measuring up: recalibrating systems of assessment
FOR THREE OPPORTUNITIES TO INCREASE THE IMPACT OF YOUR WORK ON STATE BOARDS OF EDUCATION

save the dates!

legislative conference
April 4–5, 2016
Liaison Capitol Hill, Washington, DC

new member institute
July 21–23, 2016
Ritz-Carlton, Pentagon City, Arlington, VA

annual conference
October 19–22, 2016
Sheraton Kansas City Hotel at Crown Center, Kansas City, MO
features

6 Not as Easy as It Sounds: Designing a Balanced Assessment System
Having learned that one type of assessment cannot do it all, states must figure out how to fit many pieces together to make a complete picture.
Rajendra Chattergoon and Scott Marion

10 Accountability for What Matters
Data dashboards can help states and districts better track the indicators and outcomes they deem most important without drowning in data.
Robert Rothman

14 Aligning the K-12 Academic Journey with a Postsecondary Destination
States can track how well their students are progressing toward career readiness, and North Carolina shows how it’s done.
Malbert Smith III and Gary L. Williamson

19 Leveraging Principal Evaluations to Scale Up Instructional Improvement
States and districts can get more bang for the buck by giving serious attention to developing great principals.
Benjamin Fenton

26 Georgia’s Teacher Performance Assessment
A focus on assessing new teacher candidates on classroom performance goes statewide.
Anne Marie Fenton and Pamela Wetherington

30 New Hampshire’s Accountability 3.0
How New Hampshire is overhauling its system for school accountability.
Paul Leather

34 Making Sense of New Science Assessments
New state science standards that integrate core ideas, cross-cutting concepts, and practices cry out for assessments that do likewise.
James W. Pellegrino

departments

2 editor’s note
3 news & notes
4 we, the media
5 NCOSEA voice
40 the NASBE interview: Linda Darling-Hammond
44 from the director’s desktop

NASBE
National Association of State Boards of Education

www.nasbe.org
NASBE folk have been singing their own rendition of “A Change Is Gonna Come” as regards education reform for a while now, as well as a K–12 version of “Are You Ready?” Given the policy reset reflected in the newly minted Every Student Succeeds Act, NASBE is pleased to deliver a timely issue that focuses on one thing the change means for state boards of education: greater responsibility for assessments and determining their role in state accountability systems. The authors argue for states to take a fresh, vigorous look at how all the pieces fit together.

For your consideration as you navigate these waters, Rajendra Chattergoon and Scott Marion describe the qualities of a coherent system of assessments and how you will know if you have one. The Alliance for Excellent Education’s Robert Rothman writes on how the potential muddle of multiple measures can be sorted with data dashboards. And New Hampshire’s deputy commissioner of education, Paul Leather, describes the goals and process by which his state is redesigning its accountability system. Performance assessments play a leading role.

This edition deals not only with student and school performance. Ben Fenton at New Leaders makes the case for robust implementation of principal evaluations to help school leaders develop their skills, and Anne Marie Fenton and Pamela Wetherington sketch Georgia’s path toward statewide adoption of a performance assessment for new teacher candidates.

Another NASBE refrain that may sound familiar is our call for state boards to ensure that statewide tests align with statewide standards. While there are aligned tests to choose from for Common Core adopters to assess English language arts and math skills, states that recently adopted Next Generation Science Standards have only just begun to think how assessments might connect to them. Jim Pellegrino details what states should look for in aligned science assessments.

Although states have embraced the challenge of preparing their students for college and careers, it’s fair to say that many would be hard pressed to say whether their students are displaying career readiness. MetaMetrics’ Malbert Smith III and Gary L. Williamson show that such assessment is possible, and, as a case in point, they show how North Carolina has married learning progressions in K–12 reading against distinct measures of text complexity that reflect what careers and colleges will require of new entrants.

We round out the issue with an interview with Linda Darling-Hammond on the benefits of student performance assessments and how state boards might think about incorporating them into their assessment systems.

You are keenly aware of how your communities react to assessments, particularly if they appear to them unconnected from what they expect of their schools and students. These authors make the case that a well-knitted system can assure stakeholders that what is measured matters—and that educators and policymakers alike remain focused on what matters most in helping all students achieve.
On December 10, 2015, President Obama signed the Every Student Succeeds Act (ESSA), the latest reauthorization of the Elementary and Secondary Education Act. This bipartisan bill makes significant changes in education governance that will dramatically affect the work of state boards of education. Here are five things you need to know about ESSA:

- **Testing will continue.** ESSA continues the requirement that states test all students annually in reading and math from grades 3 through 8 and that students are tested once in high school. As under No Child Left Behind (NCLB), results must be broken out into subgroups—English-language learners, students in special education, racial minorities, students living in poverty. The law will also require identification of students with a parent who is a member of active-duty armed forces so the progress of these students, who typically move more frequently than their peers, can be adequately measured.

- **ESSA does not prescribe a certain set of college-ready standards but requires states to establish “challenging” academic content standards in mathematics, reading or language arts, and science for all students. These standards must be aligned with real-world requirements in college and careers.**

- **States will get much more discretion and authority in designing their state accountability systems.** They still must set goals, but they have greater flexibility. Goals must address proficiency on tests, English-language proficiency, and graduation rates. And the goals still must set an expectation that groups that are furthest behind close gaps in achievement and graduation rates. But Annual Yearly Progress (AYP) is gone. States must also include at least one additional student-performance measure as well as at least one measure related to school success, such as student engagement or school climate.

- **States must identify and intervene in the lowest performing schools** (the bottom 5 percent on their accountability system) as well as schools where fewer than two-thirds of students graduate. The state can design the intervention, however.

- **States have more flexibility in evaluating teachers.** Gone is NCLB’s Highly Qualified Teacher requirement. So is the requirement that states link teacher evaluation to test scores, although states may do so if they want.

As recent headlines bear out, students of color and students with disabilities are more likely to be arrested or referred to law enforcement for nonviolent offenses compared with white peers (see figure). This national trend disrupts learning opportunities for many students of color and disabled students by criminalizing minor offenses. Ways that policymakers and educators can promote nondiscriminatory policies and practices are detailed in a recent NASBE report, “Advancing School Discipline Reform.” According to the American Institutes for Research, only a handful of states require training for school-based law enforcement. Such training could usefully cover child and youth development, implicit bias, trauma, and restorative justice.
Let’s talk numbers.

204: The total number of people who attended NASBE’s Annual Conference 2015 in Baltimore.

860: The number of tweets using the NASBE Annual Conference hashtag #NASBE15.

75: The number of people live tweeting NASBE’s Annual Conference 2015.

These numbers show tremendous growth. This year’s conference drew double the number of people who came to Denver the previous year, and the flurry of activity on social media during the conference (even though the numbers are an approximation using free analytics tools) indicates that attendees were deeply engaged.

Now consider these numbers:

68: The number of retweets NASBE receives for every 100 tweets.

12 million: The total impressions—that is, how often NASBE tweets showed up in Twitter followers’ streams in the past year.

I gathered that last data point from a digital media agency, Social Driver, which did a complimentary audit of NASBE’s social media efforts. I was blown away. How could one tweet have such potential?

Emily Rasowsky, Social Driver’s director of marketing strategy, explained that a tweet can have a big impact when influential people share it. In our case, this happened when Indiana State Board Member Byron Ernest retweeted this NASBE tweet during the conference:

Ernest, who just joined NASBE’s board of directors as its new member representative, has over 280,000 Twitter followers—more than any other person tweeting about #NASBE15 during the conference. Ernest’s followers alone, not to mention those of other live tweeters such as NASBE President Mary Lord and Massachusetts’s Mary Ann Stewart, gave NASBE’s tweet 25 times the exposure it would otherwise have had. That’s huge for an organization with just under 11,000 Twitter followers!

Emily had a few other ideas. If just a handful of NASBE members had the potential to reach millions with a single tweet or two, imagine what the combined efforts of the entire membership could accomplish? Reaching a sizable audience is only a starting point. Engagement—when followers interact with your content—is where real impact lies. The Every Student Succeeds Act returns great authority to the states, and when it takes effect, state boards of education will be even more essential. It will be important to help the education community understand who state boards are and what they do. We must share state success stories and engage the public in state policymaking. Social media can be a great way to leverage NASBE’s combined influence as a membership organization.

In the coming months, NASBE will be taking a close look at our membership to identify social media superstars: power users like Byron Ernest and Mary Ann Stewart, who have a strong following and great potential to help raise NASBE’s national voice and share our unique viewpoint with an expanded audience.

NASBE has such an active, inspired membership! I look forward to tapping its potential.
This issue of the State Education Standard provides a wealth of information for state boards of education and other education leaders to consider in designing systems for assessment. Here, I provide the common legal framework that courts will use in reviewing decisions about assessment. As a general rule, state boards have latitude—both in selection of assessment instruments and in using assessment data to inform their leadership decisions. Reviewing courts will not disturb a board’s decision regarding assessment unless the decision is “arbitrary, capricious, or unreasonable.” Chappell v. Commissioner of Education, 343 A.2d 811, 814 (N.J. Super. Ct. App. Div. 1975). A decision will be upheld unless no reasonable person could find it to be correct, based on the facts and the law. If reasonable minds could differ about assessment policy, a court will not substitute its judgment for the board’s.

In decisions about whether to conduct a statewide assessment in the first place, a state board has considerable latitude unless the law requires assessment of a particular content matter or grade level. For example, the Elementary and Secondary Education Act (ESEA) requires assessments of students in certain grades and subjects. While a board’s decision not to assess a required grade and subject would certainly be unreasonable, its decision about whether to add grades (first grade) or subjects (social studies) would be left undisturbed unless entirely without legal or factual foundation.

In decisions about what assessment instruments to use, courts will defer to state boards’ substantive judgments. This accords with the long-standing principle of administrative law that requires deference to matters within an agency’s specialized expertise. See Porter v. Seattle School District No. 1, 248 P.3d 1111 (Wash. Ct. App. 2011), which held that adoption of a mathematics textbook was not an abuse of discretion. State boards, not courts, are the “experts” on assessment. A court would be extremely hesitant to invalidate the selection of an assessment on substantive grounds—but might do so on nonsubstantive grounds, such as violations of state bidding law.

Once the state board has assessment data—about the state, a subgroup, a school district, or a school building—it has latitude in how to respond. Some responses may be required (such as by ESEA), but most responses are within the board’s discretion. Boards must be wary, however, of failing to act when data would suggest that a reasonable person would take some action. For example, at some point, statewide declining achievement scores in math would suggest some board action would be reasonable, such as by engaging in a root cause analysis about the decline, by providing additional statewide professional development in mathematics, or by raising teacher license requirements in mathematics. Failure to take some action when a reasonable person would do so may be an abuse of discretion. See, e.g., Litterer v. Judge, 644 N.W.2d 357 (Iowa 2002).

In sum, boards have significant latitude regarding assessment. When it is used wisely, courts will defer to boards. More important, good decisions will be made for students and educators.
Many states and school districts are rethinking how they do educational assessment. A few are going further: attempting to build “balanced,” “comprehensive,” or “next-generation” assessment systems. At the same time, practitioners and researchers have long mulled the purposes such systems should have. But just a few have thought deeply about what a system of assessment actually means. In the words of educational psychologist Theodore Coladarci: “A collection of assessments does not entail a system any more than a pile of bricks constitutes a house.”¹ This article focuses on building the house.

For states that are building, evaluating, or modifying their assessment systems, we offer three criteria to look for: coherence, a theory of action, and efficiency. We developed these criteria after reviewing literature on educational assessment systems, and we have been applying them in work with state and district leaders.

Coherence

There are two aspects to consider when evaluating coherence within an assessment system.² First, the assessments in a system must be compatible with the models of how students learn content and skills over time. Second, curriculum, instruction, and assessment must be aligned to ensure that the entire system is working toward a common set of learning goals.
Underpinning both of these features of coherence is a well-articulated theory of how students learn. Several researchers have written about these two dimensions of coherence. However, most have only examined coherence among a relatively small group of classroom assessments and not within a state assessment system.

Models of student learning describe how knowledge and skills progress as students are exposed to instruction and other learning opportunities. They build upon content standards by describing a sequence of milestones that students should reach as they progress toward a learning target. Some states have started to adopt documents that describe progressions of student understanding of a topic across grades. The development and validation of these learning progressions often require rigorous research into student cognition and assessment. As such, it may be challenging for states to draft their own models of student learning from scratch. It might be more promising for states to collaborate with one another and with researchers to identify promising models that may then be adopted statewide. Evaluations of state assessments should identify and evaluate the appropriateness of the models of student learning that a state endorses.

A major advantage of having well-articulated models of student learning is that such models can inform teachers’ formative assessment practices. And when curriculum, instruction, and assessment align in this way, the distinction between an assessment system and a classroom-level instructional system evaporates. One way to get inside the “black box” is to adopt or develop classroom observation protocols that are focused on holistically evaluating this alignment.

One example of such a protocol is the Danielson Framework. A teacher’s curriculum and its connection to models of student learning are evaluated as part of a “planning and preparation” domain. The domains of “classroom environment” and “instruction” address the quality of a teacher’s interactions with students in a classroom. Although there is no separate domain for evaluating teachers’ assessment practices or assessment literacy, some of the actions ascribed to strong formative assessment practices are evaluated in subsections. When coupled with an evaluation of the appropriateness of the models of student learning that a state uses, the results from teacher observations can provide a gauge of the system’s coherence.

Both aspects of coherence are grounded in a theory of how students learn. While the research community has over the years engaged in lively debate about the merits of different theories, there is emerging consensus that learning happens through cognitive mechanisms within a social context. This consensus matters, because different models of student learning point to different assessment tasks. A balanced assessment system should include a range of cognitively complex tasks. Consider, for example, a state assessment system in which a majority of tasks only require recall of basic facts. Such an assessment system cannot tap into students’ abilities to learn information in collaborative or social contexts.

Theory of Action

The second criterion for a high-quality assessment system is that a well-articulated theory of action guides interaction among its component parts. Such a theory of action includes the following:

- identification of an overall purpose for the system;
- identification of the stakeholders and what they need from assessments;
- a map of how assessments are being used to meet stakeholder needs and are valid for intended purposes;
- specification of the mechanisms (or processes) by which the assessments are designed to work and what needs to happen in order for the purposes to be realized; and
- evaluations of the extent to which stakeholders’ needs are met.

A set of assessments, even if they cohere, will not fulfill the intended purposes if the information never reaches the intended user. Multiple sources of assessment information should be integrated as part of the assessment system design—not after the fact.

By developing a theory of action for its assessment system, a state defines its overall purpose and a plan for achieving that purpose. We work with a state that is piloting a new accountability system that incorporates local and common
An assessment system has to provide users with a clear picture of what students know and are able to do. It can’t do so if parts of the system are not working together efficiently. Assessment efficiency means getting the most out of assessment resources and eliminating redundant, unused, and untimely assessments. Attention to efficiency enables each assessment to do what it is designed to do. For example, if a state wants to give educators information to help them adjust instruction, its assessments must be tied to the curriculum that is being used. These assessments should in turn yield timely, detailed information about the knowledge and skills being assessed at the local level.

Another partner of ours, a large school district in the southern United States, is seeking to streamline its assessments. Teachers in the district must administer more than a dozen assessments throughout the academic year, and many of them multiple times each year. For example, during October, high schools must administer a college entrance exam (e.g., the PSAT, SAT, or ACT) and an interim assessment, alongside any local assessments that teachers or teams of teachers develop. Recognizing that this structure was unwieldy, the district has begun initial planning and evaluation to make the assessment program more efficient.

States need to focus on assessment efficiency to accomplish a couple of things:

- Efficiency

- Figure 1. Hypothetical Information Flow for an Assessment System
Identify and reduce assessments that do not cohere with the state’s models of student learning and that are not mandated. 10 Carefully defining the intended purposes and uses for each assessment in the system and what will be required for each assessment to be able to fulfill the intended purposes will be a prerequisite for this task.

Define the minimum number of assessments necessary to meet the information needs of all stakeholders.

Although this district has not formally endorsed a specific model of student learning, the state has provided some limited resources to help districts think about learning progressions. In the absence of clear models, it is difficult for a district to decide which assessments to shed. However, one way to increase efficiency is to evaluate how the information generated from assessments is used. For example, how does the district use information from the October interim assessments? Can this information be obtained from other assessments within the system?

Advancing Assessment Reform

We have described three characteristics of good assessment systems to aid states in modernizing their own assessments. Even though we have described them discretely, successful reform movements will likely be attending to all three concurrently. Coherence is the kernel that should be present for any reform to succeed. Without coherence, it does not make sense to speak of pathways in which information flows or an efficient set of assessments.

A few words of caution follow. An assessment system inherently serves multiple purposes: accountability, evaluation, instructional improvement, and monitoring, for example. However, meeting all these purposes requires thoughtful planning about which data will be preserved at each level. For example, what types of assessment data will be used for accountability and which will be reserved for, say, evaluation? Many states and school districts use large-scale summative tests for accountability and other assessments for the other purposes. Given the need for comparability across districts in making state policy decisions, this allocation makes sense. However, serious unintended, negative consequences can ensue from layering too many requirements onto the state-mandated summative test. There is a long, unhappy history of accountability tests driving and narrowing the curriculum as educators focus their efforts on the tests that get counted. 11 More important, statewide summative tests may create an unnecessary divide between state and local assessments. As states revisit their assessment programs, they would do well to think about how changes will affect districts, schools, and classrooms and how coherent, balanced assessment systems could best be supported.


12We draw on the definition of coherence in an assessment system provided by the National Research Council, Knowing What Students Know: The Science and Design of Educational Assessment (Washington, DC: National Academies Press, 2001).


14For example, the Department of Public Instruction in North Carolina is working on providing learning progressions for Common Core English language arts (http://elaccss.ncdpi.wikispaces.net/) and math (http://maccss.ncdpi.wikispaces.net/Home).

15North Carolina’s Office of Early Learning is using “construct progressions” as part of the professional development in formative assessment that they are providing to school districts. For details about this project, see http://r5k3formativeassessmentsupport.ncdpi.wikispaces.net/.

16For a description of Charlotte Danielson’s Framework, see https://danielsongroup.org/framework/.


20Organizations are developing resources to help inventory state and district assessment systems. For example, Achieve has developed an “assessment inventory” for school districts: http://www.achieve.org/assessmentinventory.

For more than a decade, states have evaluated school performance largely through a single measure—test scores—and rated schools on whether they improved students' performance in reading or math. The idea was to focus schools' attention on the outcomes that mattered most and to focus states' attention on the schools that needed the most help in raising those outcomes.

Recently, though, that idea has come into question, and measures of school performance are getting a new look. With their newfound authority under the Every Student Succeeds Act (ESSA), states will be using a broader range of indicators of school performance and displaying them in ways that give school communities, parents, and district and state officials a clearer picture of how a school actually is doing.

These new systems, often called “data dashboards,” function the way a car's dashboard does—by displaying multiple measures that affect how a school is performing. This collection of information enables educators to monitor their school's overall performance while simultaneously identifying specific problems, allowing them to focus resources toward individual needs and address all of the issues that affect school performance. Just as a driver will fill up his tank before the gas gauge turns to “E” to keep the car operating effectively, a school using a data dashboard can work on improving, say, school climate before it affects how well students learn.

At the same time, states are using dashboards to shift the focus of accountability from zooming in on low performance to developing a system of continuous improvement. A driver would never look at the E on a gas gauge and say that his car is failing.

State-level dashboards are relatively new, but education systems in other countries as well as a number of local US school districts have used dashboard-type systems for some time. This article will describe some of the issues involved in establishing dashboard systems, using examples from existing systems.

Although states have held schools accountable for student performance for decades, No Child Left Behind (NCLB) established a uniform system of accountability for the first time. It required states to test all students in grades three through eight and once in high school in reading and mathematics and to set standards that would indicate “proficiency” on the tests. The law also required states to set targets for test performance so that all students would reach proficiency by 2014. Schools that failed to meet annual targets—either for the school as a whole or for subgroups of students within the school—would be designated “in need of improvement” and subject to an escalating set of sanctions and interventions.

The law also authorized states to monitor additional measures of school performance (usually graduation rates for high schools), but performance on these measures could not compensate for low performance on tests.

According to most observers, NCLB successfully focused attention on previously underserved student subgroups, and student
achievement has risen somewhat since its passage in 2001, particularly in elementary school mathematics. But high school performance has remained flat, and relatively few students have been able to demonstrate the deeper learning competencies—using knowledge to solve novel problems, thinking critically, communicating effectively, and collaborating with peers—that are necessary for success in college and careers. Moreover, NCLB produced some unintended consequences, such as an overemphasis on tests. While the tests were tied to high stakes in order to drive instruction, they also failed to assess or encourage some of the most important skills students need.

The shortcomings of the system that NCLB built led state officials and advocates to consider new approaches to accountability, opening the door to ideas like data dashboards. In 2011, the US Department of Education began issuing waivers to states to allow them to develop their own accountability systems. Some 43 states received waivers, and their new systems vary in many ways from the NCLB model. For example, some states, such as Kentucky, broadened their measures of school performance significantly. In addition to test scores, Kentucky includes measures of teacher performance in its measure of school performance.

And California, which has not received a waiver, instituted a new accountability system for all districts. Under a state law that altered how the state funds schools, California required districts to develop their own accountability systems in addition to the state system. The law required districts to measure school performance along eight dimensions: student achievement, student engagement, college and career readiness, school climate, parent involvement, basic services (such as access to materials and adequate facilities), implementation of state standards, and access to rigorous coursework.

Under ESSA, states have more leeway on which indicators to incorporate, but they can’t use a single measure. They have more flexibility to decide how teachers will be evaluated and how to address low-performing schools. Regardless of the particulars of a state’s plan, there will be a fair amount of data for states to assess.

Why Dashboards?

Current state accountability systems have a number of advantages. They focus attention on outcomes and measure whether schools have achieved learning goals. They allow parents and policymakers to compare schools to one another easily. And they can be administered with available data systems.

However, research has revealed a number of shortcomings in the systems. For example, states that use only test scores to measure school performance create incentives to focus on the relatively low levels of performance the tests tap. Alternatively, states that combine multiple indicators into an index or letter grade mask low performance on individual indicators.

In addition, the accountability systems tend to provide limited guidance to schools and districts. They might show that a school is low performing but then fail to tell educators what to do about it. Moreover, they tend to rely solely on state measures, such as tests and graduation rates, that are calculated too late to allow schools to address problems before they worsen.

Data dashboards can alleviate some of the problems with current systems. They provide an array of indicators in a transparent way so that school staff, parents, and public officials know how schools are performing on all important measures. They also can include measures of school operations and practices so that schools know what to address if performance is low. And they can allow districts to add “real-time” measures to supplement the state data and enable local leaders to monitor performance throughout the year.

Toward a Data Dashboard

Creating a data dashboard is not as simple as collecting and displaying all available information. Districts that have used the systems effectively suggest there are at least four important considerations that must go into the development of the systems:

Choosing the right indicators. Designers of dashboard systems believe that, in order to drive improvements in performance, the indicators in a dashboard need to reflect the most significant measures of a school’s performance. To that end, they rely on what they consider the most critical outcome measures, as well as research on the factors that contribute to high performance.

“It helps focus attention,” says Maggie Glennon, a former assistant superintendent in Monroe County, Georgia, who is now a consultant with the Georgia Leadership Institute for School Improvement. “You have to prioritize. You can’t do everything.”
Many dashboards group their indicators into a handful of categories, which show clearly the areas of focus. In many cases, these categories include so-called inputs, as well as traditional student outcomes. For example, Monroe County’s “balanced scorecard” provides indicators in four categories: student learning outcomes, organizational effectiveness, public engagement, and professional learning. Altogether, the scorecard comprises 70 measures of performance, but grouping them in these categories enables school and district leaders to see areas they need to address that might result in lower student performance down the road, according to Glennon. “You can't improve student learning if the classrooms are dirty and teachers don't get their paychecks on time,” she says. “If teachers are not happy, they will not do their best job teaching.”

Similarly, school reports in Alberta, Canada, provide measures in six categories: safe and caring schools; student learning opportunities; student learning achievement; preparation for lifelong learning, world of work, and citizenship; parental involvement; and continuous improvement. The indicators within these categories include a mix of outcome measures (like test scores and graduation rates) and survey results (e.g., teacher and parent satisfaction with parental involvement in decisions about their children’s education).

School districts in California include resource measures as well in their local accountability plans. For example, the reports include measures of basic services, such as access to materials and adequate facilities, and access to coursework needed for college and career readiness. In that way, districts hold themselves accountable for providing the resources necessary for schools to succeed.

The measures of student performance on dashboards tend to be broader than test scores alone. For example, the districts in California’s CORE (California Office to Reform Education) network, a group of 10 large districts that won a waiver from NCLB, include measures of social and emotional well-being in their accountability system. This reflects the growing body of research on the importance of student motivation and academic mind-sets in student learning, according to officials from the network.

Indicators in that category include results from surveys on school climate, suspension and expulsion rates, and chronic absenteeism.

Districts such as Dallas Independent School District, the School District of Philadelphia, Pittsburgh Public Schools, San Jose Unified, and New Visions schools in New York City, meanwhile, have participated in the College Ready Indicator Systems (CRIS) project to develop and report indicators of college readiness on three dimensions, based on a framework developed by David Conley, the former director of the Educational Policy Improvement Center and an expert on college readiness. The dimensions include academic readiness, academic tenacity, and “college knowledge,” or an understanding of how to navigate the admissions process and make a smooth transition to higher education. The districts collect data on these dimensions at the student, school, and district levels.

**Identifying the most critical indicators for special scrutiny.** While the dashboard systems include an array of indicators, administrators often focus on a few key indicators to ensure that schools continue to monitor them and address the most critical problems in school performance. For example, Monroe County Schools monitors attendance each year, even if schools continue to perform highly on that indicator, so that it won't slip off the radar, according to district officials.

San Jose's OPSTAT system (for “opportunity statistics”) measures school performance on 11 indicators, but individual schools are particularly responsible for four key indicators—one for elementary schools (third-grade reading), one for middle schools (eighth-grade mathematics), and two for high schools (Advanced Placement and International Baccalaureate performance and completion of coursework required for admission into the University of California system). Regardless of school performance on other indicators, schools must develop plans for improving results on those measures, according to Jason Willis, assistant superintendent for accountability and community engagement. These indicators encourage broader access to high-quality learning opportunities as well as tracking how students are doing when they receive these opportunities.

**Setting appropriate targets or benchmarks.** In addition to presenting information on the current status of school performance, data dashboards also contribute to school improvement by showing how well schools raise their performance over time. In order to show progress, leaders set targets for improvement and hold
themselves accountable for meeting the targets. 

In Monroe County, the scorecard’s color-coding scheme is based on whether schools meet targets in each of the categories. When schools meet or exceed a target, the indicator is marked in green. If schools are close to a target, it is marked in yellow. If schools fall short, it is marked in red.

Target setting is based on negotiations between principals and district leaders. District officials want to set ambitious but reasonable targets, according to Glennon. The district wants schools to show steady gains but does not want to set targets that are so out of reach that teachers and principals grow discouraged. So, for example, the district seldom sets a 100 percent target in any category because leaders recognize that any student could have a bad test day and cause the school to miss that goal. On the other hand, district leaders insist that schools raise their targets if a principal’s initial proposal is too low.

Using the tools for improvement. In addition to presenting data in a different way, districts using dashboard systems have transformed the way they approach accountability. Rather than simply present data or threaten intervention, district officials work with schools to improve performance.

Philadelphia’s school report cards show how dashboards can contribute to improvement. The school report card scores each category of indicator (achievement, progress, climate, and college and career readiness) at one of four levels: model, reinforce, watch, and intervene. The schools with “model” levels of performance are showcased throughout the city. But the district provides assistance only to the schools at the “intervene” level and does so only to address the aspects of the school that appear problematic. The district and the schools monitor performance at the “reinforce” and “watch” levels.

San Jose also shifted from accountability as finger pointing to accountability for improvement, according to Willis. Since the district shifted to its OPSTAT system in 2013, principals regard it much more favorably, he says. Principals and district leaders develop improvement plans together based on the data and report regularly on progress on the plans.

“It’s the polar opposite of the culture of mistrust” that characterized the previous system, Willis says. “It’s not intended to fire principals or label schools. There is an ongoing dialogue.”

A Comprehensive System

States are entering a new era in accountability. Under ESSA, states will be experimenting with systems aimed at providing information to schools and communities that can help them improve performance for all schools. This information is particularly important for state boards of education. They need valid, reliable information on how states are progressing to determine if policy shifts are warranted. And boards are in an important position to gather public input on the measures states should be held accountable for.

As they revise their systems, states are showing a growing interest in data dashboards. However, those are not the only elements of new systems. States also are creating school quality review processes, in which educators examine a school’s practices in depth to get a better picture of teaching and learning and recommend improvements. Such systems are already in place in a number of states, including Rhode Island and Georgia.

In addition, states are developing and adopting performance assessments to supplement state tests. These assessments provide students and teachers with a broader view of student performance by tapping abilities end-of-year tests seldom measure, such as the ability to write extended essays in a variety of genres or solve complex, multistep mathematics problems. The assessments also support instruction by providing information on student performance during the year—while teachers can act on it—instead of waiting until the end of the year.

States also are creating professional growth systems and tying them to their accountability systems by connecting teachers and school leaders who are struggling in particular areas with other educators who have demonstrated success in those areas. The CORE districts in California suggest one way this can work. Under their plans, if a school dashboard suggests that a school is having difficulty with the reading performance of its English learners for instance, the district will link that school’s teachers with those from another school that performs well in that area.

At best, these elements should work in tandem. Together, they can transform an accountability system from one aimed at singling out low-performance schools to one that supports continuous improvement for all schools.⁴


⁴David Conley, Redefining College Readiness (Eugene, OR: Education Policy Improvement Center, 2007).
How will state boards of education know their students are building the skills needed for careers?

Although debate about the adoption and implementation of the Common Core State Standards (CCSS) has been extensive, there is unanimity on the importance of college and career readiness standards. States that never adopted the CCSS—such as Texas, Virginia, Alaska, and Nebraska—have fully embraced the notion that standards must prepare students for the real-world demands of college and careers. Even states that are reconsidering their initial acceptance of the CCSS have not backed away from this overarching goal.

Evidence of the growing interest in college and career readiness abounds. According to Google Trends, web searches on this topic have increased dramatically each year since 2010 (figure 1). Likewise, top literacy experts rate “college and career readiness” as one of the most important goals in education.1

An abundance of research laid the foundation for this consensus by showing that too many students graduate ill prepared for the rigors of college or the workplace.2 Researchers at Stanford framed the problem starkly:

[H]igh school assessments often stress different knowledge and skills than do college entrance and placement requirements. Similarly, the coursework between high school and college is not connected; students graduate from high school under one set of standards and, 3 months later, are required to meet a whole new set of standards in college. Current data systems are not equipped to address students’ needs across systems, and no one is held accountable for issues related to student transitions from high school to college.3

Consider reading skills, for example. When one compares measures of text complexity, there are statistically significant gaps between high school texts and texts required in US community colleges, universities, and the

Aligning the K-12 Academic Journey with a Postsecondary Destination

by Malbert Smith III and Gary L. Williamson
workplace: Quantified in Lexile® measures, the gaps ranged from 130L for the workplace to 265L in the university. If statewide assessments are not aligned to these workplace and college requirements of reading and comprehending more complex text, then states will keep promoting a false sense of security that US high school graduates are college and career ready when in fact they are not. No Child Left Behind—which allowed every state to develop its own assessment and set its own proficiency standards—resulted in a psychometric tower of Babel, creating major confusion and public doubt when state proficiency levels differed markedly from those revealed in the National Assessment on Educational Progress, or NAEP. In addition, this illusion of proficiency “was shattered when ‘proficient’ kids got to college and needed remediation, or entered the workforce with substandard skills.” In response, state boards of education have increasingly raised their proficiency cut scores to be more aligned to NAEP.

What more can a state board member do to assure the public that the state assessment program is measuring college and career readiness? Can this program yield data on whether a student has reached a level of reading commensurate with the demands of college and careers? Additionally, does it provide grade-level analysis that tracks a growth trajectory toward that goal?

**Begin with the End in Mind**

In his 1989 classic *The Seven Habits of Highly Effective People*, Stephen Covey exhorts readers “to begin with the end in mind.” When it comes to literacy, the “end in mind” is the ability to comprehend texts that one is required to read at work or college. Alignment with this goal is the litmus test for the assessment of reading in K-12.

Fortunately, states now have access to more sophisticated measurement models that permit the measurement of growth over the developmental lifespan of the learner and across educational institutions. A number of states use The Lexile® Framework for Reading, which places reading ability and text complexity on the Lexile scale. These states can report Lexile reading measures for their students and examine their growth in relationship to the empirical reading demands expressed in Lexile text measures.

**Reading Trajectories in North Carolina**

North Carolina is one such state. It has aligned its assessments to the “endpoint,” raised standards and performance, and built student growth trajectories. North Carolina has long focused on documenting student growth, incorporating that growth into its accountability system, improving its longitudinal data systems, and using multiple analytical strategies. This history made it possible for the state to create developmental growth curves that illustrate the long-term progression of student performance. As their measurement systems and longitudinal data collections have matured, North Carolina has extended its understanding of growth from the simple year-to-year gains it employed 20 years ago, to growth curves spanning grades three through eight (in efforts that began around 10 years ago), to descriptions spanning grades 3 through 11 (in its most recent efforts). We can use these data to juxtapose the relationships between academic standards and student growth (figure 2).
Figure 2 shows two North Carolina reading growth curves that represent different student cohorts; within each cohort, the same students were measured on successive occasions. The lower curve spans grades 3 through 8 during 1995 to 2000, and it summarizes the growth of 48,790 students who had complete data across all six grades. These students constitute 53.6 percent of the students in the cohort group—a respectable match rate, given the decentralized nature of the student information systems and the absence of a uniform student ID in those days. The higher curve spans grades 3 through 11 from 2006 to 2014 and represents 103,334 students, 98.0 percent of the cohort in the state (including those with partial data). This impressive data collection benefited from the implementation of a uniform student ID system and increased attention to constructing longitudinal data collections designed for easy analysis.

Three results are immediately apparent. First, on average, North Carolina reading growth was substantial and consistent in both cohorts, with similar arcs rising rapidly from left to right. Second, student performance increased systematically across all grades and substantially in absolute terms—that is, the more recent curve is significantly higher on the scale than the earlier curve. This is especially notable because the more recent curve is based on a more inclusive and representative sample. Third, North Carolina can now measure student growth over longer periods. The more recent curve incorporates a higher percentage of student records and extends well into the high school years.

**Student Performance Rises to Meet Standards**

Over 20 years, North Carolina implemented several editions of assessments as its curriculum standards were revised. With each edition, the state board of education (SBE) raised standards. Proficiency cut scores are depicted in figure 2 as small bars arrayed in columns at grades three through eight. For the first three editions of its end-of-grade (EOG) tests, North Carolina had four achievement levels; Level III was deemed to represent grade-level performance. In its fourth edition of tests, called NC READY, the state defined five levels of performance, with the two highest being “sufficient” and “solid.” The Level III cut points are shown for the first and third editions of EOG tests; the sufficient and solid cut points are shown for NC READY. Because all the cut scores have been translated onto a
common scale, one can readily see that the SBE dramatically raised reading standards over the four editions of assessments.

With student growth and state achievement-level standards expressed on a common scale, it is possible to compare student aggregate performance directly to achievement-level standards without resorting to a relative scale (e.g., percent proficient). When one examines the cut points in relation to the student growth curves, it becomes apparent that student growth and performance rose to meet the higher expectations that the SBE set.

**Text Complexity Standards**

Over the last five years, the National Governors Association's Center for Best Practices and the Council of Chief State School Officers have helped states collectively arrive at general agreement regarding text complexity levels that are aligned with college and career readiness. The dashed lines in figure 2 represent text complexity standards for grades 3 to 12.

Clearly, North Carolina's achievement-level standards for the first edition of EOG tests were not well aligned with the text complexity ranges (which of course were defined years later). However, the SBE's efforts over time to raise achievement-level standards brought their achievement levels for subsequent editions of tests into very good alignment with the text complexity requirements. Concomitantly, North Carolina student achievement improved so that student growth on average now tracks well with college and career reading demands.

**Postsecondary Reading Demands**

Our research on the difficulty of postsecondary reading materials in colleges, universities, the workplace, the military, and other citizenship endeavors has revealed that the typical level for postsecondary reading materials is 1300L. Additionally, the middle 50 percent of sampled reading materials in our studies ranged from 1200L to 1380L. A yellow box at grade 13 depicts these results, with a diamond at 1300L (figure 2). This convention summarizes the reading demand of a wide variety of postsecondary materials, many of which directly relate to postsecondary educational experiences and workplace reading requirements.

Two things stand out. First, text complexity standards adopted by many states align very well with the postsecondary reading materials at the boundary between grades 12 and 13. Second, expressing all these measures on a common scale shows clearly that North Carolina student growth on average aligns well with postsecondary reading requirements.

**Career Text Requirements**

We have determined the typical reading difficulty of samples of reading material from a variety of "Bright Outlook" occupations, which got this designation because they are characterized by rapid growth, numerous job openings, or are newly emerging occupations. Median text complexity measures for selected career text collections are plotted in figure 2 by the number of years of required education associated with career entry. The "grade" where these measures are plotted represents years of required education (e.g., grade 12 represents that a high school diploma is required for career entry; 14 means that two additional years of education are required beyond high school—namely, an associate's degree; 16 corresponds to a bachelor's degree; 18 to a master's; and 20 to a doctorate).

Several findings are notable. Generally, a systematic increase in career reading demand accompanies increases in years of required education. At each level of education (i.e., grade), there is variability in reading demand across careers. Many careers demand comprehension of reading materials that are more complex than typical postsecondary reading materials: Note how many dots lie above the median postsecondary reading materials represented by the diamond plotted within the box at grade 13. The single dot at grade 12 represents the reading demand (1130L) associated with careers that require only a high school diploma. It represents the lowest level of reading ability required by any career examined to date.

Increasingly, we have been able to populate figure 2 with measures of text requirements and performance standards related to college and career readiness. However, the measurement of student growth and performance tends to cease around grade 11 in high school, as indicated by the fact that the growth curves do not extend beyond that point. A provocative question: How
can we extend measurement of student ability into the postsecondary years to confirm that a well-aligned journey in fact leads to success in the many possible endeavors that students may choose?

**Key Considerations for State Boards**

Figure 2 demonstrates how one can longitudinally chart the progression of students’ reading ability and whether their trajectory is one of college and career readiness. The two student growth curves represent the average scores of 48,790 and 103,334 North Carolina students respectively. While for purposes of overall state-wide performance this aggregation is fine, it is also possible to build individual growth trajectories. For example, it is possible for students and parents to tell whether their individual growth trajectories align to their postsecondary education goals.

Other states can glean a key lesson from North Carolina’s example regarding three fundamental factors for discerning their students’ career readiness: data systems that permit longitudinal analyses, a common scale for student reading ability and text complexity (e.g., the Lexile scale), and empirical, objective measures for documenting the reading requirements in careers and postsecondary education.

One happy, serendipitous outcome revealed in the North Carolina analysis is that student performance improved as the state board raised standards. By using a common scale, state boards can monitor the trends in both performance and expectations.

The gap or disconnect between K-12 and the postsecondary world of college and careers is not a new phenomenon. Almost 100 years ago, Wilson and Hoke opined thusly:

The college instructor blames the high school teacher, the high school teacher complains of the grade teacher, each grade teacher above first grade finds fault with the poor work of the teacher in the grade below, and the first grade teacher in turn is chagrined at the shortcomings of the home training. Must this go on indefinitely? Whose opinion shall prevail? Is it not possible to get away from personal opinion to an agreed-upon consensus of opinion?

May we not replace the constantly conflicting subjective standards with definitely defined objective standards?

The North Carolina example gives us confidence that this perennial finger pointing and complaining can be finally put to rest. By examining reading growth longitudinally and connecting this growth to the real-world reading demands of occupations and higher education environments, state board members can have confidence that their state’s assessments are aligned to college and career readiness.
Strong school leaders play a vital role in accelerating student achievement across a state. Research has found that principals account for 25 percent of a school’s total impact on student achievement,¹ and an above-average principal can raise student achievement by as much as 20 percentage points more than an average principal² — equivalent to seven months of additional learning.³ Teacher surveys support these findings, with more than 90 percent saying that great leadership has a very important impact on student achievement.⁴

Rodney Rowan, principal of Cherokee Elementary School in Memphis, exemplifies
what great school leaders can do. Cherokee was failing when Rowan took the helm. He established a culture of high expectations for all students and—through meticulous use of data, strong curriculum, smart hiring, and intensive feedback and guidance for teachers—turned those expectations into reality. Since he arrived three years ago, student proficiency quadrupled in math and doubled in reading. Today, Cherokee is a “Reward” school—among the top 5 percent in Tennessee for progress.

Given the powerful multiplier effect of great school leaders, states must put recruiting, developing, and supporting great principals at the center of any school improvement strategy, with principal evaluation as a keystone. Recently, the question of how to raise the quality of teachers has dominated the public discourse, but states have overlooked the fact that expanding principal capacity is crucial to achieving that goal. As a result, state education agencies have devoted a lopsided proportion of their time and resources to developing and implementing new teacher evaluation systems but have invested little in the principals who are largely responsible for carrying out those evaluations.

In fact, teacher quality and principal quality are deeply intertwined. As illustrated in figure 1, strong school principals are vital in building teacher capacity. They provide effective instructional leadership, create a vibrant and positive school culture, and provide teachers with opportunities to grow professionally, collaborate with their peers, and take on leadership roles. All of these actions can reduce teacher turnover, motivate average teachers to improve, and encourage great teachers to remain in the profession. Indeed, 97 percent of teachers cite school leadership as a deciding factor in whether good teachers remain at a school.

If we want teacher evaluations to play a significant role in improving instructional capacity among our teaching corps, districts need principals who can conduct meaningful observations, accurately identify strengths and weaknesses, provide useful feedback, and facilitate support that will help teachers get better. Unfortunately, reports are rolling in from teachers across the country that their principals do not have these skills and that they are not being held to the same level of scrutiny. This deficit is breeding widespread cynicism and distrust in the new evaluation systems. To ensure that the work invested in teacher evaluations does not go to waste, states must also invest in developing principal evaluation models that hold principals to high standards of performance and help them improve in their work supporting teachers.

My organization, New Leaders, has spent 15 years preparing, observing, and studying school leaders. We have learned much about what great principals do and how to help them become even better. Over this same period, a number of states and districts have worked hard to get more great leaders at the helm of their schools, honing in on improving principal evaluation systems as one important way to achieve this goal. While this work is promising, we have observed that these states and districts frequently spend much time and energy designing evaluation tools but invest too little in implementation.

Ultimately, the goal of principal evaluations is to improve principals’ instructional leadership capacity. For this to occur, principal evaluations must put in motion a true cycle of goal setting, frequent observation and feedback, self-reflection, and adjustment in practice. Principals must have a thorough understanding of the expectations and standards they will be held to and be active participants in their self-improvement. And just as effective teacher evaluations depend on knowledgeable principals to carry them out, principal evaluations require skilled principal supervisors (such as superintendents and assistant superintendents) to enact them effectively.

Principal Evaluation as Performance Management

An effective principal evaluation process does two things at once: It assesses principal performance, and it helps principals develop the knowledge and skills they need to be highly effective. Unfortunately, this developmental role tends to get short shrift in the push for accountability.

A strong principal evaluation system should have the following key attributes:

- a shared understanding among district offices, principals, principal supervisors, and preparation programs of what effective school leadership looks like, translated into a focused set of principal standards that are grounded in research, aligned with school
accountability goals, and that inform all policies and programs related to principal quality (preparation, licensure, recruitment and hiring, in-service training, and evaluation);

- a strong and simple evaluation system that meaningfully differentiates levels of performance (we recommend a minimum of four levels);

- a design that is based on both inputs (principal actions) and outputs (school and student outcomes); and

- a robust system for observing, giving feedback, and providing support to principals. This requires significant investment in principal supervisors so they understand principal standards and expectations deeply and can provide accurate, actionable feedback.

For such systems to stick, states and districts need buy-in from the people who will be carrying out this work. Therefore, from the start of any effort to revamp principal evaluations (and as the work continues over time), it is important to bring in effective principals and principal supervisors to help craft the system. These educators can help you develop a system that feels manageable, that prioritizes the standards that resonate with them, and that will enable them to receive the support they need to become the strongest leaders possible.

**Developing Strong Principal Standards for Your State**

The first step toward a robust evaluation system is establishing a shared vision among all stakeholders of what great school leadership looks like and developing performance standards aligned to that vision. These standards should be grounded in principal actions that research has shown to increase student learning. Based on our extensive and ongoing review of research on principal effectiveness, and direct study of New Leaders principals, my organization has developed a list of standards that characterize great school leadership. Of these, three

---

**Figure 1.**

- Lending group learning activities
- Creating a professional climate of shared accountability for student learning
- Conducting observations w/useful feedback
- Staffing up
- Ensuring accountability
- Individualizing roles and responsibilities
- Fostering “Teacher Learning Communities”
- Building a culture of respect
- Instituting a student code of conduct

**Creating a Great Place to Work**
key areas of practice are the most critical:

- **Instructional Leadership.** Implementing high-quality, rigorous curriculum and instruction and using data to continually improve classroom instruction.

- **Talent Management.** Increasing the quality and capacity of teaching staff through effective evaluation, accurate feedback, professional learning, and shared leadership.

- **Culture Building.** Creating a strong school culture that includes high expectations, student support, a consistent approach to behavioral management, family engagement, and shared responsibility for school goals.

Many state standards lump the first two areas under one banner of teaching and learning, which discourages careful scrutiny of the actions principals take to increase the quality and capacity of their staff. In addition, many state standards do not include school culture as a separate standard from other instructionally focused practices. To ensure that principals get feedback and support on all three aspects of their leadership, New Leaders strongly advises differentiating between these three strands. The recently released Professional Standards for Educational Leaders, which update the earlier Interstate School Leaders Licensure Consortium (ISLLC) standards, include discrete standards focused on all three key areas.7

As state boards of education work to establish and prioritize principal standards, they should give educators a role in the process. For instance, state boards can convene leading principals, principal supervisors, and representatives from effective principal preparation programs to review different versions of principal standards, which could include a state’s current standards, the 2015 Professional Standards for Educational Leaders, and others such as New Leaders’ recommended principal evaluation standards.8 Then boards can prioritize those they determine will resonate and align best with their states’ improvement goals and accountability systems.

These expectations set for principals—regardless of which ones a state selects—should also inform principal preparation programs and licensure requirements. Ideally, states will have systems for communicating to preparation programs their graduates’ ratings so these programs can improve their training. At a minimum, states should align expectations across preparation, licensure, and evaluations.

### Designing an Evaluation Tool

Beyond establishing a clear set of priority standards, it is important to have a strong and simple evaluation system and tools. Creating a good evaluation tool can be time consuming and often distracts from the core work of implementation. To avoid this common pitfall, states should develop a model that districts can adapt while keeping the core elements. This approach has the added benefits of increasing the probability that all districts will use a high-quality tool and of making it easier to compare principal capacity across the state.

Historically, most evaluation tools have sorted principals into only two levels of performance: satisfactory or unsatisfactory. This binary rating does not give principals the nuanced, comprehensive feedback they need. Instead, there should be at least four levels: excellent, satisfactory, performance that needs improvement, and performance that is clearly unacceptable. The tools should use multiple measures, significantly weighing both professional practice and school and student outcomes. While clearly the ultimate aim for any school leader is to improve outcomes for students, an annual review of principal practices provides insight into why they may or may not be effecting gains in student achievement and helps to pinpoint areas where they may need support.

Teachers and other school staff observe their principals every day and therefore are an important source on their principal’s talent management, instructional leadership, and culture-building skills. States should administer annual teacher surveys and incorporate that information into the professional practice portion of the evaluation. Note that several universities and organizations have developed strong surveys focused on principal practice that states can adopt or adapt.9

In the best systems, ratings based on student outcomes use measures of both absolute attainment and gains, thus encouraging growth for all students rather than just those approaching proficiency. They set targets that close achievement
development. This level of participation from both parties enables deep, practice-based conversations and ensures that evaluation does not devolve into an exercise in compliance.10

As illustrated in figure 2, evaluation begins with data analysis, self-reflection, and goal setting. Throughout the school year, the plan is carried out and evidence is collected. The cycle continues with a midyear formative review and further implementation. Then principals get a chance to formally self-assess, review interim data, and reflect on progress, a step that informs the summative evaluation. Evidence from the summative evaluation and self-assessment feeds the principal’s subsequent goal setting, as the cycle continues into the following school year.

While principals will be rated on the full spectrum of standards designated by the state, principals demonstrate stronger growth when they and their evaluators select two or three areas for focus. Observations and feedback can then be targeted to those areas, supporting deeper development. Critically, the chosen areas of practice should tie clearly to student learning goals.11

Implementing Robust Evaluations

The principal evaluation process should animate a rigorous cycle of continuous improvement. Principals and their evaluators collect evidence about school practices and student results over the course of a year; deliver, review, and reflect on feedback; calculate a final rating; and identify goals and recommendations for the following year. For this process to work, principals’ supervisors must spend more time in schools observing practice and providing actionable feedback, and principals must play an active, engaged role in their own professional development. This level of participation from both parties enables deep, practice-based conversations and ensures that evaluation does not devolve into an exercise in compliance.10

As illustrated in figure 2, evaluation begins with data analysis, self-reflection, and goal setting. Throughout the school year, the plan is carried out and evidence is collected. The cycle continues with a midyear formative review and further implementation. Then principals get a chance to formally self-assess, review interim data, and reflect on progress, a step that informs the summative evaluation. Evidence from the summative evaluation and self-assessment feeds the principal’s subsequent goal setting, as the cycle continues into the following school year.

While principals will be rated on the full spectrum of standards designated by the state, principals demonstrate stronger growth when they and their evaluators select two or three areas for focus. Observations and feedback can then be targeted to those areas, supporting deeper development. Critically, the chosen areas of practice should tie clearly to student learning goals.11
Bringing Educators into the Process

Educators must be involved in developing a strong evaluation system, and this involvement must go beyond participation in the initial design. Instead, states can create structures to collect ongoing feedback from those who are carrying out the work.

Delaware has done this well. Over the past few years, Delaware developed and implemented a new evaluation system for principals, in partnership with New Leaders, the University of Delaware, and the Delaware Department of Education. Principals and principal supervisors served on the design committee and also received initial training in implementation, including how to develop and carry out an evidence collection plan, prepare for and conduct effective midyear conversations, and carry out self-assessments and summative judgments based on collected evidence. But their participation didn’t end there. They continued to serve on the committee and, after the first year of implementation, shared what they had learned from the experience to adjust and improve the system. Perhaps most important, the principal supervisors who best leveraged the evaluation process to improve principal practice were assigned to train other supervisors in the state.

Building Principal Supervisor Capacity

Even if states spend substantial time establishing strong standards, designing an effective evaluation tool, and developing a clear structure for carrying out the process, little will change if the evaluators themselves do not schedule regular school visits or cannot deliver ongoing feedback and coaching to make principal evaluations a truly developmental experience.

Traditionally, principal supervisors’ job was to ensure that school leaders complied with local, state, and federal rules and regulations. Transitioning from compliance officer to coach requires effort and investment. Aside from the role in monitoring compliance, the principal supervisor position varies across states and districts, and there frequently is no clarity about responsibilities and expectations.12 Often, school visits are not anyone’s responsibility at the central office. Some principals in New Leaders’ client states have reported that they never had a school visit from their supervisor and that their evaluation consisted of little more than submitting a binder and receiving a rating. And some New Leaders principals have been promoted to supervisory positions where the responsibilities are vague, so that every supervisor in their district is left to interpret job requirements.

I cannot stress enough the central importance of developing capacity among principal supervisors. States and districts can begin this process by setting out clearly what principal supervisors should know and be able to do (see the forthcoming CCSSO 2015 Model Principal Supervisor Standards as a starting point), rewriting the job description so that observation and coaching is a core responsibility, and shifting responsibilities to other central office employees so that principal supervisors—whether they are superintendents or assistant superintendents—have time to conduct multiple school visits. Finally, states or districts must invest in upfront training for both principals and principal supervisors to ingrain leadership standards and expectations and a strong understanding of the evaluation process. Then they will need ongoing support to develop their skills in observation and instructional leadership coaching.

Many principal supervisors were school leaders themselves and are familiar with the job’s daily challenges. That familiarity, however, does not automatically translate to skills at identifying growth areas in the leaders they supervise and in delivering accurate, actionable, and results-driven feedback. Like everybody else, principal supervisors need practice and coaching in adult leadership skills.

Recognizing the need to strengthen the role, the Cleveland school district hired New Leaders to establish a community of practice that prioritized developing a shared understanding of the principal supervisor’s responsibilities. Once the group came to a consensus about the role, the New Leaders instructor led principal supervisors on site visits to practice collecting evidence and in group discussion on what was observed, what constituted effective practice, and the kind of feedback that would help principals improve. New Leaders led exercises in conducting difficult conversations with principals and reviewed sample school data with the supervisors to develop their skills in establishing appropriate growth targets for the principal and the school.
All states can play a similar role in ensuring that their principal supervisors receive and complete high-quality training around evaluation implementation. Alternatively, or in addition, they may want to create a more formal certification process for principal evaluators, as many states have done for teacher evaluators.

Questions State Boards of Education Should Ask

Principals are at the front line of national reform efforts and are a key lever for ensuring their success. They lead implementation of new, more rigorous college- and career-ready standards and assessments and carry out a far more complex and in-depth teacher evaluation process than in the past. Yet policies and programs around principal quality and development have not kept up. Most states spend far more on teacher evaluations and professional development than they do on principals—a counterproductive omission. We cannot improve teacher quality at scale without great principals.

It is heartening that at least some states and districts are spending time and money to improve principal evaluation tools. While this is the right instinct, there are troubling signs of a misplaced emphasis on design. There are already a number of high-quality principal evaluation models available for free that can easily, and with minimal cost, be adapted to align with local accountability systems and improvement priorities. Instead, states and districts should spend far more time and money on implementation. This entails training principals and principal supervisors in the new evaluation system, building principal supervisor capacity, and ensuring that principals have high-quality, job-embedded opportunities to build the skills they need to help their students achieve dramatic academic gains.

State boards of education can play an important role. To determine where to focus attention and resources, state boards can ask themselves the following:

- Do our policy discussions about educator effectiveness focus enough on principals?
- Are our standards for principals adequately focused on the most important aspects of a principal’s role?

- Does our state have a strong model for principal evaluation that districts can adopt or adapt?
- As we consider new systems, how are we engaging practitioners—particularly principals and their supervisors—in design and implementation?
- Are we thinking beyond the initial design of new systems and considering how we will support good implementation?
- What other policies (e.g., licensing rules, principal preparation regulations) help or hinder efforts to increase principal effectiveness through improved evaluation?

---

2Based on a 2005 study by McREL that analyzed 70 studies from more than three decades of research.
4Marzano et al., School Leadership That Works.
9Examples of teacher surveys focused on principal quality include 5Essentials survey from University of Chicago and VAL-ED from Vanderbilt University. The New Teacher Center also has a survey tool, TELL and the Cultural Insight Survey, that can be obtained upon request.
11For more detailed information, see New Leaders, The Principal Evaluation Handbook (New York City, 2012).
Georgia’s Teacher Performance Assessment

Like most states, Georgia until recently depended on an assessment of content knowledge to award teaching licenses, along with a licensure recommendation from candidates’ educator preparation programs. While the content assessment reflected candidates’ grasp of subject matter, licensure decisions did not hinge on direct, statewide assessment of teaching effectiveness. This method of assessment began to change in Georgia in 2012, when the state began to explore using a content-specific performance assessment for teacher candidates.

Today Georgia is implementing edTPA in all 59 initial teaching preparation programs across the state. This shift affected the entire continuum of educators, from recruitment to preservice preparation, certification, induction, and retention. Not only has edTPA changed how Georgia’s educators think about teaching and learning, it has built bridges and stronger partnerships across the state’s education community.

edTPA, a subject-specific performance assessment developed by Stanford University and the American Association of Colleges for Teacher Education, has been operational since 2014.1 Forty-one states and 626 educator preparation providers participate in edTPA nationwide. In six states, edTPA is required for teacher licensure, certification, or program completion.2 As with the National Board Certification process for veteran teachers, edTPA focuses on analysis of classroom practice, supported by video documentation, to assess and foster growth in the core areas of teaching that have long been priorities in most preparation programs.

edTPA is one of several performance-based assessments designed for beginning teachers. When Georgia began exploring performance assessments, edTPA was far along in its development, and Georgia sought a content-specific pedagogy assessment aligned to multiple preparation fields.

As a performance-based measure, edTPA does not rely on candidates’ answers to multiple-choice tests or a faculty member’s point-in-time observation of a candidate’s student teaching. Rather, it assesses how a candidate demonstrates real classroom teaching against 15 rubrics covering three core areas: planning, instruction, and assessment.

Teacher Educators Propel Shift

The impetus for edTPA’s adoption in Georgia came not from a state mandate but because Georgia’s teacher educators recognized in it their own aspirations for what teacher preparation ought to look like.

The path toward statewide adoption began in fall 2011, when a few Georgia teacher educators began to explore edTPA and some attended the national edTPA Implementation Conference to learn more. One of the attendees was Deirdre Greer, then department chair at Columbus State University (CSU). She called a colleague from the conference to relay her reaction: edTPA assessment was measuring what they had long felt were the most important qualities of effective teaching.

In spring 2012, Georgia was invited to participate in a national edTPA field test, in which seven educator preparation providers explored edTPA with their candidates. CSU was one of those, and while their edTPA journey is only one example of the focused efforts toward educative implementation by Georgia preparation providers across the state, it is a good example.

For the national field test, CSU selected 10 teacher candidates from a pool of volunteers to go through the full process. The candidates completed the requirements for edTPA’s portfolio in addition to their existing assignments. When the detailed score reports for these initial 10 came in at the end of the semester, faculty pored over the data. It was clear that some preparation methods needed attention. For example, the reports revealed that candidates needed more preparation for developing relevant, authentic assessments of student learning. They also needed more support as they analyzed their students’ learning and used results to inform future instruction.

As Georgia preparation providers were exploring edTPA, the Georgia Professional Standards Commission was developing a tiered certification structure for teachers. A
diverse task force informed this work, including preparation providers, preK-12 schools and districts, regional education service agencies (RESAs), educational organizations, the Georgia Department of Education, the university system of Georgia, and others. Programs that had participated in the edTPA field test presented their experiences to the task force; their positive accounts as well as additional examination prompted the task force to recommend that the commission consider edTPA for use statewide.

In the 2012–13 academic year, CSU decided to do a more extensive pilot involving about half its candidates. To help faculty learn in detail what edTPA requires, CSU faculty used materials developed by the Stanford Center for Assessment, Learning, and Equity (SCALE) to evaluate the assessment locally, which helped them better understand edTPA’s rubrics and process. CSU faculty and candidates dug deep that year. Faculty called the 2012–13 cohort of candidates the Pioneer Group, awarded them certificates, and celebrated candidates’ completion of their edTPA portfolios.

Other Georgia preparation providers, both public and private, traditional and nontraditional, were similarly getting their feet wet. As with any transition in education, the providers needed time to evaluate edTPA and develop their own structures to support it.

Taking a Collaborative Path

Several preparation providers convened a summit later in 2012 in Macon, Georgia, to share the paths their programs were taking. For many attendees, this was the first in-depth conversation with their counterparts about course sequences, candidate portfolio components, and evaluating instruction using the edTPA rubrics. In fact, Georgia’s education agency delayed edTPA implementation to give time for an organic approach to emerge based on providers’ interests. Eventually, the agency hosted a series of local, regional, and statewide meetings and began creating support structures for programs and their preK-12 school partners. To enhance communication during this period, each provider appointed an edTPA coordinator as a statewide point of contact.

This groundwork helped preparation providers tackle edTPA implementation. Participants gained insight from their colleagues at these forums and developed strong, self-sustaining relationships. Another benefit for teacher educators was edTPA’s detailed tasks and standards, which provided a common language for talking about teacher preparation. From state agencies to deans and faculty and their school-based partners, this shared lexicon helped clarify goals and expectations among the providers, the state agencies, and preK-12 school leaders.

To maintain this collaborative spirit, the state formed a Policy and Implementation Advisory Group consisting of preparation providers, preK-12 representatives, RESAs, and state agencies. Part-time regional consultants also offered technical assistance to colleagues across the state.

One of the last tasks before full implementation in 2015 was to set the performance standard or cut score needed to pass. Following an evidence-based process similar to that used by SCALE with a large national panel of preK-12 educators and teacher educators in 2013, the Georgia Professional Standards Commission proposed ramping up the cut score over time, starting at 35 and moving to 38 for the 2017–18 academic year. With advisory group input, the state decided to examine data each year over the next three years before setting a timeline to move toward the nationally recommended professional performance standard. That score, 42, will require continued dialogue and familiarity with edTPA among candidates, their program providers, and preK-12 partners—an ongoing project for Georgia.

“Better at a Faster Rate”

Meanwhile, CSU faculty were finding that edTPA helped address long-standing quandaries: What’s the best way to prepare candidates who are weak instructors or poor classroom managers for student teaching? What steps ensure that faculty have honest conversations with candidates about their performance before they’re eligible for student teaching?

To foster such frank conversations, CSU reorganized its education program around “critical assignments,” which serve as gateways for continued progress in the program. For example, in the second semester of their junior year, candidates take on assignments that help
prepare them for edTPA's task 2, Analysis of Effective Teaching. They watch videos of their own teaching, evaluating it using accepted standards for excellent teaching. This assignment builds candidates' comfort in being recorded, helps them understand student-privacy policies, encourages them to closely examine their own practice, and increases their awareness of the many dimensions on which student learning occurs.

CSU candidates must pass three such formative assignments before they begin student teaching. All candidates fashion professional development plans with their advisers; these prompt faculty to talk with candidates about their progress and to suggest changes if they are struggling with a critical assignment. Thanks to edTPA, this feedback is rooted in faculty supervisors' comparisons of their observations of candidate performance with characteristics of effective instructional practice.

The rigors of edTPA catch some candidates off-guard. In fall 2014, one CSU candidate, who had just begun student teaching, sat sullenly, arms crossed in a meeting with CSU faculty. He expressed frustration at the demands the assignments were making of him, on top of his student teaching. His instructor reiterated that the candidate's work was being evaluated against an accepted yardstick for great teaching so that he would be fully prepared for his first day as a new teacher. Eventually, the candidate said, "I get it. You guys want us to become better at a faster rate." He's now a successful teacher in Gwinnett County Schools, outside Atlanta.

Generally, candidates see edTPA as a boon: They feel more confident in their abilities heading into the classroom and are clearer about their strengths as well as areas for growth. edTPA's focus on differentiated instruction contributes to this increased confidence. It requires candidates to see the classroom as a group of individuals who each have needs and strengths rather than a mass of students.

Schools and districts also see a positive impact. "We have noticed a significant difference in the levels of preparedness of CSU pre-service teachers and first-year teachers since edTPA has been introduced," said Justin M. Finney, principal at Harris County's Mulberry Creek Elementary. "CSU pre-service teachers and first-year teachers now make an immediate positive impact in our school. They now come to us prepared to provide high levels of instruction that result directly in student learning and growth."

"The edTPA performance assessment raises the bar for our preservice teachers and better prepares them for the real world of education," said Richard Rogers, director of personnel with the Houston County School System and part-time professor at CSU. "In addition, P-20 institutions are now working together to understand, support, and be accountable to one another."

Wendy Sands, a teacher at Richards Middle School in Muscogee County said, "The edTPA has streamlined the expectations for preservice teachers to better prepare them for the necessity of documentation when they are in their own classrooms."

Celebrating candidates as they go through edTPA, as CSU did with its Pioneer Group, has emerged as a means of maintaining candidate investment in the process and of welcoming them into the teaching profession. Professors, friends, and family members all celebrate prospective attorneys who pass the bar exam. The same goes for doctors, nurses, accountants, or other professionals who pass their professional entrance exams. Fully professionalizing teaching also means publicly celebrating those who attain this high bar for entrance.

Peach State Spreads the Word

In August 2014, the Alabama State Department of Education invited the Georgia Professional Standards Commission to a convening of Alabama educator preparation providers at Auburn University to discuss Georgia's implementation, as well as requirements and options for state certification. Turnout was impressive: 21 of Alabama's 27 state institutions were represented.

Rhonda Morton, associate dean of education at the Auburn University Montgomery, was among the attendees. She booked a trip to the next edTPA Implementation Conference in Los Angeles, where she met with Andrea Whittaker, SCALE's director of teacher assessment and a national leader of edTPA implementation.

Energized by her experience, Morton became deeply involved in Alabama's edTPA push.
To broaden the base of support in the state, she invited a mix of new and veteran faculty members to the edTPA Southeast Regional Summit, held in Savannah in April 2015. Testimonials from Georgia candidates about the value of the assessment convinced many of Morton’s colleagues of the benefits of edTPA’s focus on performance.

Georgia faculty members at the summit were also candid about the challenges they encountered and how they overcame them. These include the need for careful mapping of edTPA to the curriculum, integrating a common understanding of academic language into each subject area, and strengthening communication across state agencies, educator preparation providers, and preK-12 districts and schools.

Alabama’s department of education offered grants to institutions for technical assistance, faculty training in scoring, and conference attendance to support edTPA pilots. According to Morton, edTPA was “just a little twist” on what Alabama was already doing—since many state institutions were already using an electronic portfolio model that included videos. But the programs needed time to adapt and to increase their comfort with the idea of outside national scorers evaluating their candidates’ performance.

As in Georgia, edTPA sparked increased collaboration in Alabama. Small meetings of edTPA coordinators now precede gatherings of the Alabama Association of Colleges for Teacher Education (ALACTE). Thomas Bice, Alabama state superintendent of education, has briefed superintendents, who have held discussions with principals, who have had conversations with cooperating teachers. Alabama educators are now beginning to “pay it forward”: ALACTE has met with counterparts throughout Alabama.

Georgia’s edTPA implementation reconfirmed the value of collaboration for Georgia’s educators. Whether through technical assistance meetings and conferences, webinars, or monthly calls for edTPA coordinators (held regularly since 2013), edTPA has resulted in more conversations about all aspects of effective teaching, and it is a common topic of conversation at local, regional, and statewide meetings.

Another benefit is the intentional interaction between preparation providers and preK-12 schools. The nuts-and-bolts demands of edTPA, particularly the videotaping of lessons, have necessitated many conversations about the means and ends of student teaching. Conversely, new graduates can use their edTPA experience and data to articulate their strengths and where they need support. Employing districts and principals are finding edTPA sheds light on one of their biggest questions about new teachers: How can I best support them during their first years of teaching? District induction programs can now make better-informed support decisions.

“We believe that adding a performance assessment for new teachers is the right step for Georgia preK-12 students,” said Kelly Henson, executive secretary of the Georgia Professional Standards Commission. “We also know that change is not easy. That is why we have taken time to build partnerships across the state and at all levels of education, creating a community of learning and collaboration with the state, providers and their faculty, and P-12 school districts and schools to support each other during this transition and beyond.”

Although edTPA was required for teacher candidates’ program completion only with this academic year—and Georgia is still building capacity to fully support it—the state is already seeing positive outcomes, as expressed by one CSU candidate, Kimberly Adkins, who wrote, “It’s comforting to know that edTPA is in fact making me a better teacher.”


Anne Marie Fenton is director of assessment at the Georgia Professional Standards Commission, and Pamela Wetherington is edTPA coordinator at Columbus State University in Columbus, Georgia.
New Hampshire’s Accountability 3.0

Accountability must be a reciprocal process. For every increment of performance I demand from you, I have an equal responsibility to provide you with the capacity to meet that expectation. Likewise, for every investment you make in my skill and knowledge, I have a reciprocal responsibility to demonstrate some new increment in performance. 1

—Richard Elmore

Education policymakers over the past few decades have made two critical decisions that I believe have had deep, deleterious effects on the overall public education system. First, by demanding that each school be evaluated based on a single externally determined assessment or set of measures, we branded public educators and schools as mediocre at best and failures in the eyes of many in their communities. Second, we took responsibility, if not accountability, out of the hands of local school boards, educators, parents, and students.

The result is a one size fits all system, where instead there should be vibrant schools in which educators develop their own mission statements, theories of action for teaching and learning, robust courses of study, assessments based on their own learning paradigm, and accountability based on research-based indicators of excellence and success. It is no wonder so many have been so taken by charter schools, where responsibility has been largely given over to charter boards. Yet overall outcomes for charters have been surprisingly similar to those of public schools.

At the same time, a return to local educator responsibility would not relieve the state education agency and the state board of education of their responsibility for accountability, transparency, and equity. They still must ensure that students graduate high school knowing how to read, write, and calculate. They need these basic skills not just to succeed in college but also in careers. Additionally, students need the agility to adapt to a rapidly changing world. Developing this adaptability requires a retooling of the system.

How is this circle to be squared? I believe New Hampshire has found a third and better way—one that deprives neither local educators and communities of their ability to drive learning in their schools nor states of their need to ensure that public education is working for all students. Rather, it focuses on model competencies jointly developed by local districts and the department, intensive professional development around performance tasks, and a coherent system of assessments, also developed with a joint local and state imprint.

In 2011, the New Hampshire State Board of Education adopted this vision for the state’s public schools:

To harness community resources and technology to provide a world class, personalized, student-centered education in a flexible, innovative learning environment that promotes active engagement to maximize the potential in every individual.

Adoption of this vision spurred the state education agency and school districts to direct schools toward competency-focused education, blended learning, and student-centered practices. The board did not adopt this vision out of the blue: It is consistent with board statements and leadership as far back as 2005, when it adopted rules eliminating the Carnegie unit and requiring that students master competencies in order to earn credit toward high school graduation.

At the beginning of her tenure, Commissioner Virginia Barry launched a collaborative dialogue with state board members, community leaders, superintendents, principals, teachers, parents and students, and faculty of institutes of higher learning. She asked them to make a plan for the next several years. As a result, the New Hampshire Department of Education adopted this theory of action:

If we expect that all students will be college and/or career ready by the time they graduate from high school, then we see the need for a system of personalized learning that has clear and high expectations, delivered through a rigorous competency-based course of study.

One of the architects of New Hampshire’s overhaul of its school accountability system explains what drove the change and how the state is doing it.

by Paul Leather
career success. With Accountability 3.0, teaching and learning is becoming more personalized and more competency and project based, and the system more directly balances formative and summative assessments around a student-centered approach.

New Hampshire’s new model, Performance Assessment of Competency Education (PACE), dovetails with the work of Linda Darling-Hammond, Gene Wilhoit, and Linda Pittenger. In their 2014 paper, they describe a continuous improvement and innovation model that focuses on richer, deeper learning that requires the development of the professional capacity and resources needed to support this learning at every level.

PACE balances state standardized assessments with locally developed performance-based assessments and portfolios. The standardized assessments are used to validate local assessment results; the performance-based assessments are used to garner more robust test results and inform teaching and learning. New Hampshire now administers state assessments once in each grade span. With local educators leading the way, the SEA has created complex, curriculum-embedded performance assessments in the intervening grades. These assessments were designed to support a system in which students are engaged in increasingly complex tasks based on state and national standards.

New Hampshire began using PACE in a few pilot districts—first in four and in eight this year. In order to join the pilot, districts had to demonstrate they had put several “guardrails” in place:

- adoption of state-model competencies aligned with college and career outcomes as the main learning targets;
- an instructional system to support student learning of competencies, including strategies to personalize learning;
- an assessment system to measure student achievement and growth related to the competencies;
- use of vetted performance assessments in the local assessment system; and
- administration of the Smarter Balanced assessment at least once each in elementary, middle, and high school.

New Hampshire invested in intensive professional development for its educators, with a particular emphasis on professional learning to develop rich performance tasks.
Once these communities demonstrate readiness to join the pilot, their district leaders joined the SEA in constructing the overall system of assessments. Currently at the table are lead educators from eight districts and schools, lead staff from the department, and New Hampshire’s core technical partner, the Center for Assessment, which is based in Dover, NH. Together, we developed key design principles, which include the following:

- a focus on college and career outcomes and promotion of deeper learning for all students;
- a clear commitment to the achievement of educationally disadvantaged students;
- a clearly described internal accountability process supported by the local board of education;
- commitment of resources necessary to ensure the plan’s success by both the state and local districts; and
- development of leadership and educator capacity to design, implement, support, and sustain the system.

District and state staff also worked together, with technical advice from the Center for Assessment, on actual construction of common performance assessments. The operative maxim of the state’s support of district work is “trust but verify.” State support includes these elements:

- assistance in preparing an application to the department for inclusion in the pilot;
- help in conceptualizing local assessment system design;
- design and selection of assessments and other local measures;
- data analysis of student results;
- review and implementation, including calibration of student results on performance assessments; and
- reports on progress and adjustments to the system based on monitoring of student outcomes.

As part of these supports, the Center for Assessment is developing comparability studies that will weigh the results of the grade-span state assessments against local competency determinations, with common performance tasks as checks in the process. In this way, the department can “validate” the use of rich, curriculum-embedded tasks in its annual accountability determinations over time.

While those checks are crucial to the ultimate success or failure of the PACE pilot, in the interim, department staff are encouraged by the stories they have heard along the way. Teachers in one district declared, “This is incredibly hard work, but it’s the right work to support our students’ learning.”

PACE’s pilot phase ends with this school year. The department projects that full implementation throughout the state will take five to ten years.

State boards of education everywhere are entering a new day in terms of assessments and accountability. Key principles can guide their initiatives:

- Decisions about assessments and even accountability measures should be done in partnership with local educators and communities, with the state providing technical support.
- Readiness to partner matters, and state education policymakers should assess and support the state agency and local agencies’ readiness to engage.
- One size does not fit all.

These principles are driving New Hampshire’s efforts. As New Hampshire puts education back into the hands of educators, with checks along the way, I believe our state will see richer student work, higher student performance, and a new era for learning.

---

Making Sense of New Science Assessments

What we choose to assess in science is what will end up being the focus of instruction. US science standards once treated content and inquiry as fairly separate strands of science learning, with content standards stating what students should know and inquiry standards stating what they should be able to do. In its content coverage, these standards were also deemed to be a “mile wide and an inch deep.” Assessments followed suit.

By contrast, the National Research Council in its Framework for K-12 Science Education Standards presents a very different way of thinking about science proficiency. The framework articulates three interconnected dimensions of competence: disciplinary core ideas, crosscutting concepts, and scientific and engineering practices.

The framework focuses on core ideas in four areas: (a) life sciences, (b) physical sciences, (c) earth and space sciences, and (d) engineering, technology and the application of science. In so doing, it attempts to reduce the long, often disconnected catalog of factual knowledge that students have been expected to learn. Core ideas in the physical sciences include energy and matter, for example, and core ideas in the life sciences include ecosystems and biological evolution. Students are supposed to encounter these core ideas over the course of their school years at increasing levels of sophistication, deepening their knowledge over time.

The second dimension is crosscutting concepts, of which the framework identifies seven that have importance across many science disciplines: patterns, cause and effect, systems thinking, and stability and change, for example. Eight key science and engineering practices— the third dimension—include asking questions, defining problems, planning and carrying out investigations, developing and using models, and engaging in argument from evidence.

The disciplinary core ideas and crosscutting concepts serve as thinking tools that work together with scientific and engineering practices to enable learners to solve problems, reason with evidence, and make sense of phenomena.

Adopting this view of science competence also requires a change in how learner proficiency is measured. Figure 1 shows a task that could be used to assess middle school students’ understanding of the properties of substances that are associated with chemical identity. It does so in the context of constructing an argument about which substances are the same or different based

New state science standards that integrate core ideas, cross-cutting concepts, and practices cry out for new assessments that do likewise.

by James W. Pellegrino

Figure 1. A Physical Science Assessment Task for Middle School

Steven found four different bottles filled with unknown pure liquids. He measured the properties of each liquid. The measurements are displayed in the data table below. Steven wonders if any of the liquids are the same substance.

<table>
<thead>
<tr>
<th>Liquid</th>
<th>Density</th>
<th>Color</th>
<th>Volume</th>
<th>Boiling Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0 g/cm³</td>
<td>Clear</td>
<td>6.1 cm³</td>
<td>100 C°</td>
</tr>
<tr>
<td>2</td>
<td>0.89 g/cm³</td>
<td>Clear</td>
<td>6.1 cm³</td>
<td>211 C°</td>
</tr>
<tr>
<td>3</td>
<td>0.92 g/cm³</td>
<td>Clear</td>
<td>10.2 cm³</td>
<td>298 C°</td>
</tr>
<tr>
<td>4</td>
<td>0.89 g/cm³</td>
<td>Clear</td>
<td>10.2 cm³</td>
<td>211 C°</td>
</tr>
</tbody>
</table>

Use the data in the table to:

1) Write a claim stating whether any of the liquids are the same substance.
2) Provide at least two pieces of evidence to support your claim.
3) Provide reason(s) that justify why the evidence supports your claim.

on patterns of data. A teacher could use this task to gauge how well her students have understood which properties are associated with chemical identity as well as their ability to use evidence to construct a scientific argument.

The central argument of A Framework for K-12 Science Education Standards is that competence is realized through performance expectations about what students should know and be able to do. These statements of performance expectations integrate the three dimensions and move beyond vague terms such as “know” and “understand” to more specific statements—“analyze”, “compare”, “predict”, and “model.” Finally, Framework makes the case that competence and expertise develop over time and increase in sophistication and power as the product of coherent systems of curriculum, instruction, and assessment.

How Will We Know What Students Know?

The framework’s three-part structure—practices, crosscutting concepts, core ideas—signals an important shift for science education. It also presents a challenge for the design of both instruction and assessment—finding a way to describe and capture students’ developing competence along these intertwined dimensions.

In developing the Next Generation Science Standards (NGSS), Achieve and its partners described how students at each grade are expected to apply the practices and crosscutting concepts to the knowledge of core ideas they are expected to have. Each performance expectation asks students to use a specific practice and a crosscutting concept in the context of an element of disciplinary knowledge. Across each grade-level set of expectations, each practice and crosscutting concept appears multiple times.

But standards and performance expectations, even as explicated in the NGSS, do not provide sufficient detail to create assessments. The designer of valid, reliable science assessments must be developed for specific purposes, and that intended use will drive their design. In the classroom context, instructors need formative assessments to make decisions about next steps for instruction, give students feedback about their progress, and motivate them. They need summative assessments to help determine whether a student has attained a certain level of competency after completing a particular phase of education.

Large-scale assessments—which are administered at the direction of users external to the classroom—also provide information about the attainment of individual students, as well as comparative information about how one individual performs relative to others. But the results seldom help teachers or students make day-to-day or month-to-month decisions about teaching and learning. Another common purpose of these assessments is to help administrators, policymakers, or researchers form judgments about the quality and effectiveness of educational programs and institutions. Just as with individuals, the quality of the measure determines the validity of these decisions.

The purpose of an assessment determines priorities, and the context of use imposes constraints on the design. A one-size-fits-all fallacy is often leading to inappropriate choices of assessments for instructional, evaluation, or research purposes that in turn can lead to invalid conclusions regarding persons, programs, or institutions.

A Balanced System

Since one form of assessment cannot serve all the needs of the various actors in an educational system, multiple assessments will inevitably be required. Given the many assessments already used in schools, it is not surprising that educators are often frustrated when such assessments appear to have conflicting goals and yield inconsistent results. Such discrepancies can be meaningful and useful, as when assessments are explicitly aimed at measuring different school outcomes. More often, however, conflicting goals and feedback from the tests shed more heat than light. Thus it is critical that state education policymakers develop a vision for a balanced, coordinated system of assessments to promote effective science teaching and learning that takes into account both classroom assessments and large-scale monitoring assessments.
The Classroom Component. Classroom assessments are integral to instruction and learning and will be both formative and summative. Assessment activities will tell students what is expected of them and give them opportunities to reflect on their performance. Teachers will get information with which they can adapt their instruction. Instruction that is aligned with the NRC framework and associated standards will naturally provide many opportunities for teachers to observe and record evidence of student thinking, such as when students develop and refine models; generate, discuss, and analyze data; engage in both spoken and written explanations and argumentation; and reflect on their own understanding.

The Monitoring Component. This system component is used to answer a range of important system-level questions about student learning (see table 1). In the United States, there are two predominant types of data collection. The first is a fixed-form test, in which all students take the same form of the test on a given occasion. The science assessments that states employ to comply with No Child Left Behind (NCLB) and now the Every Student Succeeds Act (ESSA) are examples: Each public school student at the tested grade level in a given state takes the full test. ESSA requires that these tests be given to all students in the state at least once in each of three grade spans (K-5, 6-8, 9-12). They address questions about student-level performance (first column of table 1). The scores are also aggregated as needed to provide information for the monitoring school-, district-, and state-level performance (the three right-hand columns).

A second type of test administration makes use of matrix sampling, which is used when the primary interest is group- or population-level estimates (i.e., schools or districts) rather than individual-level estimates (the middle two columns of table 1). No individual student takes the full set of items and tasks. Instead, a sufficiently large, representative sample of students completes each task. This method makes it possible to gather data on a larger, more representative collection of items or tasks for a given topic than any one student could be expected to complete in the time allocated for testing. In some applications, all students from a school or district are tested (with different parts of the whole test). In other applications, only some students are sampled for testing.

Matrix sampling is a powerful, economical, and relatively straightforward option that has not generally been possible in the context of

<table>
<thead>
<tr>
<th>Types of Inferences</th>
<th>Levels of the Education System</th>
<th>Policy Monitoring</th>
<th>Program Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Individual Students</td>
<td>Schools or District</td>
<td></td>
</tr>
<tr>
<td>Criterion-referenced</td>
<td>Have individual students demonstrated adequate performance in science?</td>
<td>Have schools demonstrated adequate performance in science this year?</td>
<td>How many students in state X have demonstrated proficiency in science?</td>
</tr>
<tr>
<td>Longitudinal and comparative across time</td>
<td>Have individual students demonstrated growth across years in science?</td>
<td>Has the mean performance for the district grown across years? How does this year’s performance compare to last year’s?</td>
<td>How does this year’s performance compare to last year’s?</td>
</tr>
<tr>
<td>Comparative</td>
<td>How does this student compare to others in the school/state?</td>
<td>How does school/district X compare to school/district Y?</td>
<td>How many students in different states have demonstrated proficiency in science?</td>
</tr>
</tbody>
</table>

Table 1. Questions Answered by Monitoring Assessments

Reprinted with permission from Developing Assessments for the Next Generation Science Standards, 2014 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, DC.
state testing in the last decade because of the requirements of NCLB for individual student reporting.

Both fixed-form and matrix sampling approaches can be combined in a single test. For example, one test could include a fixed-form component for estimating individual performance and a matrix-sampled component to estimate performance at the school level. States such as Massachusetts, Maine, and Wyoming used such a hybrid design before NCLB was implemented.

Regardless of the form that monitoring takes, the tasks must address the progressive nature of learning, include multiple components that reflect three-dimensional science learning, and include an interpretive system for the evaluation of student products. Such assessments also must be designed so they can be given to large numbers of students, are sufficiently standardized to support the intended monitoring purpose, cover an appropriate breadth of the standards, and are cost effective.

Opportunities to Learn. Another system component—perhaps less obvious than the classroom and monitoring components—is a series of indicators of opportunity to learn. Such indicators will enable states to evaluate the equity of students’ opportunity to learn science that is aligned to high-quality standards. Without these indicators, there can be no way to adequately and appropriately interpret the assessment results obtained at each level of the system—especially critical if the results are to be used for accountability.

A Coherent System

State boards of education and state education agencies need to understand and plan for development and implementation of new science assessment systems in stages, over a span of years. Many innovative assessment programs floundered in the 1990s in part because they were implemented far too rapidly (perhaps to meet political exigencies). In many cases, developers were not given sufficient time to implement major changes or to modify assessments as they learned from experience. Some have cited this rush to implement at scale as a key factor in the lack of sustainability of many such efforts.

Any new assessment system has to evolve alongside other elements that are also changing: curriculum, instruction, professional
development, and the other components of science education. Coordination is needed not just because assessments have to be embedded in curriculum and instruction but also because students ought not be assessed on material and kinds of learning they have not had the opportunity to master.

With regard to the opportunity to learn, many schools and districts reduced the amount of science instruction, particularly in early grades, partly in response to the accountability demands of NCLB. If they are to implement the new standards effectively, many jurisdictions will need to reintroduce science in the early grades and review and revise policies that limited the time available for science. Frequently, schools that serve the most disadvantaged students are those in which the opportunity to learn science has been most reduced. Even in schools and districts that maintained strong science programs at all grade levels, neither students nor teachers have had experience with instruction that involves applying the practices as envisioned in NRC’s Framework.

Given the magnitude of the change needed across multiple aspects of the science education environment, policymakers would do well to adopt a “bottom up” orientation toward assessment systems development—that is, one grounded in the classroom—rather than grounded in needs for monitoring, accountability, and teacher evaluation (“top down”). Placing the initial focus on developing high-quality, valid assessments that are as close as possible to the point of instruction will be the best way to identify successful strategies for teaching and assessing science in ways that promote deep learning. Such strategies can lay the groundwork for subsequently developing assessments for monitoring and accountability.

One follow-on implication is that continued use of large-scale science assessments that states developed under NCLB is neither appropriate nor advisable. These instruments are not aligned to college- and career-ready science standards and thus will not support the desired changes in teaching and learning.

Interim solutions will be needed that can simultaneously satisfy federally mandated testing requirements and allow space for change in classroom practice. When external, on-demand assessments predominate in an assessment system and are the sole basis for monitoring and accountability, curriculum and instruction are most likely to narrowly reflect only the material that is tested.8

The Road Ahead

It is worth noting that there is very limited evidence that accountability policies driven by large-scale assessment have led to improved student achievement.9 In contrast, the positive relationship between classroom assessment and student learning outcomes is well established.10 Assessment that closely aligns with curriculum and instruction and that engages students in the kinds of science learning described in NRC’s Framework will return the focus to what is most important—the direct support of students’ learning.

If they are to implement the new standards effectively, many jurisdictions will need to reintroduce science in the early grades.
How does the conversation around measures of success account for the needs of disadvantaged students? Is subgroup accountability a realistic goal when it comes to the deeper measures of success that you, Pedro Noguera, and Diane Friedlaender have written about?

The deeper measures of success, which often include performance assessments and various kinds of performance tasks, are explicitly focused on equity in several ways. Over the 20th century, a “thinking” curriculum was reserved for a limited number of students who were thought to be going into more knowledge-based occupations while others were just getting basic skills to prepare them for factory line or menial work. Disproportionately, low-income students and students of color were not getting access to problem-solving, critical-thinking curricula. An approach that brings this curriculum to all students and engages them all in research, analysis, writing, scientific inquiry, mathematical modeling, and problem solving creates more access to the 21st century curriculum that virtually all kids need today.

Over the years, we have learned how to score these kinds of measures reliably—both in the US and in other countries where they are more commonly part of the assessment systems. We can score performance tests reliably for large-scale purposes, and we can score them reliably for small-scale and local purposes. And they can be reported with individual and collective scores overall and by subgroups, just like any other assessment can be. So you can also monitor progress for all groups on their performance within these types of measures, just like you could with a multiple choice test.

Will performance assessment results be as easily communicated to parents in those populations so that it is actionable for them?

It certainly can be. Advanced Placement courses and tests in world languages, English language arts, and the arts all [have] open-ended tasks, essays, portfolios, and other types of tasks that are scored reliably and are reported in ways that people understand. Performance tasks are being introduced in the AP science courses now, and they have long been present in the International Baccalaureate program. You can analyze these data by different groups of students, by different schools, etc. There is nothing about performance assessment that makes it “un-scoreable” or “unanalyzable.”

Think about Singapore’s high school examinations: Those students design and conduct science investigations over a couple of months. They get numerical scores for those investigations, the same way they would for sit-down essay tests. These data can be analyzed by the kinds of students or kinds of schools, by the kind of instruction they have. And that is what we need. It does cost a bit more to score these kinds of tests, because most of the time, real people have to do it, but the cost is also producing benefits for student learning and teacher learning.

Part of the benefit is that, as teachers are involved in scoring that kind of task or test, they become more knowledgeable about the standards and what they mean in operation, how to teach to the standards, what kind of work is proficient, and how to develop that
work. So they become better teachers as they become involved in scoring. A lot of studies show that teachers find scoring of performance assessments to be excellent professional development. It helps them design their instruction and their curriculum in the classroom in ways that will enhance student’s success.

What steps should states be taking, or thinking about taking? What’s the low-hanging fruit?

If state boards are interested in developing performance assessments in their states, they are probably going to want to engage in some state-supported work and some local work. The Innovation Lab Network created the Performance Assessment Resource Bank that a number of states are using—New Hampshire, Colorado, California, Virginia, and others. Some states have teachers developing tasks that come into the bank; they and others can also download tasks. They can use those for state assessments or local assessments as part of a new system, and some teachers simply use them in their classrooms. These are matched to standards, topics, and concepts and have been through a quality control process. So there are tools to get started.

Policymakers need to think about where this fits in their assessment systems. In New Hampshire, which has a waiver from the federal government under ESEA flexibility, pilot districts use a traditional state sit-down test in some years (the Smarter Balanced test) and then they also have what they call common tasks. These are performance-based assessments used across all the districts that are included in the waiver, which are eventually going to be used statewide. They have kids engaged in these rich tasks; they then organize teachers to score the tasks in a way that will produce consistent scores. Beyond that, the state expects local districts to choose existing tasks or create their own tasks for other subject areas—for example, in history and social science, or science in years when there is not a state assessment. So the kids are getting a steady diet of performance-based, project-based learning. They are engaged in investigations and inquiry in multiple disciplines, writing, speaking, communicating, and using technology.

So a short-term step would be tapping into an existing bank, a medium-term step is thinking about how this fits within a state’s comprehensive assessment and accountability system, and a long-term step is making sure educators are using this information to improve instruction.

Right. A lot of states have started to create graduation competencies, and a lot of districts have begun to create a list of competencies that reflect the 21st century skills they want to be sure students have. A number of schools and networks of schools and districts have begun to create assessments to measure those competencies and portfolios of student work that are evaluated against those competencies. When they reach a high enough standard to be part of the portfolio, they use that as evidence that they have accomplished certain competencies. Quite often, states will say to local districts, “Look, here are the competencies we will agree to. You need to develop a system locally to assess these competencies,” and often that takes the form of a portfolio. And so another thing that [can be] shared are portfolio systems. That is another way states can begin to encourage higher order thinking and performance.

Having teachers score these kinds of performance tasks together, as many schools do, is helpful. For example, where they are involved in a portfolio assessment [group] like the New York Performance Standards Consortium, teachers are creating a shared practice around the work that students are engaged in and the standards that are used to judge that work. They calibrate that with each other and share curriculum ideas to improve the work. You can encourage better practice through that lens. You can further improve collaboration and collegiality because teachers are designing and scoring the work together. A good example of this is the Quality Performance Assessment group in New England, where many educators are doing this kind of work schoolwide.

How do you think this focus on performance assessment affects how we evaluate teachers and leaders and how we support them?

Districts should use some standards-based observation tool to look at teachers’ practices, in order to have good evidence of what they are doing. We should also want to understand something about what kids are learning. And we should want to understand how [teachers and leaders] are working collegially to improve
the school. This kind of assessment addresses all three of those areas.

You can also use the data from performance assessments to inform the evaluation of teachers and leaders, where they are setting goals and collecting evidence of student learning, which illustrates that they are making progress toward those goals. These kinds of tests can provide some of that evidence.

When we are talking about teacher and leader effectiveness, we not only want to measure it, we want to develop it. So we want the tools used in evaluation to cause teachers to deepen their instruction, improve their curriculum, and support students more completely.

Is this the instrument to deal with questions of hiring and firing at the district level?

Districts have different data that they use to hire teachers initially. But subsequent personnel decisions have to be made based on a holistic appraisal of a teacher’s or leader’s accomplishments. We got off on a problematic foot when districts were asked to rate teachers based on a value-added test score. Researchers have found those scores to be unreliable, unstable, and inaccurate, particularly for teachers whose kids are in the groups that traditionally score above or below the grade level. In addition to the fact that there are many things that affect test score gains beyond teachers—for example, administrative and instructional supports, curriculum materials, class sizes, parent supports, and home contexts—the learning gains of some groups of students cannot be measured fully on current state tests because they lack items above and below the grade level.

Instead, we are rating teachers against each other with a tool that is not accurate. We have so many unexpected and unintended surprises. For example, in some states and districts virtually all the teachers of "gifted and talented" students were rated low on value added because the kids had already maxed out on the tests and couldn’t go any higher than 100 percent. The decision of retaining or firing a teacher should be based on a holistic appraisal of what they do in the classroom, in the school, how their students learn on multiple indicators, and how they interact with students, parents, and others. Those processes [should be] well informed by rich data. If we just focus on getting one thermometer score, for example, we can miss some very important things and fail to diagnose the problem, if there is one.

During the pre-NCLB era, Massachusetts continued to evaluate student thinking, writing, and communication on its assessments. Connecticut and Kentucky held out for a long time with their curriculum-embedded performance assessments, even though the law and the then Department of Education’s way of implementing [ESEA] were making it very difficult. Right now, we are rebooting. The assessments that Smarter Balanced and PARCC have created are richer than the tests most states were using; they include more performance-oriented tasks.

California, New Hampshire, and Iowa were involved in a project that had teachers score the Smarter Balanced performance tasks together and reflect on the implications of those tasks for their instruction and curriculum. Teachers gave that experience rave reviews. More than 95 percent said they learned more about the standards, they learned more about the assessments, that it helped them improve their instruction and their curricula, and they feel that teachers should be involved in scoring these tasks. States can consider this kind of practice when they issue contracts for administering their state tests. In California, for example, teachers will be asked to score the performance tasks, and this will be emphasized as a part of the new system.

So you envision this as the beginning of the system of assessment for states and not the whole system of measures?

Exactly. For many states, that will be the spine of the system, which will include other, deeper measures that supplement and augment what you can learn from a sit-down test. States like New Hampshire, Kentucky, and Colorado are augmenting those systems with performance assessments that may be the harbingers of the next era of deeper learning. In the real world, people do not give you a sit-down test to gauge and shape your activities. You need to know how to plan and implement some kind of investigation or problem-solving activity, including sometimes a collaborative activity. We need
these kinds of assessments in our systems as well because students need to learn what they are going to be doing in college and careers.

There is a very interesting new Advanced Placement senior project/seminar, in which kids design and conduct a collaborative project. There are different papers, projects, and activities that students need to do that are scored by teachers. The folks at the College Board are training teams of teachers to offer this new course, and they are overwhelmed by the number of requests to bring this new AP course into a range of schools. They can only incorporate a certain number of teachers each year because they have to train teachers to do it, so they have more demand than they can meet right now. And for the first time, the demographics of the students being served by this course match exactly the demographics of the students in the country. So this new course is getting much wider reach. That is the direction of the future. You see it already in International Baccalaureate, which is really the gold standard for project-based learning and performance assessment. These kinds of tests help teachers to be better instructors, allow kids to do more complex work, and help create systems to incentivize and evaluate national progress toward more productive learning.

**How can states avoid making it seem like performance assessment is just a fad that people can ride out?**

If you are going to initiate complex strategies like this, it is going to take time, and people need to understand that. There will be folks who will raise their hands right away to volunteer, saying, “Please, we want to be involved in the early piloting and testing.” That’s a particularly good strategy if people can watch and be informed by their colleagues across schools about what is working. It is important that these and other assessments not be used for sanctions and punishment and are [couched] clearly in terms of improvement.

What I have seen in so many of the schools that I have worked with are lots of kids saying, “This is so much better than a multiple choice test. We really like doing these projects, we really feel like we are learning.” As long as they have their learning supported and have the time to do it well, they are enjoying it, and their teachers feel enriched by the experience.

**In your 51st State accountability paper, you discuss “reciprocal accountability.” Can you elaborate on it here?**

We are coming out of an era where accountability was conceptualized as very top down: The state holds schools and teachers accountable for achieving test score gains or other outcomes. Reciprocal accountability says that everyone in the system has some responsibility. The state’s responsibility is to provide useful guidance and adequate, equitable distribution of resources for schools and districts to meet the standards that the state has articulated. The districts then have to organize those resources in intelligent ways to accomplish the goals and evaluate how they’re doing. Further, they have a responsibility to look at data around students and schools so that they can manage and continue to improve the system. Schools have to use the resources well and fairly on behalf of the kids, to invest in developing high-quality instruction, and to manage the process of engagement with parents. Teachers have the responsibility to teach productively toward the standards, to take advantage of the professional learning that the states and districts are responsible for providing, and to continue to help the students to reach further. Students and parents have to take advantage of educational opportunities that are available to them and to work collaboratively with educators. Reciprocal accountability is thinking about how everyone in the system is accountable and responsible for the parts of the system they can control, and how everyone is responsible for continuous improvement.

**Are there examples of places moving in that direction?**

Singapore is a really interesting place, and Finland is another. They have this idea of continuous improvement and ongoing evaluation, and they are always [collecting] data about what’s working, what’s not working, and how it is working. They are always giving that data and information back to the school-based educators and those who lead the system, including the ministry of education. They are always thinking about how they can support better, stronger, more equitable learning. Those are good examples of places where we can see tremendous gains in achievement due to the commitment to continual learning.
The Power of Community

The first lesson is sometimes the hardest. New state board of education members often begin their service with great ideas and enthusiasm. They’ll change policy. They’ll affect what students learn. They’ll make things better.

And, of course, they can do all those things. They just can’t do them alone.

As an individual, a state board member has no more authority than any other citizen in the state. But when members are acting as part of the board, they can make policy decisions and influence education in their state.

We find that lesson applies to many other aspects of service on a state board of education. There is power in community.

At our recent Annual Conference in Baltimore, our record-breaking crowd told us they loved our sessions, but they also found tremendous value in the conversations they had over meals and in the halls. It’s great to hear from a state board member in another state about an initiative they tried that was a wild success.

Hearing about what did not work can also be helpful. State board members are willing to share their “bark-off” stories: We did X, but we’d never do it again.

There’s power in community on other levels as well, and the reauthorization of the Elementary and Secondary Education Act (ESEA) offers a case in point. For the last several years, education groups have tried to persuade members of Congress that this version of the bill, which became law in 2001, was outdated.

There were several efforts at reauthorization over the years. But if a bill could get through the House, it would die in the Senate. If it got the support of conservatives, then liberals would oppose it.

Which brings us to 2015. When Senators Lamar Alexander (R-TN) and Patty Murray (D-WA) outlined their plan to tackle ESEA reauthorization again, they were met with more than a little skepticism. Some of that skepticism might have come from me.

But they kept at it. And as it became clear that a bipartisan deal might actually be in the works, something even more astonishing occurred among the education groups. We put aside our differences and decided not to let the perfect be the enemy of the good. After all, we reasoned, if Congress could work together, surely we could, too.

So over the last few months, education groups representing teachers, superintendents, and school board members at the state and local levels, combined with parents and other groups of educators, delivered a common message. The bill, we all agreed, was not perfect. But passing it was so important that we were willing to support compromise wording.

Now we know the story has a happy ending. With passage of the 2015 Every Student Succeeds Act, we created powerful change by working together. We nudged our paralyzed national capital into acting for our children’s good. And if we can do that, there’s no telling what else we might achieve together. ■

from the
Director’s Desktop

Kristen Amundson
Executive Director
My state board of education is a NASBE member. So what do I get?

- **NATIONAL MEETINGS** held every year: Annual Conference, Legislative Conference, New State Board Member Institute
- **REGIONAL MEETINGS**, such as ones held recently in St. Louis on standards-based leadership and in Pittsburgh on science standards
- **CONVENINGS** of states that receive competitive NASBE grants
- **NASBE STAFF VISITS** tailored to the needs of specific state boards: on standards-based leadership, school climate, student data privacy, deeper learning, leadership development, board governance issues, strategic planning, and more
- **CONNECTIONS WITH EXPERTS** through publications such as the State Education Standard, webinars, e-newsletters, conference calls, and face-to-face meetings
- **NATIONAL VOICE** on federal education matters before the administration, Congress, and the US Department of Education
- **OPPORTUNITIES TO SERVE** on association committees and NASBE’s board

“We don’t know what we don’t know. In joining NASBE, I discovered what it meant ‘individually’ to be a responsible, accountable member of a state board of education and how ‘collectively’ we can be a more effective, efficient board in our state. I had the opportunity to network with knowledgeable individuals from all over the country on common issues and challenges and have built lasting friendships.”

—Gayle Manchin, past president, West Virginia Board of Education

Be NASBE. Be engaged.

Contact Kristen Amundson, NASBE’s executive director at kristen.amundson@nasbe.org, with membership questions.

www.nasbe.org
How can states help ensure every school has an effective principal?

A NEW REPORT WITH AN INFOGRAPHIC OFFERS SOME IDEAS.

Developing Excellent School Principals to Advance Teaching and Learning: Considerations for State Policy

School principals are “invaluable multipliers of teaching and learning in the nation’s schools,” according to this report by political scientist Paul Manna, but to date it’s been unclear what state policymakers could do to boost their effectiveness. This report aims to fill that gap by offering guidance to states that want to take action.

Download this report and other resources, free at WALLACEFOUNDATION.ORG