports on state standards, *The State of State Standards—and the Common Core—in 2010*, *The State of State U.S. History Standards 2011*, *and The State of State Science Standards 2012*, at http://standards.educationgadfly.net/; "Quality Counts 2011: Uncertain Forecast—State Report Cards," *Education Week*, January 2011, http://www.edweek.org/ew/qc/2011/16src.h30.html; and American Federation of Teachers, *Sizing Up State Standards 2008* (Washington, D.C.: American Federation of Teachers, 2008), http://www.aft.org/pdfs/teachers/sizingupstandards0308.pdf.

[14] For example, while Massachusetts eighth graders scored higher than any other state's eighth graders on the 2009 reading and mathematics National Assessment of Educational Progress (NAEP), California's eighth graders scored third-to-last among the states in reading and fourth-to-last in mathematics. See the NAEP website at http://nces.ed.gov/nationsreportcard/. See also Gary Phillips, *International Benchmarking: State Education Performance Standards* (Washington, D.C.: American Institutes for Research, October 2010), http://www.air.org/files/AIR_Int_Benchmarking_State_Ed__Perf_Standards.pdf.

[15] See the Florida and Indiana profiles in "Quality Counts 2011: Uncertain Forecast—State Report Cards," *Education Week*, January 2011, http://www.edweek.org/ew/qc/2011/16src.h30.html.

[16] Matthew Ladner and Dan Lips, "Demography and Destiny? Hispanic Student Success in Florida," *Education Next* 9, no. 3 (Summer 2009):20-27, http://educationnext.org/demography-as-destiny-2/. See also the NAEP website at http://nces.ed.gov/nationsreportcard/.

(though these have flattened of late) as well as set the standard for the country to test students regularly in the core content areas, disaggregate performance, and tie accountability measures to those outcomes.

Colorado is recognized for its robust system of school choice. Its capital city also claims a sophisticated teacher evaluation and compensation system, one of the nation's first pay-for-performance pioneers. In addition, in May 2011 it passed a comprehensive legislative package (S.B. 191), which reformed teacher tenure, defined teacher effectiveness based primarily on student performance, and permitted districts to lay off teachers based on performance rather than seniority.

Massachusetts)[17] may now be outdated. (See the sidebar *ESEA Waivers Alter State Accountability Systems* for a discussion of overall patterns in the approved waivers). We gathered information on state accountability systems from multiple sources, including state department of education websites; interviews with present and past employees of the state education agencies,[18] district superintendents and/or other district staff, and on-the-ground stakeholders; and extant data sources.[19] Interviewing multiple insiders at the state and local levels provided a window into how state- and district-level accountability policies and practices interacted (and sometimes collided) on the ground. In addition, we reached out to representatives of national organizations—including the Council of Chief State School Officers, the U.S. Education Delivery Institute, Achieve, and the Aspen Institute—in order to gain a better understanding of the current accountability landscape and add context to our state reviews. In total, we interviewed over forty education representatives between summer 2011 and early winter 2012.

In short, our examination yielded six themes—three key strengths in state accountability systems and three key weaknesses. First, the strengths:

1. **A number of states have developed clear and comprehensive means for rating district and school performance.** Robust accountability systems turn multiple data points into user-friendly labels. For instance,

[17] Ohio has applied for a waiver in the second round of requests.
[18] Texas is the one exception—the Texas Education Agency declined our request to be interviewed for this paper, and thus the Texas profile includes district-level, but not state-level, input. For Massachusetts, we interviewed personnel associated with both the Executive Office of Education and the Department of Elementary and Secondary Education.
[19] In particular, we consistently drew information from the National Council on Teacher Quality's *2011 State Teacher Policy Yearbook*. Other sources are cited in the individual profiles.

Florida and (more recently) Indiana use A-to-F grades to evaluate performance. These identifiers are informed by an array of achievement measures, including student proficiency and growth—but also by such other indicators as graduation and high school dropout rates, attendance, participation in AP/IB courses, achievement gaps, and performance on the SAT and ACT. Florida, for example, includes in its measures the performance of the lowest 25th percentile of students at each school, while Texas considers schools' college-going rates and college-level remediation rates.

2. **States are collecting a variety of data and making progress toward rendering them more "transparent" and "user-friendly."** States must provide information to educators, parents, and taxpayers in order to help them understand and evaluate school and district results. And indeed, many states ensure that such stakeholders are awash in data relative to state, district, and school performance as well as sundry other areas (see above). Some states also publish results from various diagnostic and progress monitoring tools, as well as discrete data from online courses. Presenting these myriad data in a user-friendly fashion is a continuing challenge, but states such as Ohio are making inroads. In addition to its school report card, Ohio also issues a web-based interactive Local Report Card (iLRC) that easily allows comparison among schools and districts.

3. **Teacher accountability is improving.** More and more states, such as Florida, Indiana, and Colorado, are requiring student achievement measures to feature prominently in teacher evaluations, which then inform employment decisions. More states are also questioning automatic tenure after two to three years and experimenting with various performance-pay options.

Now, for the weaknesses:

4. **States still struggle with how to meld state and federal accountability systems.** Since the introduction of the No Child Left Behind Act, states have grappled with how best to integrate state and federal requirements so as to avoid conflicting messages. Of course, the Obama administration's waivers are intended to help address this problem. We hope that, in time, ESEA reauthorization will address it as well. While a few states have seamlessly combined their federal and state systems, most struggle to operate two parallel, disjointed systems—resulting in dual performance standards, designations, and sanctions for schools and districts. Such a twofold structure is not just onerous for schools and districts, but also confusing for educators and parents, particularly when a school meets one system's performance standards but not the other.

Massachusetts is an exemplar here. It designed one system that integrates federal and state accountability provisions. This allows the Bay State to report one marker of school performance based on *both* state and federal measures, and to set individual growth goals that also meet federal requirements for each school and district.

5. **Most states have only weak incentives to drive school, district, principal, and student performance.** Generally, states have policies that require intervention on the behalf of low performers—whether they are districts, schools, or students—but few policies that incentivize or reward high performers. Further, recent budget reductions have largely eliminated performance-based incentives for schools, teachers, and principals. (Accountability policies, in fact, tend to ignore positive incentives for school leaders altogether.) States do a *tad* better rewarding students; a couple of them offer high-performing students automatic college admission or financial aid.

6. **Sanctions are non-existent or ineffective.**[20] We're big supporters of transparent data both to motivate and to shame schools, districts, and individuals to action. But transparent data must be coupled with tangible incentives that reward high performance (as in #5 above) and meaningful sanctions that penalize low performance. Previous research has shown (and those we interviewed agreed) that NCLB's cascade of interventions was unsuccessful, in part, because districts took the road of least resistance. How might low performance be handled differently this time around? There's heated discussion about the capacity of state education departments to turn around low-performing schools—whether that be 5, 15 or 50 percent of them—but there's also a need to think more fundamentally about which approaches are right for which schools and what entities should be charged with overseeing change. State education departments, districts, business communities, parents, nonprofits, for-profits, and other entities likely all have a role in helping struggling schools, students, and adults. But very few states are recruiting and equipping multiple partners in the challenge.

[20] Again, "sanctions" implies the use of the stick, not the carrot. We do not include here state efforts to offer supports to schools.