

The State Education

September 2020

# Standard

*The Journal of the National Association of State Boards of Education*

Test-Based Accountability  
COVID and Through-Year Tests  
Alternate Assessments

Pros and Cons of  
Performance Tasks

Innovative Pilots

$$\frac{m_1}{m_2} = e^{v/I}$$

## NEXT-GENERATION ASSESSMENT





“Above all, we will apply a lens of equity and excellence to everything we do so that our focus and that of our member boards remains fixed on ensuring that all students get the resources they need to succeed.”

—“Our Members, Our Mission: A Strategic Plan for the National Association of State Boards of Education” (Fall 2018)

# NASBE

National Association of  
State Boards of Education

## Mission Statement

**NASBE develops, supports, and empowers citizen leaders on state boards of education to strengthen public education systems so students of all backgrounds and circumstances are prepared to succeed in school, work, and life.**

## Core Values

**We believe in equity and excellence:**

We believe that students of all races, genders, and circumstances deserve the supports they need to thrive in school.

We believe all students can learn at high levels and must have the opportunity to do so through state policies that address their diverse learning needs.

**We prioritize state leadership of public education:** We believe that state boards of education, in partnership with chief state school officers, governors, and legislatures, are best positioned to craft, promote, and oversee state education policies.

**We empower citizen leaders:** We build the capacity of state board members to question, convene, and act boldly with and for students, educators, and families.

**We elevate evidence in policymaking:** We ground our resources, tools, and services in the strongest evidence available and help our members make informed decisions in the best interest of students.

**We build community:** We facilitate productive dialogue and nonpartisan exchange among members with diverse perspectives, creating lasting channels for the sharing of evidence-based strategies and solutions.

**We collaborate:** We promote respectful collaboration, both among our members and between our members and the broader education community.



# Standard

VOLUME 20 ■ ISSUE 3

## features

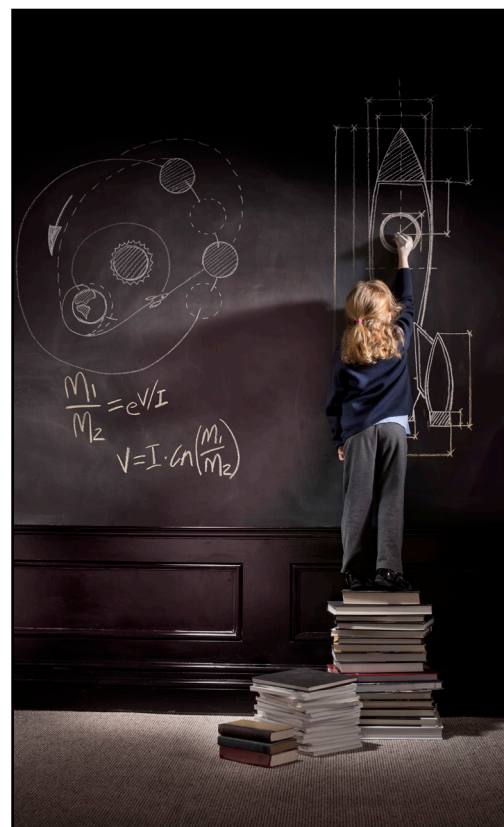
- 6 **A Shifting Landscape for State Testing**  
*COVID-19 adds complexity to an already fragmented assessment market and states' willingness to try new approaches.* LYNN OLSON
- 12 **Four Test Questions for State Boards**  
ABIGAIL POTTS
- 13 **Test-Based Accountability in Distressed Times**  
*Stick with results-based accountability, because few if any other strategies produce such improvements in achievement.* CHESTER E. FINN JR. AND ERIC A. HANUSHEK
- 18 **Breakthrough or Breakdown? School Accountability in Flux**  
*Time to steer systems toward better balance and coherence.* CHRIS DOMALESKI
- 24 **How a Crisis Can Transform Learning, Teaching, and Assessment**  
*Through-year assessments give teachers news they can use and provide a hedge against disruptions.* ABBY JAVUREK AND JASON MENDENHALL
- 31 **We Should Listen to the Canaries**  
*Lessons can be learned from existing efforts to provide alternate assessments for English learners and students with disabilities.* ELLEN FORTE
- 35 **Performance Assessments: Promises and Pitfalls**  
*Performance tasks add great depth but come at a cost.* MARIANNE PERIE
- 40 **States Experiment with Assessments through Innovative Pilots**  
*Five states take federal officials up on their offer to support assessment innovation.* JOSEPH HEDGER

# NASBE

National Association of  
State Boards of Education

## departments

- 2 editor's note
- 3 news & notes
- 4 ncosea voice
- 5 student voice
- 51 we the media
- 52 from the president's pen



Cover Photo Credit: iStockphoto

# Editor's Note

Valerie Norville  
Editorial Director

Bowing to the realities of the pandemic, states halted summative testing in spring 2020, with a federal blessing. For some, this was cause for celebration in the midst of misery—a strike against reviled standards and the tests they spawn. For others, it was wind in the sails of changes already under way in statewide accountability systems—to measure things other than math and reading performance, for instance, and to give teachers timely information on student gains within a school year.

But regardless, for the handful of state education leaders responsible for monitoring how the system overall is providing equitable instruction, there is no substitute for a comparable test that compares how schools, districts, and subgroups of students are doing at a snapshot in time and allows states to make decisions on how to direct scarce resources. In the face of such flux, how will state boards make decisions about assessment and accountability?

The authors provide useful context for these discussions. Lynn Olson of FutureEd frames the issue with a succinct scan of 50 years of federal policy, its attendant backlash, and the fracturing of the testing marketplace, and she projects how learning loss due to school closures will drive what's ahead.

Chester E. Finn Jr. and Eric Hanushek combine research and state board expertise to make their case that state boards ought to stand pat and resume their assessment regimens in spring 2021. To do otherwise “means flying blind, uncertain whether schools are following a flight plan and getting close to the intended destination,” they write.

Two articles press for a shift from status quo. Chris Domaleski from the Center on Assessment urges states to rebalance their systems, with attention to

coherence, reciprocal support, differentiated accountability, and assessment utility. Abby Javurek and Jason Mendenhall from the nonprofit assessment provider NWEA urge state boards to welcome the chance to rethink how their states handle assessment. They write that state leaders can map out a system that recognizes that students progress at different paces, addresses learning loss, and lets teachers modify instruction during a school year in response to interim assessments.

Another set focuses on what state boards can learn from different types of assessment. Ellen Forte of edCount argues that alternate assessments for students with disabilities and English learners can open a window for what assessment could become for all students. “Instead of debating how many tests are appropriate,” she writes, “state leaders might instead ask why there are not more tests that involve teacher-student interactions, administered at times and in ways that directly support students’ learning.”

Marianne Perie of Measurement in Practice details the history and state of play for performance assessments, and she suggests questions state boards should ask before adopting them statewide.

NASBE’s Joseph Hedger offers a glimpse of the range of plans brewing under the federal Innovative Assessment Demonstration Authority program. Participating states include Louisiana, New Hampshire, Georgia, North Carolina, and Massachusetts. And Abby Potts, also of NASBE, provides some questions to usefully frame the conversations that state boards will inevitably have about the assessments they are asked to approve and the role they play in state accountability systems. ■

The State Education

# Standard

Volume 20, Issue 3 September 2020

## NASBE Staff

President and CEO: *Robert Hull*  
Editorial Director: *Valerie Norville*  
Communications Director: *Renée Rybak Lang*  
Associate Editor: *Joseph Hedger*  
Art Director: *Gina Addison*

## Officers

Chair: Brooke Axiotis, *Iowa*  
Chair-Elect: Byron Ernest, *Indiana*  
Past Chair: Rachel Wise, *Nebraska*  
Secretary-Treasurer: Angelika Schroeder, *Colorado*

## Area Directors

Central: Vic Lenz, Missouri; Christine Benson, *Illinois*  
Northeastern: Karen Williams, *District of Columbia*; Audrey Noble, *Delaware*  
Western: Janet Cannon, *Utah*; Ronald McNinch, *Guam*  
Southern: Stephanie Bell, *Alabama*; Sarah Moore, *Arkansas*  
New Member Representatives: Jason Dean, *Mississippi*; Fern Desjardins, *Maine*

## Ex Officio Members

National Council of State Board of Education Executives (NCSBEE) President: Alicia Williams, *Arizona*  
National Council of State Education Attorneys (NCOSEA) President: Tiffany Lucas, *North Carolina*

## Editorial Advisory Board

Stephanie Bell, *Alabama*  
Wilson Hess, *Maine*  
Peggy Hill, *Kansas*  
Patsy Koch-Johns, *Nebraska*  
Donna Leak, *Illinois*  
Martin West, *Massachusetts*

The *State Education Standard* is published periodically by the National Association of State Boards of Education, 123 N. Pitt Street #350, Alexandria, VA 22314. Copyright ©2020 by NASBE. All rights reserved. ISSN 1540-8000. The opinions and views expressed in this journal do not necessarily represent those of NASBE.



# News & Notes

Congress adjourned for the long summer recess without providing additional funding to help states and localities respond to the pandemic, including schools. Legislators hoped to reach an agreement with the White House on a new emergency response package that would provide more help to schools prior to the beginning of the new school year. Negotiations between the parties broke down, however, and congressional leaders sent members home in early August. At press time, Congress had not yet resumed talks about how and if the federal government should provide additional emergency funding to communities.

In July, the House of Representatives approved a fiscal year 2021 appropriations bill for Labor, Health and Human Services, Education and Related Agencies. It would provide \$73.5 billion for the U.S. Department of Education and related programs (a proposed increase of \$716 million from FY20 enacted levels). Most of this increase would go to ESEA Title I and IDEA Part B, but modest increases were also allocated to the Student Support and Academic Enrichment block grant and other K-12 programs. However,

Senate Appropriations has not yet acted on its spending bill for ED, and the Senate may not have enough time to approve it and resolve differences with the House bill by September 30, in which case legislators will be forced to extend the agency's funding at FY20 levels and carry the budget process over to a lame-duck Congress (or later).

Former Vice President Joe Biden's presidential campaign released recommendations in July on key domestic policy issues, including education. The Democratic National Committee worked to translate aspects of these recommendations into the party platform, unveiled ahead of the August convention.

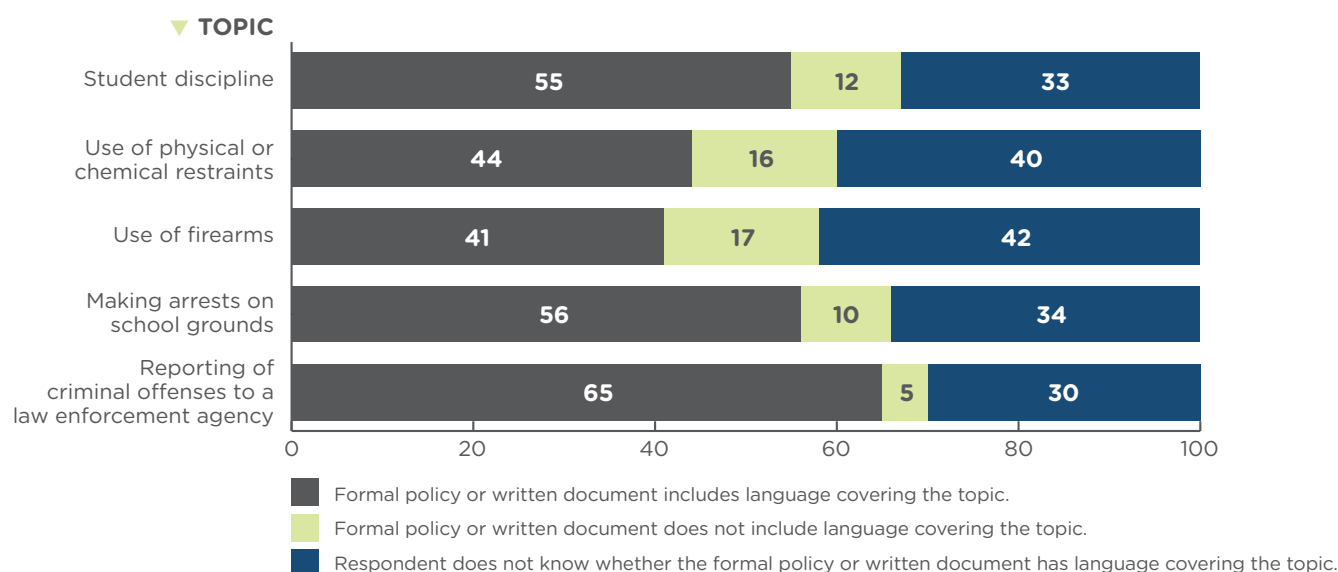
Biden proposes tripling Title I funding under the Every Student Succeeds Act, adding restrictions on charter schools, and revisiting federal assessment and accountability requirements. In June, the Republican National Committee voted to retain the party's 2016 platform, which strongly supports school choice, local control, and reducing the overall federal footprint in education.

The Centers for Disease Control (CDC) released supplementary school guidance regarding the reopening of K-12 schools. Following a weeks-long push by the Trump administration to urge schools to physically reopen and questioning existing CDC guidance for schools, CDC Director Robert Redfield committed to providing "additional" guidance for K-12 communities.

In July, the Department of Education's Institute of Education Sciences invited public comment before September 14 on a study of the Innovative Assessment Demonstration Authority (IADA) Pilot Program, authorized by the Every Student Succeeds Act. The IADA lets the department exempt certain states from some ESSA testing requirements, provided they pilot new assessments within the state. This study will produce a "progress report" and a "best practices report" to help the department target technical assistance in pilot states. The department also plans to provide a guide for other states to develop new assessments.

*Thanks to Foresight Law + Policy for this update.*

**Figure 1. Formal Policy on Roles of Officers in Schools, 2017-19 (percent of responding principals)**



Source: U.S. Department of Education. National Center for Education Statistics (NCES), 2017-2018 School Survey on Crime and Safety (SSOCS), 2018.





**Holly Farris**

Past President, National Council  
of State Education Attorneys

# NCOSEA Voice

## *The Silver Lining of COVID-19*

In the face of the stress and anxiety caused by COVID-19 and the adjustments to daily life that are required of each of us, there have been countless examples of people coming together, helping neighbors, and trying to find ways to make the unfamiliar part of a daily routine.

Lawyers are no different. I have been privileged to see the community of members of the National Council of State Education Attorneys (NCOSEA) come together to confront the challenges presented by the pandemic and the resulting social and legal changes. In the immediate aftermath of so many states and school districts closing schools, first temporarily and then for the remainder of the school year, there were many questions. Foremost was to ask how schools could ensure that students were receiving required services, in addition to the supports that many students and families rely upon, such as meals or special education services. With the uncertainty that COVID presented on so many fronts, there were many questions with no precedents. Others were time-sensitive: State officials needed to make sure that states provided uninterrupted services to the best of their abilities.

In late March and April, I saw the NCOSEA listserv light up with questions, information sharing, and offers to provide feedback. As we moved further into the pandemic's timeline and as the CARES Act moved through Congress, our conversation only intensified. It eventually ended up where everyone seemed to end up during the pandemic: on Zoom, with NCOSEA collaborating with the National Association of Attorneys General to conduct two webinars with the

largest turnout for NCOSEA programming in many years.

NCOSEA members were selfless, sharing template documents or guidance issued by their state education agencies (SEAs) as informational resources, with the only goal being to provide as much help to states and students as possible, and not for credit or professional acclaim. For example, guidance was shared on states' responses to the U.S. Department of Education's proposed rules regarding the distribution of CARES Act funding. Members dissected the unique considerations for accommodations of special education students during the pandemic. Now, as we move into the new school year, members are tackling the legal ramifications of potentially requiring face masks or face shields in schools.

While it took a global pandemic to muster the network to this degree, it was incredibly gratifying to see the way NCOSEA members responded to help not just their SEAs but each other. It shows on a larger scale what I have experienced on an individual basis as an NCOSEA member since I joined in 2014. As an attorney in a small state where SEA attorneys are few, I can personally attest that a resource like the NCOSEA listserv and the network of attorneys across the country it represents are invaluable.

I hope you have seen similar silver linings in your networks as well. Stay safe and healthy! ■



# Student Voice

## *Teaching Students to the Test*

As a high school freshman in a magnet STEM school, I felt the pressure to succeed from all sides. In the rat race for good grades, I never enjoyed class. Later that year, I realized that every high school student suffered the plague of stress culture—cramming for tests and feeling their self-worth was bound up in test scores.

Education should be fun and collaborative. Instead, it produces anxiety and fails to encourage students to love learning. Even when adults ask students for their perspectives on school climate, student feedback is never well received. Thus student governments are essentially party planners, making decisions on prom decor and homecoming floats but with no power over curriculum, coursework, or the pace of learning.

This system continues to fail millions of high school students. Four reforms will help transform the outdated education system.

**Leverage students' optimal learning styles.** Mine involved kinesthetic, visual, and auditory learning, which were part of elementary school but hardly ever revisited in later schooling. Use of pictures, projects, and hands-on learning ought not be phased out as students progress to middle and high school. Education researcher Hilary G. Conklin stresses students' needs for creativity through play, which is stifled as students get older.<sup>1</sup> Even in the era of COVID-19, where teachers and students have had to adapt to online learning, there are ways to cultivate a love of learning rooted in student preferences.

**Cater to student needs.** The classroom should be a safe space for students to take risks, without fear that their ideas will

be ridiculed or dismissed. It is crucial to recognize that everyone may not achieve the same level of understanding from concept to concept and to encourage everyone to try their best. Mandating benchmark tests against a set of standards at the beginning and end of the year risks deemphasizing individual student needs. The key is for teachers to use test scores to see where students began and to use subsequent exams, projects, or other hands-on mechanisms to determine student progress toward understanding.

**Use positive metrics.** Some schools group students based on their levels of understanding. Honors courses or teachers' announcements of the highest class test scores may encourage some students to excel, but it will not produce a growth mind-set in all students. I never made the top five in honors or AP calculus, and I felt my attempts were never good enough, despite earning high B's. Teachers could foster engaged, reflective learning by letting students set independent goals. They could share only average scores to encourage classes to grow as collaborative communities.

**Be responsive to student feedback.** Students must help shape their education. Providing a place for continuous student feedback on content pacing and understanding is paramount to engagement and success. Students should have a voice in curriculum and standards development as well. Until the system begins to treat students like people, students will continue to be left behind. ■

<sup>1</sup>Hilary Conklin, "Playtime Isn't Just for Preschoolers—Teenagers Need It, Too," *Time* (March 3, 2015).



**Dorcas Olatunji**

Former Student Representative  
on the Delaware State Board  
of Education

*This column reflects her views alone  
and not the views of the Delaware  
State Board of Education.*







# A Shifting Landscape for State Testing

Last spring's sudden shift to remote learning in response to the COVID-19 pandemic disrupted K-12 education in an unprecedented way. Upon Education Secretary Betsy DeVos's waiver of federally mandated state testing for spring 2020, it also triggered the first nationwide break in state testing in half a century.

With students returning to school for the 2020–21 school year, state boards of education will need to get a handle on exactly where students are in their learning in order to target resources and support. States also will need to restart annual assessments mandated by the federal Every Student Succeeds Act (ESSA) in order to track progress over time, particularly for vulnerable subgroups of students.

Yet even before the pandemic, two recent FutureEd reports found that state testing systems were already in transition. A national analysis by FutureEd found that between 2014 and 2019, lawmakers in 36 states passed legislation to respond to concerns about overtesting, including by reducing testing in a variety of ways.<sup>1</sup>

That trend is likely to continue, as opponents to standardized testing eye the coronavirus crisis as an opportunity to leverage more cuts—and even end state testing entirely. At the same time, interest is growing in innovations that could make state testing systems more useful for teachers and students by moving away from a single, end-of-year test to more frequent assessments that better reflect classroom curriculum and instruction.

As state boards think about the right questions to ask about testing in order to make smart decisions moving forward, it is important to understand the history of state summative assessment in the United States, the increasing fragmentation of the state testing landscape, and the larger political context.

## A Half Century of Testing

Standardized testing has been part of the K-12 education landscape for the past 50 years, with both Democratic and Republican leaders at times pushing for higher standards and greater accountability. During this time, state tests have been viewed as an essential component of educational improvement—providing a window into school performance, helping policymakers map strategies and allocate resources, and ensuring that the needs of underserved students are being addressed.

In 1969, with federal dollars flowing to schools under the Elementary and Secondary Education Act (ESEA), Congress mandated a federally funded snapshot of student performance, the National Assessment of Educational Progress. The same year, Michigan launched the first statewide testing program.

In the 1970s, concerns about the performance of high school students in particular led to the “minimum competency” movement and expanded state testing to ensure that students graduated with the requisite basic skills.

In the 1980s and 1990s, as the nation transitioned to a knowledge-based economy, political leaders began to push for higher educational standards and national goals, as well as efforts to hold schools and districts accountable for results, particularly for the nation's most disadvantaged students. The Clinton administration's reauthorization of ESEA in 1994 required states, for the first time, to adopt state standards that would be the same for all students and to test all students' progress against those standards in at least three grades.

Not all states responded to the 1994 requirements with equal rigor. So when George W. Bush took office, he decided to place significantly more emphasis on tests

*COVID-19 added a layer of complexity to a fragmented market and states' interest in different approaches.*

---

**Lynn Olson**

**NCLB was designed to shed a bright light on educational inequities to counter the “soft bigotry of low expectations.”**

and test results in the next reauthorization of the law, with the goal of ensuring that the needs of students furthest from opportunity were being addressed. The No Child Left Behind (NCLB) Act of 2001 mandated that states test every student every year in reading and math in grades 3-8 and once in high school and in science at least once in elementary, middle, and high school. NCLB also required states, districts, and schools to publicly report test data by race and income. And it set strict timelines for schools to get every student to the proficient level on state tests or face an escalating series of supports and sanctions.

NCLB was designed to shed a bright light on educational inequities to counter what President Bush described as the “soft bigotry of low expectations.” But it had unintended, negative consequences, as states, districts, and schools frequently responded to the pressure in ways that jeopardized student learning and kindled antitesting sentiment.

Schools emphasized instruction in tested subjects at the expense of untested subjects and stressed test-taking skills.<sup>2</sup> School districts piled on new benchmark tests to gauge how students would perform on end-of-year exams. Many states began to rely heavily on multiple-choice tests because they were cheaper and easier to administer in the face of tight testing timelines. And many states lowered their testing standards to get more students over the proficiency bar.

Recognizing the need for more consistent and ambitious standards, a group of states began collaborating to identify the knowledge and skills needed for college and career readiness by higher education and employers. The work laid the foundation for a 2009 agreement by the National Governors Association and the Council of Chief State School Officers (CCSSO) to jointly develop demanding voluntary standards in English language arts and math—what became the Common Core State Standards. That work coincided with the Obama administration’s Race to the Top initiative, which provided billions of dollars in education funding to states to help address the 2008–09 economic crisis.

By making the competitive grants contingent on states adopting the reforms, the \$4.3 billion program incentivized states to adopt

tougher academic standards and more rigorous tests aligned with those standards. Most states quickly embraced the Common Core standards and joined one of two voluntary state consortia to develop Common Core-aligned tests with federal funding: the Smarter Balanced Assessment Consortium or The Partnership for Assessment of Readiness for College and Careers (PARCC).

## The Backlash Begins

The administration’s decision to leverage billions of dollars in federal funding on behalf of higher standards, harder tests, and test-based consequences for teachers brought the national teachers’ unions and Tea Party conservatives to the barricades, if from opposite directions. The Tea Party and its Republican congressional allies condemned the Common Core as a federal usurpation of traditional local control in public education, even though state-facing organizations led the development of the new standards. The unions targeted the new teacher evaluation systems and the increase in teacher accountability they represented, spurred by rank-and-file members’ outrage that their livelihoods suddenly depended on how well their students performed on brand new standards and tests.

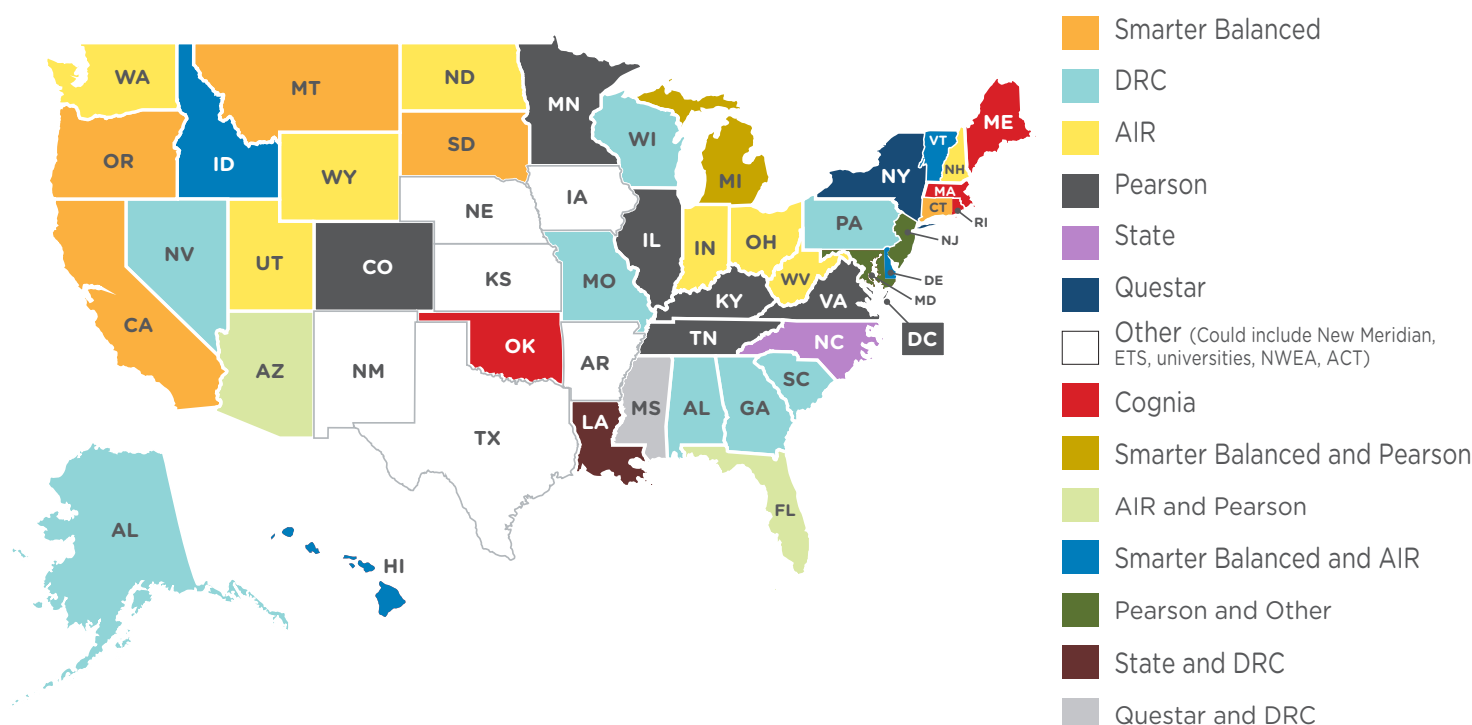
The avalanche of opposition forced the Obama administration to retreat. In August 2015, U.S. Secretary of Education Arne Duncan announced his department’s decision to grant states with NCLB waivers a one-year delay in incorporating student test scores into teacher evaluations. CCSSO and the Council of Great City Schools, representing the nation’s large urban school districts, announced a joint project to throw “their collective weight behind an effort to reduce test-taking in public schools, while also holding fast to key annual standardized assessments.”

In December of that year, the president signed ESSA, which replaced NCLB. Democratic leaders fought hard to keep the requirement for annual state testing of every student in reading and math in seven grades and results disaggregated by race, income, English-learner status, and disability status.

The new federal law gave states and districts far more power to craft their own education



**Figure 1. Testing Vendors for Grades 3–8 by State, 2019**



Source: Lynn Olson, “The New Testing Landscape: How State Assessments Are Changing under the Federal Every Student Succeeds Act” (Washington, DC: FutureEd, September 2019), 4.

solutions. It abandoned NCLB’s requirement that states impose escalating sanctions on underperforming schools. It jettisoned Duncan’s earlier push for states and school districts to use students’ standardized test scores in teacher evaluations. It permitted states to use college-admissions exams—the SAT and the ACT—as substitutes for state high school standardized tests. And it allowed states to explore new ways of testing students under an Innovative Assessment Demonstration Authority.

## The New Testing Landscape

To get a handle on the state of state summative assessments under ESSA, FutureEd conducted a scan of state testing programs across the 50 states and Washington, D.C., based on publicly available data and interviews.<sup>3</sup> We found that the marketplace for summative assessments in grades 3–8 has fragmented as an increasing number of states met ESSA’s accountability requirements with tests designed to reflect their individual state’s content

standards (figure 1). Many states abandoned the two Common Core–aligned state assessment consortia. In 2010, 45 states planned to use tests developed by the consortia. By the 2019–20 school year, only a dozen states remained in the Smarter Balanced consortium, and only D.C. was using PARCC exams.

At the high school level, more states are embracing the ACT and the SAT as their assessments despite concerns that the tests are not fully aligned with state standards. Language permitting this option under ESSA has accelerated the trend, enabling states to offer tests that parents and students actually care about. (It will be interesting to see if this trend continues now that the University of California and other higher education institutions are no longer requiring the exams for admissions.)

While states and the federal government still prioritize the ability to compare student test results within a state, comparing results across states has faded as a dominant goal. That said, because of PARCC and Smarter Balanced, there’s evidence that both the quality and rigor

**Because of PARCC and Smarter Balanced, there’s evidence that both the quality and rigor of state tests have risen.**

**States have shown growing interest in assessment systems that better reflect and support the daily work of students and teachers.**

of state tests have risen. Cut scores—the benchmarks that states set for student performance—appear to be holding the line in most places, but it is harder to get a handle on test quality.

In this atmosphere, churn in state testing systems is a big concern. While no one is tracking how many states have changed their tests or assessment vendors multiple times in the past five years, many of those interviewed mentioned it. Constant changes in state assessment systems make it harder to track performance over time, create problems for state and district accountability systems, and send mixed messages to educators, diminishing their morale and ability to focus instruction.

### **Growing Interest in Innovation**

At the same time, states have shown growing interest in designing assessment systems that better reflect and support the daily work of students and teachers in classrooms. Such systems would deliver faster turnaround of test results, as well as greater use of end-of-unit tests, performance-based tasks that ask students to apply what they know and can do, and tests that are more closely linked to the curriculum. Such efforts could provide better ongoing information about student progress, while giving teachers more guidance on how to adjust instruction. To date, five states—Georgia, Louisiana, New Hampshire, Massachusetts, and North Carolina—have been approved for the federal innovative assessment pilot (also see article, p. 40).

Louisiana, for example, is developing ELA/social studies tests to be given three times a year at the end of units that can be rolled up into an end-of-year summative score. The online assessments will enable districts to choose from a set of texts organized around key topics. Students will respond to writing tasks that require them to make meaning of texts they have already studied or of texts that are new to them but closely related to the curriculum. This dovetails with the state's efforts in the past decade to adopt higher standards and an optional text- and content-rich ELA curriculum (ELA Guidebooks 2.0) with units built around general themes, knowledge domains, and “anchor texts.”

New Hampshire is taking a different approach: piloting classroom-embedded

performance tasks in English language arts, math, and science to replace annual state tests with the goal of providing students a richer, more individualized learning experience. Schools and districts participating in the Granite State's innovative assessment pilot—Performance Assessment of Competency Education, or PACE—supplant much of the traditional end-of-year state testing with teacher-developed performance tasks. These include one common task in each grade and subject combination without a state test, which participating districts agree to collaboratively develop and administer.

### **Legislative Response to Overtesting Concerns**

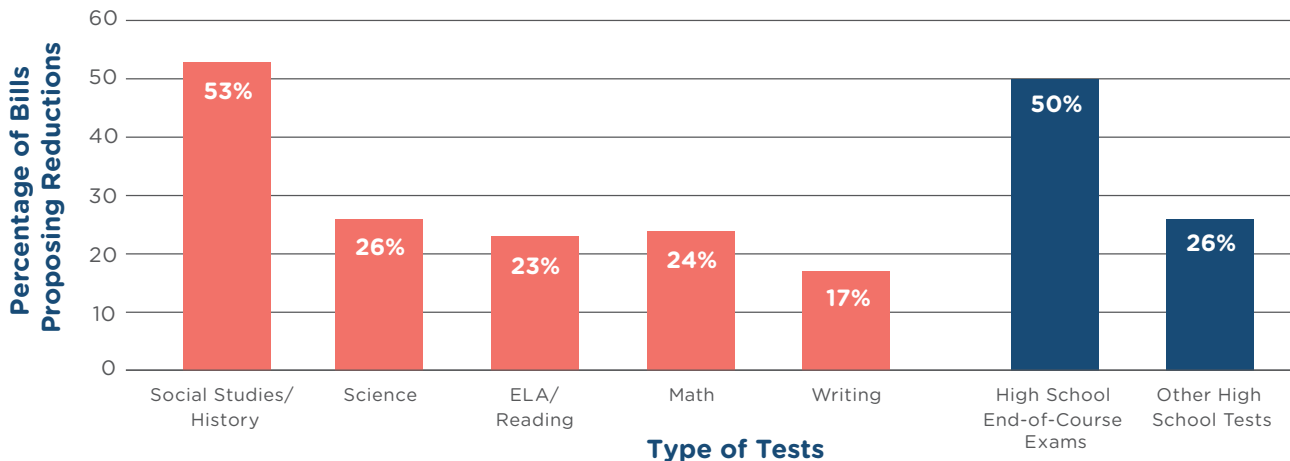
Given the testing climate in recent years, ESSA has become a bulwark against further reductions in the measurement of school performance, even as Secretary DeVos suspended the law's requirements for 2019–20. But a close analysis of the political landscape of standardized testing makes clear that unless a new generation of tests can play a more meaningful role in classroom instruction, and unless testing proponents can again convince policymakers and the public of the value of state testing for school improvement and educational equity, annual state tests and the safeguards they provide are at risk.

From 2014 through 2019, lawmakers introduced no fewer than 426 bills and 20 resolutions in 44 states in response to critics' claims of overtesting (figure 2), and measures were adopted or enacted in 36 states.<sup>4</sup> There were more bills in 2019 than in 2018, an indication that antitesting sentiment remains strong in state capitals five years after the signing of ESSA. This analysis excludes dozens of parental “opt-out” bills that in most instances granted students unrestricted rights to sit out state tests. And it doesn't reflect moves to reduce testing in many states in recent years by governors, state boards of education, and state education agencies.

Lawmakers' most common legislative response was to reduce the number of state tests students must take. In other instances, they shortened the length of tests, capped standardized testing time in schools, required public reporting of testing time, or directed



**Figure 2. Targets of Measures to Reduce State Testing, 2014–19**



Source: Lynn Olson and Craig Jerald, “The Big Test: The Future of Statewide Standardized Assessments” (Washington, DC: FutureEd, April 2020), 10.

state agencies or local school districts to limit testing. A quarter of the 167 bills to reduce state testing demanded the discontinuation of every test not required by ESSA. Others targeted tests in grades and subjects not covered by federal law—particularly social studies. More than half the test-reduction legislation involved at least some social studies tests or high school exams in social sciences such as history. And half targeted high school end-of-course tests.

FutureEd’s analysis confirmed the bipartisan nature of the legislative action against standardized testing. Sixty-eight of the measures introduced in 2019 were sponsored by individuals rather than legislative committees. Of those, Republican legislators authored 41 percent, Democrats authored 44 percent, and 15 percent were bipartisan.

## Testing amid COVID

Given the months of time spent outside of school buildings during the 2019–20 school year, the most immediate concern this fall is how to capture where students are in their learning. One study estimated that the average student could lose roughly half of what they were expected to learn in math during the 2019–20 school year and close to a third of what they would have gained in reading.<sup>5</sup>

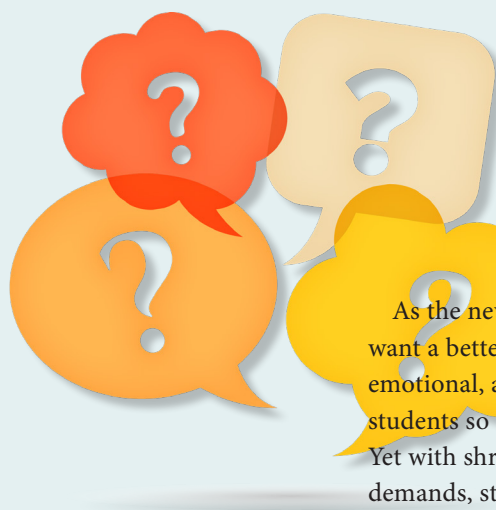
As schools reopen, educators should turn to diagnostic assessments and other measures of reading and math proficiency to help teachers identify individual student needs and know where to pick up instruction. These early-year assessments would be best kept closely connected to the school’s curriculum and used solely to inform instruction and allocate resources.<sup>6</sup>

In the spring, states should restart their annual assessments mandated by ESSA. Just as widespread coronavirus testing will guide a return to normal life, state testing systems have a valuable role to play in helping leaders map education strategy, track progress, and back the nation’s neediest students. But if federal waivers permit, states should not use these results for high-stakes accountability decisions. Attaching stakes to test results too quickly—as schools, teachers, and students are readjusting to the “new normal”—would play into the hands of accountability opponents at a time when we need smart testing more than ever.

Deploying high-quality assessments to support the nation’s students and teachers and to inform policymaking in the wake of the COVID-19 crisis is one of the most significant challenges facing the education sector. I hope the articles in this issue of the *Standard* will help guide state policymakers as they face critical decisions about testing. In addition, FutureEd

**State testing systems have a valuable role to play in helping leaders map education strategy, track progress, and back the nation’s neediest students.**

Lynn Olson is a senior fellow at FutureEd.



# Four Test Questions for State Boards

As the new school year begins, states will want a better understanding of the social, emotional, and academic learning needs of students so they can respond to them rapidly. Yet with shrinking budgets and increasing demands, states are grappling with tough decisions about how to balance giving students time to adjust to a new normal of schooling and assessing the scope and extent of learning loss due to the COVID-19 pandemic.

For example, state boards over the summer wrestled with the prolonged negative impact of learning loss in reading, especially in the early years. Some states, such as Indiana, chose to make early reading assessments optional for the fall, in hopes of giving teachers more time to focus on helping students adjust to new school contexts. Others, such as Colorado, plan to continue administering assessments to ensure that students received interventions to address learning needs.

Four key questions will help guide decision making during these unprecedented times.

## What is the purpose?

An essential first step for a state board is defining an assessment's intended purpose, which is vital to ensuring the reliability and validity of the data. No one assessment can achieve all goals. For example, formative assessments are better suited to exploring individual student learning on discrete standards and guiding the learning process, while summative assessments are better suited for reliably measuring average school performance across student groups. With a clear purpose statement, state boards will be well equipped to ask whether a given assessment validly supports its use for the intended purpose.

## How should the data be used?

Inherent in an assessment's purpose is a clear articulation of the actions or changes

in behavior expected as a result of the data. State boards can build narratives around how to make the data actionable. Who is using the information? What actions are taken as a result? How does this action help achieve the strategic vision of education in your state? Testing by itself does not mitigate achievement gaps. Rather, the actions teachers and administrators take based on the data to better support students are the key levers to achieving the state's vision for education. If the data do not support these actions, then the time spent on assessment is not well used and runs the risk of crowding out instructional time, with negative impacts on students.

## Is the test fair and accessible?

As conveners and questioners, state board members serve a vital role in ensuring the fairness and accessibility of an assessment. As assessments are developed or selected, state boards can pave the way for accessibility by ensuring that experts on students with disabilities, English learners, and culturally diverse perspectives guide the process. By asking about the diversity of design and review committees, state boards can set a state up to identify and resolve issues efficiently rather than adopting the costly strategy of trying to retrofit an assessment that was not designed with the needs of some students in mind.

## What resources and capacity building empower educators to act on data?

Data can be complicated and complex. Supporting teachers' assessment literacy is vital. States can encourage professional learning that explores assessment data and how they can be integrated with other information. State boards can also draw attention to resource inequities and catalyze action to address them. ■

— Abigail Potts





# Test-Based Accountability in Distressed Times

When COVID-19 forced a hiatus in federally mandated assessments in spring 2020, it interrupted a quarter century of effort to track, disaggregate, and publicize achievement levels at the school level. The aborted school year put a big data gap where 2020 scores should have been. Combined with the already-raging assaults on testing, state education leaders find themselves in a fraught and difficult place.

Arguably since Congress passed the Improving America's Schools Act (IASA) in 1994 and most definitely since it enacted No Child Left Behind (NCLB) in 2002, test-based school accountability has played a big role in American public education. Although the latest revision of the Elementary and Secondary Education Act, known as ESSA (2015), lets states make more decisions regarding their school accountability

*State leaders should stick with their assessments because they improve student learning and school performance.*

---

**Chester E. Finn Jr. and Eric A. Hanushek**

**International evidence indicates that countries with testing programs that allow for external comparisons have students who do better on international achievement tests.**

systems—particularly when it comes to sanctions, interventions, and consequences—the assessment requirements did not significantly change in 2015 nor did the obligation to disaggregate and publicize school-level results.

Pandemic-induced assessment waivers and a testing holiday will intensify the longstanding pushback against these requirements. They will embolden test critics and sundry education interest groups to declare, “See, we don’t really need those damned tests. We can’t even use them during unusual circumstances such as at-home study—and when we do use them they distort the curriculum, cheapen instruction, worry teachers, alarm parents, and scare kids. Given the learning gaps and uncertainties presented by the present plague, it would be cruel to go back to using them and whatever results they might show in spring 2021 will surely be misleading.”

We disagree. We will offer state boards of education suggestions on how to proceed in the short and the longer run. But first, we review the history of test-based accountability as it has been practiced and the evidence of its impact.

## The History

NCLB significantly enlarged the role of the federal government in education. Its focal point was evaluation of school outcomes, and it contained incentives for ensuring high levels of overall achievement and broad impacts across all subgroups defined by race, ethnicity, poverty, and language.

The NCLB feature garnering the most attention was the requirement that all students be proficient in reading and math by 2014. As time passed, it became clear that this extremely ambitious goal was not going to be met. School staff led a simultaneous, steady drumbeat aimed at ending the accountability regime altogether. In time, NCLB became a four-letter word to many. What is often not recognized, however, is that 43 states already had their own test-based accountability system in place at the time the NCLB became law in 2002.

In many ways, NCLB had it backward: States were charged with identifying the learning standards and testing regimes to describe student proficiency while the federal government dictated what actions should be taken if

progress was insufficient. The federal government, without knowledge of the demands on or capacities of schools, clearly was unprepared to dictate how all schools should provide education. At the same time, individual states have little way to set standards for the national and international labor markets for which they were preparing students. Moreover, NCLB measured pupil performance by level of achievement without regard for the preparation and readiness with which students came to the schools.

## The Evidence

Despite this inverted structure and the mounting public uproar over the heavy-handed federal role, NCLB actually produced improvements. In both the period of state actions before its enactment and after its passage, accountability systems led to greater student achievement.<sup>1</sup> Moreover, systems that incorporated consequences for schools and personnel had the strongest effect.

These U.S. data are consistent with international evidence that indicates that countries with testing programs that allow for external comparisons have students who do better on international achievement tests.<sup>2</sup> Alternative testing approaches, including school inspectorates, do not lead to better overall student performance when they lack the ability to compare performance across students and schools.

The international evidence also highlights a generally overlooked part of assessments—that not all testing is about schools. Many countries develop testing for graduation and student placement into higher education or other institutions. These exit exams are generally designed to do two things: give students incentives to learn and provide direct comparisons of students across local grading systems. The use of such exams has generally shown beneficial effects across countries.<sup>3</sup> Importantly, the value of local school grades rises when there is an external test that can be used to calibrate the meaning of those grades.<sup>4</sup>

It is often said in arguments against NCLB-like reforms that top-ranked countries such as Finland do not use testing and accountability yet rank very highly on international assessments.<sup>5</sup> But Finland has a national exit exam that is mandatory for all students and



that determines access to tertiary education. Performance on the Finnish Matriculation Examination has strong implications for students and schools.<sup>6</sup>

The United States has also used consequential student exams for purposes other than school accountability. Twenty-four states required exit exams for high school graduation in 2013, with most of their passing scores set at least at the 10th grade level.<sup>7</sup> Moreover, a number of states require students to meet minimum test-based standards in order to be promoted to the next grade. Evidence suggests that both of these testing regimes have had somewhat positive effects on achievement, although there are some mixed findings.<sup>8</sup> Both regimes have also faced considerable political pressure, leading to slow but discernible movement away from such test use.

The objections raised about test-based accountability do not erase the fact that students learn more when there are measures of performance and when schools pay attention to levels of achievement. Some have argued that it did not work because it did not make U.S. schools the best in the world,<sup>9</sup> but that is not a legitimate criterion for evaluating accountability systems.<sup>10</sup> Others contend that NCLB narrowed classroom instruction because it only focused on reading and math. Emphasizing basic academic skills was, of course, part of the design because it is hard to argue that basic reading and math skills are not key to most subsequent learning. At the same time, transcript studies at the high school level show that the curriculum of the typical student in fact became broader and more rigorous in academic coursework with the introduction of state and federal accountability.<sup>11</sup> It's a fact, however, that more time and focus on basic academic skills necessarily means less is available for other parts of schools such as the arts, a trade-off that comes from setting priorities in schools.

One common argument against test-based accountability is that it leads to teachers spending too much time teaching to the test. This would chiefly be a problem if tests fail to reflect what we want students to know. In programs such as Advanced Placement and International Baccalaureate, for example, which are highly

regarded by most teachers, preparing students to ace the end-of-year exams is typically viewed as honorable and gratifying.

NCLB is often decried for being a “high-stakes” testing regime. Yet the state assessments used in the aggregate accountability systems of most states attach no rewards or penalties to student test takers and typically not to their teachers either. To be sure, exit exams and college admission tests carry high stakes for students, but there is generally less (though tangible) political pushback to high-stakes testing that involves students than to the kind that judge schools and educators.

It is true that the availability of state assessments can lead to larger stakes for teachers, because they facilitate linking the performance of students to specific educators. The possibility of evaluating teachers based on how much their students learn has been recognized for 50 years.<sup>12</sup> While it was not part of the original NCLB, Secretary of Education Arne Duncan included teacher evaluation as part of the department's Race to the Top initiative and in a 2012 “education flexibility” offer to states.<sup>13</sup> Continuation of this aspect of accountability—and teacher resistance to it—appears to be the most important pressure for scaling back if not eliminating regular student testing.

The use of student performance to gauge teacher effectiveness requires that the teacher's influence be separated from other factors that bear on student achievement. This separation is generally accomplished through “value-added” analysis. There is a large literature on such analysis, and two generalizations can be made: First, there are significant differences in teacher effectiveness, differences that have huge impacts on students' labor-market prospects and on the U.S. economy.<sup>14</sup> Second, nobody would argue for using value-added scores alone to evaluate teachers. Teachers have wider impacts than what can be measured on state math and reading tests. More important, only a minority of teachers—perhaps a quarter—can be evaluated at all in a value-added context because of the limited subject and grade-level testing in the schools. Thus the drumbeat against “evaluating teachers just on narrow tests” appears to be more a general pushback against any evaluation of teachers, not to the overuse of tests.

---

**Students learn more when there are measures of performance and when schools pay attention to levels of achievement.**

**In fall 2020, states and districts would be well advised to deploy whatever assessments they can tap.**

Currently, 34 states require some measure of growth in student achievement to be used in the evaluation of teachers, but this is down from 43 just five years earlier.<sup>15</sup> This decline, however, is not consistent with the research evidence on the effectiveness of personnel systems that use teacher value-added in their operations. Although understandably contentious, teacher value-added has proved to be useful in contributing to the overall evaluation of teachers when such information is available from student testing programs.

Washington, DC, offers the best example of the influence on student learning from a personnel system incorporating teacher value-added. Its IMPACT system uses value-added analysis for the quarter of the teachers where it is possible, but the largest portion of the evaluation of all teachers comes from an objective observational rating. Based on the overall evaluation, the District of Columbia gives large bonuses to effective instructors and dismisses grossly ineffective teachers, and these steps have contributed to the best achievement gains of a major U.S. city.<sup>16</sup> A similar system in Dallas, Texas, also appears to be significantly improving student outcomes, particularly when it is used to guide staffing in schools with concentrations of disadvantaged students. These are just two examples of how improved compensation policies can be usefully introduced into school policy.<sup>17</sup>

## The Year Ahead

Important technical questions arise when an assessment regime built on an annual cycle—like the familiar school calendar—gets interrupted. Although a simple resumption of testing may still accurately display student achievement at whatever point in time the tests are again given, a year (or more) of missing data will affect trend lines, confound growth calculations, and complicate applications of data derived from them. Psychometricians and testing directors will have their work cut out trying to make the necessary adjustments, adding caveats, and double-checking reliability.

For growth or value-added calculations in particular, a single year of missing data on individual students causes many complications and more than a year will render such calculations

impossible under most circumstances. That missing data will invalidate key elements of most states' ESSA accountability plans, remove valuable information from school report cards, and—if the absence of growth data places greater weight on simple achievement data—will give rise to new equity concerns.

Although ESSA is all but certain to remain the law of the land for some time, it is likely that the U.S. Department of Education will see many more requests for waivers from elements of it and that a number of states will consider amending their approved accountability plans. How that may play out depends, of course, on individual state circumstances, education-governance structures, and political dynamics, but some movements in this direction are all but certain. Meanwhile, there obviously had to be a hiatus in the consequential use of end-of-year test information for evaluating schools, teachers, and students, both because there is a big data hole and because almost no students had a full year of proper instruction.

With the resumption of school in fall 2020—in all the complex forms it is taking—states and districts would be well advised to deploy whatever assessments they can tap to gauge where individual children, groups of children, and entire schools and systems are at the year's start. The cessation of in-person instruction in spring 2020 has obviously produced significant learning gaps but, perhaps as important, has led to much larger variations in the achievement that will be found in most schools during this next round of learning. Data on baseline performance are sorely needed to shape and adapt curriculum and instruction and to align instruction more accurately to needs.

To satisfy formal requirements under their ESSA plans but, more important, to be better able to plan for subsequent years, states and districts should plan to resume their familiar assessment regimen at the end of the coming year. End-of-year data from 2020–21, if available, can be compared with end-of-year data from 2018–19 to calculate two-year changes and to chart the course of what may be a series of interrupted years.

Meanwhile, many of the familiar assessments, both formative and summative, may need to be adapted to a schooling environment at least some of which takes place outside



the school building for at least some students and at least some of which takes place outside traditional school hours.<sup>18</sup> With careful planning, these testing adaptations can yield sturdy, valid, reliable gauges of student learning with minimal glitches, such as those that beset some who took at-home Advanced Placement exams in May 2020. (The College Board deserves kudos for improvising an assessment arrangement that apparently worked for the vast majority of AP students.)

Perhaps most important for state board leaders to understand and communicate to their constituents is this: Results-based accountability for schools and students is perhaps the most impactful education policy that we have. Few if any other strategies produce such broad improvements in achievement.

All should welcome the quest for additional sources for improving student learning and school performance, as do we. But until and unless such sources prove as stable, reliable, and revealing as testing, American education cannot stop testing. To do so means flying blind, uncertain whether schools are following a flight plan and getting close to the intended destination. Put differently, no enterprise can succeed without regular, reliable data on its own performance. Testing cannot furnish all the data that educators and policymakers need, but it is an essential source of indispensable information.

While having good assessment data cannot ensure gains in student achievement, not having good assessment data can virtually guarantee no improvement. ■

<sup>1</sup>Eric A. Hanushek and Margaret E. Raymond, "Does School Accountability Lead to Improved Student Performance?" *Journal of Policy Analysis and Management* 24, no. 2 (2005): 297–327; Thomas S. Dee and Brian A. Jacob, "The Impact of No Child Left Behind on Student Achievement," *Journal of Policy Analysis and Management* 30, no. 3 (2011): 418–46; David Figlio and Susanna Loeb, "School Accountability," in Stephen Machin, Eric A. Hanushek, and Ludger Woessmann, eds., *Handbook of the Economics of Education* (Amsterdam: Elsevier, 2011): 383–421.

<sup>2</sup>Annika B. Bergbauer, Eric A. Hanushek, and Ludger Woessmann, "Testing," *NBER Working Paper* 24836 (Cambridge, MA: National Bureau of Economic Research, July 2018).

<sup>3</sup>Eric A. Hanushek and Ludger Woessmann, "The Economics of International Differences in Educational Achievement," in Hanushek et al., *Handbook of the Economics of Education*.

<sup>4</sup>Guido Schwerdt and Ludger Woessmann, "The Information Value of Central School Exams," *Economics of Education Review* 56 (2017): 65–79; Margaret E. Raymond, "The Diploma Dilemma," Hoover Education Success Initiative

(Stanford, CA: Hoover Institution, February 2020).

<sup>5</sup>E.g., Pasi Sahlberg, "Education Policies for Raising Student Learning: The Finnish Approach," *Journal of Education Policy* 22, no. 2 (2007/03/01): 147–71; Linda Darling-Hammond, *The Flat World and Education: How America's Commitment to Equity Will Determine Our Future* (New York: Teacher's College Press, 2010).

<sup>6</sup>Matriculation Examination Board (Finland), website, <https://www.ylioppilastutkinto.fi/en/>.

<sup>7</sup>U.S. Department of Education, *Digest of Education Statistics, 2017*, advanced release ed. (Washington, DC: National Center for Education Statistics, 2018).

<sup>8</sup>John H. Bishop et al., "The Role of End-of-Course Exams and Minimal Competency Exams in Standards-Based Reforms," in Diane Ravitch, ed., *Brookings Papers on Education Policy 2001* (Washington, DC: Brookings, 2001); Guido Schwerdt, Martin R. West, and Marcus A. Winters, "The Effects of Test-Based Retention on Student Outcomes over Time: Regression Discontinuity Evidence from Florida," *Journal of Public Economics* 152 (2017/08/01/): 154–69.

<sup>9</sup>Michael Hout and Stuart W. Elliott, eds., *Incentives and Test-Based Accountability in Education* (Washington, DC: National Academies Press, 2011).

<sup>10</sup>Eric A. Hanushek, "Grinding the Antitesting Ax: More Bias Than Evidence behind NRC Panel's Conclusions," *Education Next* 12, no. 2 (Spring 2012): 49–55.

<sup>11</sup>National Center for Education Statistics. 2011. *America's High School Graduates: Results of the 2009 NAEP high school transcript study*. Washington, DC: U.S. Department of Education.

<sup>12</sup>Eric A. Hanushek, "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data," *American Economic Review* 60, no. 2 (May 1971): 280–88.

<sup>13</sup>U.S. Department of Education, "ESEA Flexibility," June 7, 2012, <https://www2.ed.gov/policy/eseaflex/approved-requests/flexrequest.doc>.

<sup>14</sup>Eric A. Hanushek and Steven G. Rivkin, "The Distribution of Teacher Quality and Implications for Policy," *Annual Review of Economics* 4 (2012): 131–57; Eric A. Hanushek, "The Economic Value of Higher Teacher Quality," *Economics of Education Review* 30, no. 3 (June 2011): 466–79; Raj Chetty, John N. Friedman, and Jonah Rockoff, "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review* 104, no. 9 (September 2014): 2633–79; Eric A. Hanushek and Ludger Woessmann, *The Knowledge Capital of Nations: Education and the Economics of Growth* (Cambridge, MA: MIT Press, 2015).

<sup>15</sup>Elizabeth Ross and Kate Walsh, *State of the States 2019: Teacher and Principal Evaluation Policy* (Washington, DC: National Council on Teacher Quality, 2019).

<sup>16</sup>Thomas S. Dee and James Wyckoff, "Incentives, Selection, and Teacher Performance: Evidence from IMPACT," *Journal of Policy Analysis and Management* 34, no. 2 (Spring 2015): 267–97; Thomas S. Dee and James Wyckoff, "A Lasting Impact: High-Stakes Teacher Evaluations Drive Student Success in Washington, D.C.," *Education Next* 17, no. 4 (Fall 2017): 58–66.

<sup>17</sup>Eric A. Hanushek, "The Unavoidable: Tomorrow's Teacher Compensation," Hoover Education Success Initiative (Stanford, CA: Hoover Institution, January 2020).

<sup>18</sup>Of course, this environment also calls for many changes in staffing, budgets, curricula, technology, and the delivery of instruction.

**Testing cannot furnish all the data that educators and policymakers need, but it is an essential source of indispensable information.**

**Chester E. Finn Jr. is Distinguished Senior Fellow at the Thomas B. Fordham Institute and a senior fellow at the Hoover Institution of Stanford University. Eric A. Hanushek is the Paul and Jean Senior Fellow at the Hoover Institution.**



*Time to steer systems  
toward better balance  
and coherence.*

**Chris Domaleski**

## Breakthrough or Breakdown? School Accountability in Flux

Newton's second law of motion states that when force is applied to an object at rest, the rate of change is directly proportional to the force applied. There is little doubt that tremendous force has been applied to education recently, including assessment and accountability systems. First, COVID-19 prompted rapid shifts to distributed learning, suspension of state assessments, and waivers for school accountability. Then the collective anguish over George Floyd's murder in Minneapolis led to a sense of urgency for advancing equity in all institutions, including education.

Already under fire before spring 2020, the dominance of federally mandated test-based accountability in American education had led many educators and policymakers to decry the system as largely out of balance and to suggest that this imbalance has stifled productive local efforts toward meaningful, lasting improvements in student learning. It is not too early to conclude that all these cumulative factors will—and should—change assessment and accountability systems. But what kind of change is appropriate, and how can state boards of education support such changes?

## What Is the Promise of School Accountability?

Before state boards can help nudge their systems back into balance, they should examine the purpose of accountability systems in the first place. Broadly, school accountability can be thought of as a system that 1) signals what outcomes are valued, 2) provides information about school performance with respect to those outcomes, and 3) prescribes a system of supports and interventions based on performance. Improvement is thought to occur by incentivizing the right kinds of behaviors and actions, shining a light on areas where improvement is needed, and providing targeted supports to those areas. That may sound straightforward enough, but this portrayal is built on scores of assumptions and a vast network of actions and interactions that are best addressed in a strong theory of action.

Accountability systems may highlight goals and benchmarks and provide some useful information to guide actions, but real educational progress always has been pegged to the practice of teaching and learning that occurs daily in classrooms. In order to promote school improvement, systems must activate the conditions and supports that provide students an opportunity to learn. School improvement requires attention to the “instructional core,” with these principles in mind:

There are only three ways to improve student learning at scale: You can raise the level of the content that students are taught. You can increase the skill and knowledge that teachers bring to the teaching of that content. And you can increase the level of students’ active learning of the content. That’s it. Everything else is incidental.<sup>1</sup>

Ultimately, I argue that accountability systems can play a role in an overall plan to promote student success, but they are not a “treatment” and far from a holistic prescription for education reform. In fact, too often contemporary accountability systems are built on an impoverished theory of action—that suggests putting data in the hands of policymakers or educators will lead to strategic actions to improve schools.

What then are the key factors essential for leveraging the promise of accountability to improve schools? I will describe three:

principled design, balanced and coherent structures, and reciprocal support.

## Principled Design

There is no gold standard, no single correct approach, for developing and implementing systems that monitor and support school improvement. A system should take into account policy priorities, local conditions, and context. Only then can education leaders design a solution that supports guiding principles and values.

Engaging in a principled design process is best accomplished through a well-explicated theory of action.<sup>2</sup> A theory of action is a tool for both designing and evaluating accountability systems by clarifying goals, assumptions, and the hypothesized mechanisms to bring about the intended changes. Randy Bennett’s recommendations for an assessment system theory of action can be extended to accountability. Bennett’s conceptualization includes

- the components of the system and the rationale, grounded in research and theory, for these components;
- the claims that will be made from results or outcomes;
- the intended effects of the system;
- the mechanisms thought to cause the intended effects; and
- potential unintended or negative effects and the plan to mitigate them.<sup>3</sup>

Developing a strong theory of action starts with clarifying the highest priority goals. For example, the system can be designed to privilege closing achievement gaps or preparing students for postsecondary success. While a system can have multiple goals, focus is important. If an initiative is thought to do everything well, there is a good chance it will not do anything very well.

The value of a theory of action comes from documenting the hypothesized connections between and among the actions and the outcomes they are thought to promote. At a high level, Erika Hall and colleagues recommend portraying these elements with respect to inputs, outcomes, and evidence.<sup>4</sup>

Inputs describe the resources, actions, and conditions that will be necessary to support improvement. It includes the source—federal,

---

**Developing a strong theory of action starts with clarifying the highest priority goals.**



---

**Developing the theory of action is a shared responsibility, which should be led by policymakers and leaders.**

state, district—and nature of the support. For example, inputs might include financial resources and wraparound student support services.

With respect to outcomes, the theory of action should specify the proximal and distal effects the system will promote. For example, instructional practices will improve, family and community engagement will grow, or students will encounter and meet higher academic expectations. Importantly, the supporting conditions and rationale thought to promote the outcome should be made explicit.

Finally, the theory of action specifies the evidence to support connections among inputs and outcomes. For example, improved school climate is documented by survey data, instructional practices are appraised via interviews and observations, and tracking disaggregated growth rates provides insight on academic progress for students in traditionally underserved groups. A systematic collection of evidence can help leaders engage in ongoing monitoring, review, and refinement of the system.

Only by investing upfront in the hard work of creating a well-specified theory of action can the role of assessment and accountability in supporting improved outcomes be understood and evaluated. Developing the theory of action is a shared responsibility, which should be led by policymakers and leaders, including chief state school officers and state boards of education, in collaboration with practitioners, technical advisors, and a broad-based group of stakeholders.

## **Balance and Coherence**

Currently, nearly all the attention on school improvement is connected to state accountability systems that were designed to meet federal requirements as specified in the Every Student Succeeds Act (ESSA). These systems are overwhelmingly influenced by academic indicators in the form of performance on end-of-year summative tests in English language arts and mathematics. This represents a woefully imbalanced, incoherent system.

Creating more balance and coherence involves attending to the emphasis and interrelationships within and among systems at the federal, state, and local levels. To start, it is useful to briefly distinguish the roles at each of level.

**The Federal, State, and Local Role.** Equity is at the core of the Elementary and Secondary Education Act (ESEA) of 1965, an initiative

designed to improve educational opportunities for disadvantaged students. Early federal accountability provisions typically focused on compliance or inputs.<sup>5</sup> However, the scope has grown over the over the years, with perhaps the most pronounced pivot occurring with No Child Left Behind (NCLB), which mandated annual statewide achievement testing and school accountability systems that predominately linked consequences to attainment of proficiency on these tests.<sup>6</sup> Accountability has not been the same since.

State educational authority is drawn from each state's constitution and requirements from the legislature, the state board, and other governing bodies. Chief among the states' roles are to establish the content and rigor of the academic standards, implement the state assessment and accountability systems, and provide support and resources to help districts and schools meet performance expectations.<sup>7</sup> Naturally, federal requirements heavily influence much of a state's efforts in this endeavor, as states must meet the detailed strictures of ESSA and federal peer review before implementing state assessment and accountability systems.

Despite the prominent influence of federal and state authority, education remains primarily a local responsibility. Local school boards, responding to community priorities, and district leaders are charged with the essential front-line responsibilities: hiring and supporting educators and staff, establishing and implementing the curriculum, and managing day-to-day operations such as transportation, facilities, and food service.

**Balance.** Balance primarily refers to the development of systems that are well specified at each level and pegged to the appropriate areas of emphasis. Unfortunately, contemporary accountability systems at the district level, if they exist at all, often treat districts like a “super school” by simply aggregating all the school metrics to the district. Moreover, these school metrics typically mirror the state's standardized model. As previously noted, this model primarily comprises performance on end-of-year tests.

Districts can achieve more balance in at least two ways. First, they can design their systems to reflect the specific goals and priorities of the district and community, especially with respect to the areas under the districts' direct influence. Second, the systems can be specified at a much finer grain size that takes advantage

### Box 1. Promising Initiatives

While much work remains to improve school accountability, there are some promising initiatives. One is the Massachusetts Consortium for Innovative Education Assessment (MCIEA), a collection of school districts and partners that are reimagining assessment and accountability. Their framework explicitly links essential inputs (e.g., educator and leader development, school culture, resources to support teaching and learning) and key outcomes such as academic learning, community building, and well-being. The model puts classroom performance assessment in the foreground in lieu of high-stakes state summative testing and includes a variety of broad school quality measures that are informed by research and community input and linked to prioritized outcomes. By so doing, the model moves toward more balanced, coherent accountability.

Innovation at the state level can be more difficult, not least because of federal constraints. Some states have overcome this by pursuing initiatives outside of ESSA. Recognizing that a one-size-fits-all system does not work well for alternative high schools, the Wyoming Department of Education engaged in a multiyear initiative to produce a supplemental school accountability system.<sup>a</sup> The resulting framework was built on a distinct theory of action that recognizes the unique mission of alternative schools and incorporates a broad, flexible set of indicators selected to support prioritized outcomes. For example, the system promotes student engagement and the holistic development of skills associated with postsecondary success via mechanisms such as individualized student success plans. While this system addresses a relatively small number of schools, the process and resulting framework are instructive.

Numerous states have signaled a move toward more improved school accountability. For example, New Hampshire's Performance Assessment of Competency Education (PACE) could serve as a mechanism to more tightly link curriculum, instruction, and assessment (see article, page 39). Georgia's College and Career Ready Performance Index encourages schools to promote readiness in areas such as fine arts and languages and rewards attainment of a variety of college- and career-ready credentials. This flexibility supports balance, allows for appropriate differentiation, and ultimately provides more useful, actionable outcomes. The Louisiana Department of Education helps districts implement strong practices by 1) reviewing extant curriculum and assessment resources, results of which are publicly available, 2) providing training and instructional resources, and 3) curating a bank of model resources and instructional tools.

<sup>a</sup>Chris Domaleski and Erika Hall, "Wyoming Alternative School Accountability Framework: Recommendations from the Alternative Accountability Advisory Committee" (Dover, NH: National Center for the Improvement of Educational Assessment, 2015).

of local initiatives and inputs. For example, because the district has auspices over personnel and professional development, they are best positioned to monitor and evaluate the efficacy of programs such as new teacher induction, professional development initiatives, and mentoring programs. As another example, schools and districts may have access to data on student performance apart from statewide tests. More focused information provided in a more timely manner is more likely to influence changes in practice.

States play a part in promoting balance as well. Although federal requirements constrain state systems, states can create supplemental initiatives that need not be high stakes, insofar as they are focused on producing performance classifications. For example, states can curate model resources and research-based improvement practices or help collect and report meaningful data that goes beyond summative assessments. As another example, some states have developed differentiated accountability and support systems for alternative schools outside ESSA. Taken together, there are multiple ways for states to partner with districts to promote balance (see box 1).

**Coherence.** As Ben Forman, Charles DePascale, and I detail elsewhere, coherence in accountability addresses at least three dimensions:

- **External coherence:** Are connections among multiple accountability systems logically consistent?
- **Internal coherence:** Are comparisons within the local system logically related?
- **K-12 coherence:** As students advance from early grades to graduation, are the different levels of the system logically connected?<sup>8</sup>

---

**More focused information provided in a more timely manner is more likely to influence changes in practice.**

---

**State systems can be poorly specified, such that they thwart innovation at the local level.**

External coherence is evident when systems at the state, federal, and local levels are mutually supportive. The responsibility goes both ways. Unfortunately, state systems can be poorly specified, such that they thwart innovation at the local level. For example, a state system that places too much emphasis on participation or performance on Advanced Placement courses as a pathway to postsecondary success may create a perverse incentive for districts and schools to eschew a broader range of academic and cocurricular experiences that prepare students for success in college and careers.

In systems with internal coherence, the components are mutually supportive and aligned to the overarching system goals. Incoherence can manifest in several ways. It may occur when systems of support are not connected to primary outcomes of interest. For example, if a central objective of the system is to close achievement gaps but there is no plan to provide supports to students in historically lower performing groups, the system is not internally coherent. There are technical features that can contribute to incoherence, as well. A common example is the disproportionate focus on proficiency rates in most systems, which is ill suited to gauge progress for students below standard, in contrast to academic growth measures.

Finally, K-12 coherence, which could also be termed vertical coherence, refers to a system that is thoughtfully designed to support students' success throughout their educational pathways. For example, Scott Marion and colleagues emphasize the importance of learning progressions as the "organizing framework for connecting various assessments and learning activities in a vertically coherent system."<sup>9</sup> Unfortunately, it is all too common for schools to administer large-scale commercial assessments at regular intervals to gauge student progress with little or no information about the degree to which they are providing useful feedback on the skills most important for students to demonstrate success as they progress through the curriculum.

## **Reciprocal Support**

The idea of reciprocity as a key factor in developing effective accountability systems is not new.

Reciprocity refers to the shared responsibility to support attainment of performance expectations. As Richard Elmore explains, "For every increment of performance I demand from you,

I have an equal responsibility to provide you with the capacity to meet that expectation."<sup>10</sup>

Arguably, contemporary accountability systems have overstated classifications and underspecified the mechanisms for building capacity and providing support to educators and students. In fact, it might be more appropriate to shift the language from "school accountability systems" to "school support systems" to reflect the proper emphasis on support as the key element in any well-designed system.

Given the central importance of reciprocity in promoting improved systems, it is useful to further develop two essential elements to reciprocal support: differentiation and utility.

**Differentiation.** One might challenge the notion that schools are the sole locus of accountability. Brian Gong explores this concept using the term "differentiated" accountability:

Advocates of differentiated accountability may argue that just as it isn't reasonable to hold students accountable for meeting standards until they have been given a fair opportunity to learn, so it is not reasonable to hold schools accountable until schools have been given a fair opportunity to provide the opportunity to learn.<sup>11</sup>

He suggests that an accountability system can be considered incomplete if it lacks details about the obligations and consequences associated with all entities responsible for establishing the conditions for success. This information should be included in the guiding theory of action. It stands to reason that developing such a system will also promote more balanced, coherent structures.

**Utility.** Utility refers to the extent to which information generated in the system is appropriate and actionable. As noted previously, this requires having indicators represent the outcomes of interest and are appropriately specified to detect the desired effects.

Moreover, utility is supported when the timing and manner of reporting are sufficient to inform a helpful response. With little exception, current accountability practice provides information on distal outcomes at the end of the academic year or later. Improved systems will be characterized by data collection and reporting systems that provide more signals along the way to indicate if the desired outcomes are on track. By providing information at regular



intervals throughout the year (e.g., attendance, performance on interim assessments), personnel will be better equipped to detect and prevent a negative outcome.

It should be obvious that such data and support systems must be decoupled from high-stakes classifications if they are to be useful. To the extent that data intended for ongoing monitoring also inform summative classifications, the value of that information will almost certainly diminish due to Campbell's Law. Named for social psychologist Donald Campbell, this principle holds that the likelihood of an indicator being corrupted increases in proportion to the degree to which that indicator is used for consequential decision making.<sup>12</sup>

## Final Thoughts

It is not easy to change the status quo in American public education. But perhaps the moment is right to consider a new path forward for school accountability that is more credibly linked to improved outcomes for all students.

What are some specific actions that state boards can take to better leverage the promise of accountability?

- Work to develop a comprehensive theory of action for school improvement informed by a broad-based and diverse group of leaders and stakeholders. Regard this theory of action as a dynamic document, returning to it often to refine assumptions and guide monitoring and support.
- Use the theory of action as the foundation for reexamining the coherence and balance of the state's accountability system. Refine the system as needed to ensure that it helps the state measure and promote what matters most. Consider including a broad set of indicators that go beyond summative assessments and measures of performance that can provide useful signals of student progress during the academic year. Such indicators may be supplemental—that is, decoupled from the federal system.
- Advocate for state practices that support districts and schools to stand up strong and complementary practices for monitoring and support. These systems should reflect local values, goals, and responsibilities.

- Promote the curation and dissemination of research and resources to help districts and schools. For example, provide model assessment and instructional resources or evaluate the quality of commercial products with respect to the state's academic expectations.

To be fair, it will take more than “a better system” or sheer force of will to realize long overdue reform that promotes better outcomes. Wraparound services and supports to address the needs of historically underserved students must be vigorously pursued. State and federal laws must be friendlier to innovation. And sustained efforts to build capacity to improve teaching and learning are critical. Now is the time to shift focus away from counterproductive practices and toward more promising alternatives. ■

<sup>1</sup>Elizabeth A. City et al., *Instructional Rounds in Education: A Network Approach to Improving Teaching and Learning* (Cambridge, MA: Harvard Educational Press, 2003).

<sup>2</sup>Marianne Perie, “Key Elements for Educational Accountability Models” (Washington, DC: Council of Chief State School Officers, 2007); Erika Hall et al., “A Framework to Support Accountability Evaluation” (Washington, DC: Council of Chief State School Officers, 2016).

<sup>3</sup>Randy Elliot Bennett, “Cognitively Based Assessment of, for, and as Learning: A Preliminary Theory of Action for Summative and Formative Assessment,” *Measurement: Interdisciplinary Research and Perspectives* 8 (2010): 70–91.

<sup>4</sup>Hall et al., “Framework to Support Accountability Evaluation.”

<sup>5</sup>Susan H. Fuhrman and Richard F. Elmore, eds., *Redesigning Accountability Systems for Education* (New York: Teachers College Press, 2004).

<sup>6</sup>Margaret E. Goertz, “Implementing the No Child Left Behind Act: Challenges for the States,” *Peabody Journal of Education* 80 (2005): 73–89.

<sup>7</sup>Chris Domaleski, Damian Betebenner, and Susan Lyons, “Promoting More Coherent and Balanced Accountability Systems” (Dover, NH: National Center for the Improvement of Educational Assessment, 2018); Joanne Weiss and Patrick McGuinn, “The Evolving Role of the State Education Agency in the Era of ESSA and Trump: Past, Present, and Uncertain Future,” *CPRE Working Papers* (2017).

<sup>8</sup>Ben Foreman, Charles DePascale, and Chris Domaleski, “Local Accountability: The Forgotten Element in Education Reform” (Boston: Massachusetts Institute for a New Commonwealth, 2018).

<sup>9</sup>Scott Marion et al., “A Tricky Balance: The Challenges and Opportunities of Balanced Systems of Assessment” (Dover, NH: National Center for the Improvement of Educational Assessment, 2018), p. 3.

<sup>10</sup>Richard Elmore, *Bridging the Gap between Standards and Achievement: The Imperative for Professional Development in Education* (Washington, DC: Albert Shanker Institute, 2002), p. 5.

<sup>11</sup>Brian Gong, “The Next Generation of State Assessment and Accountability” blog post (Dover, NH: National Center for the Improvement of Educational Assessment, April 2, 2020).

<sup>12</sup>Donald T. Campbell, “Assessing the Impact of Planned Social Change,” *Evaluation and Program Planning* 2, no. 1 (1979): 67–90.

**To the extent that data intended for ongoing monitoring also inform summative classifications, the value of that information will almost certainly diminish.**

**Chris Domaleski is associate director at the National Center for the Improvement of Educational Assessment.**

*State policymakers should take the opportunity to reimagine their education systems.*

**Abby Javurek and Jason Mendenhall**

# How a Crisis Can Transform Learning, Teaching, and Assessment

When states issued the first coronavirus stay-at-home order last March, few could have foreseen how the foundations of learning and teaching would be shaken. As hopes for a brief disruption confronted reality, state and district leaders began plans to extend some form of distance learning into the 2020–21 school year. They also began to think about how to turn a daunting crisis into an opportunity to rethink policies and practices from the ground up—accountability policies in particular.

For many years, the majority of U.S. students have not met grade-level expectations, with socioeconomically disadvantaged students disproportionately affected.<sup>1</sup> The pandemic only underlined and exacerbated these disparities. The coronavirus hit historically underserved communities harder, and the digital divide has harmed students in these communities more.

These facts loom as state policymakers consider the environment that students and teachers face this school year. Despite the heroic efforts of teachers, students, and families, some students may have returned to school with as little as 70 percent of the learning gains they typically would have made in reading (figure 1) and less than 50 percent of the gains they would have made in math (figure 2), according to research from assessment nonprofit NWEA. Some may be as much as a year behind in math.<sup>2</sup>

Researchers further project that existing achievement gaps will likely widen as a result of the school closures, with some students benefiting from remote and self-driven learning environments and others—more likely those already behind, lacking learning opportunity at home, or both—losing ground.<sup>3</sup>

To put this potential variance in context, researchers from several

universities conducted an analysis of NWEA data across 10 states to understand the range of instructional levels present in a “typical” fifth-grade classroom after a regular summer break. They found students with instructional needs ranging from third grade through high school (figure 3).<sup>4</sup> Given that gaps are expected to be even larger this school year, this figure represents a best-case scenario.

What can state boards of education do in the face of this challenge? We believe they and their peers in state policymaking have an unprecedented chance to not only recover learning loss due to COVID-19 but also address the longstanding problem of learning opportunity by effecting change at the systems level. They will need to ask some hard questions: How can educators best serve the diverse learning needs of any given cohort of students? Should all learners be “ready” at the same time? How do we know the system is helping our kids grow and prepare for the future?

## Teaching and Learning

Other major disruptions, such as Hurricane Katrina in 2005, have demonstrated that helping students get back on track is a long-term challenge. Even under “normal” circumstances, it is uncommon for students to achieve a year and a half of growth in a single year.<sup>5</sup> For some, it may be tempting in the face of this challenge to do away with accountability for challenging all students to reach their highest potential. But a better response is to systematically pursue new approaches to learning, teaching, and assessment that better withstand disruption and more efficiently accelerate student achievement.

Reimagining learning should start with new structures and systems. Agrarian school calendars inherently

Figure 1. Projected COVID Slide in Reading

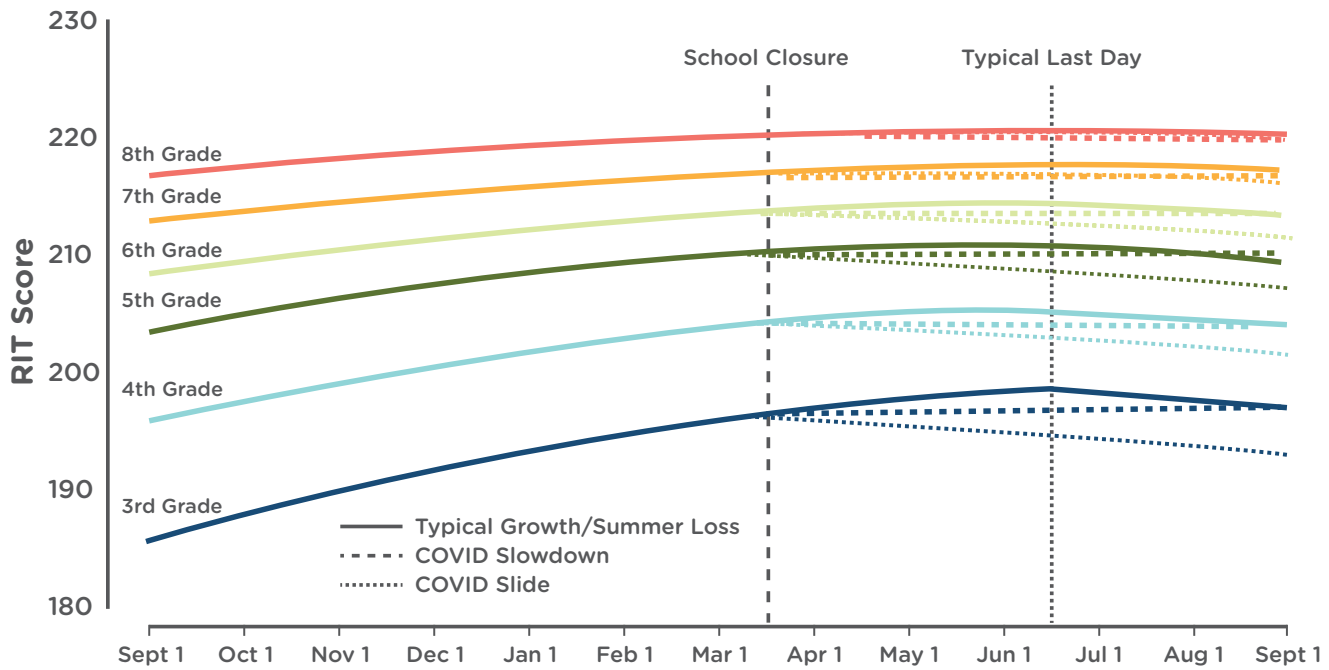
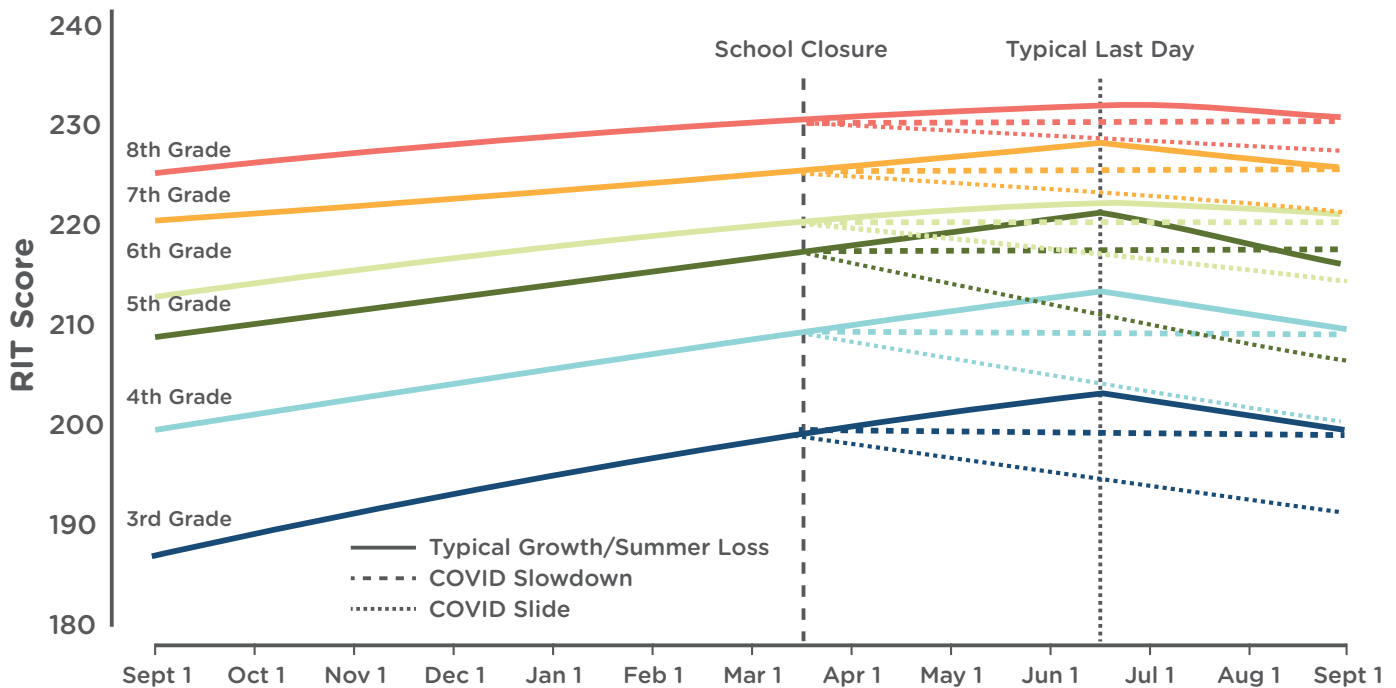
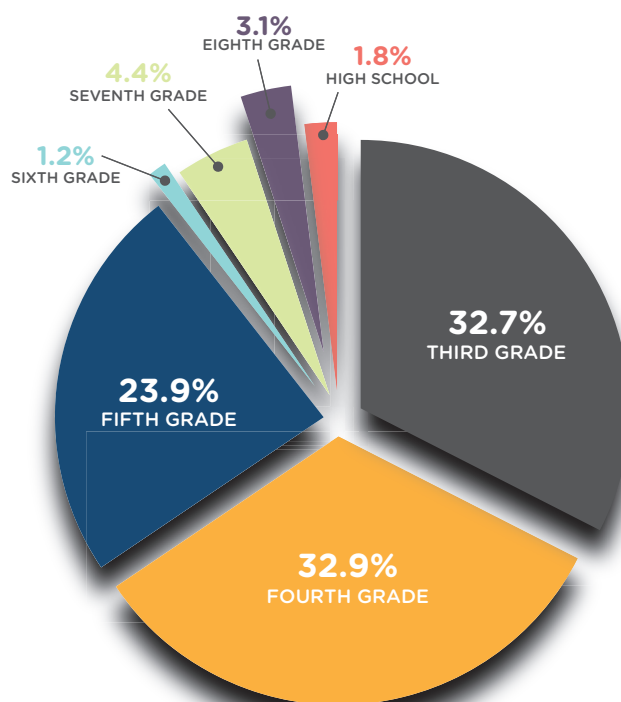


Figure 2. Projected COVID Slide in Math





**Figure 3. Instructional Needs in Average Grade 5 Class Span Seven Grades**



**State leaders must assume that blended learning is a core instructional strategy, not just an emergency response.**

disrupt learning continuity and yield learning loss.<sup>6</sup> Rhythmic, year-round, school calendars are politically difficult to implement but are a precious opportunity to mitigate lost learning by increasing learning opportunity. Evolving in this direction will not happen overnight, so we recommend that state and district leaders start thinking through what a potential transition might look like and what steps can be taken immediately to serve learners during summer months. It will be important to move beyond sending summer reading lists to more formal blended and online summer learning strategies—not only in summer 2021 but also in the summers that follow.

State leaders must invest in distance learning infrastructure and assume that blended learning is a core instructional strategy, not just an emergency response. Before the pandemic, authors of a Bellwether Education Partners report said that teachers think digital learning has a positive impact on student learning but do not always leverage it in ways that motivate and engage students. They also found that teachers value assessment data but have limited time to digest it, have too much of it, and do not always receive it in a timely fashion. They concluded that the education sector must “reimagine how

targeted intervention for students with learning gaps is delivered so as to maximize access to grade-level knowledge and skills.”<sup>7</sup>

Coronavirus school closures intensified the urgency of this reimagining. Ready or not, state and district leaders are grappling with the need to leverage digital learning tools and assessments to address the unique needs of every student. They must continue to invest in those tools and the professional learning that teachers will need to support challenging, personalized learning for every student—whether in person or remotely.

Technology cannot replace a caring, high-quality teacher, but in combination with adult and peer interactions, digital tools and data can expand teachers’ reach and increase their impact. Addressing the complexity of administering assessments remotely while protecting the validity of results and the security of test questions is central to this effort. In this new world of more flexible, resilient learning and teaching, educators still need timely, accurate information to make good decisions for each learner. Thus it is necessary, though challenging, to provide quality remote proctoring capabilities that ensure data integrity.

## Age, Grade, and Opportunity

Making good decisions for individual learners means understanding what they know and are ready to learn next—whether on or off grade level—and acknowledging that opportunities to learn are not limited to schooling. In a global digital world, learning can happen on demand, inside and outside the formal classroom. It will be important for education leaders to foster connections with supplemental and community-based learning opportunities and to support teachers with assessments that can capture the full range of learning strengths and gaps for each learner.

Grade-level standards are essential for establishing expectations for what each learner should know and be able to do at particular points in time. However, they are not a prescription for meeting a learner's needs, especially high achievers and struggling learners. And they cannot address the broad range of the pandemic's impacts on student achievement. Given that economic hardship often drives student mobility,<sup>8</sup> COVID-19 could present teachers with even more learners than usual for whom they have little context. In this environment, a student's age and grade level will be less reliable as a proxy for learning needs than ever before. Understanding the unique needs of each learner, regardless of where they fall on the continuum, will be paramount.

As personalized learning experts suggest, focusing on grade-level content in spite of lost or unfinished learning makes it harder for students to master more advanced concepts and hinders their college and career readiness.<sup>9</sup> Yet too often, well-meaning efforts to meet students where they are result in students not being appropriately challenged, which can lead to a systematic lowering of expectations for students who are behind.<sup>10</sup> Maintaining high expectations for all students is essential, but expectations alone do not outline the interventions and scaffolding required to optimize learning for each student, especially given the greater-than-usual variance in skill that teachers are expected to see this school year.

Meeting student needs and maintaining high expectations is not an either/or dilemma but rather an opportunity to implement more efficient, scalable ways to personalize and accelerate learning.<sup>11</sup> The technology and knowledge

required to achieve this is available now. Consider these ideas:

- **Make distinctions to maximize learning based on developmental needs and/or content areas.** Earlier grades may require more time (i.e., more than one school year) to master the basic skills on which everything else builds. Students in later grades may need more individualized, scaffolded support to fill critical skills gaps. Educators have an opportunity to make better, more granular, student-centered decisions to optimize learning for each student, in each subject.
- **Rethink grouping and regrouping students based on learning needs.** Especially in K-5/6 environments, cohorts of learners staying with the same teacher for all learning needs and instruction in a year make differentiating at the student level challenging, but grouping and regrouping cohorts of learners is scalable and achievable. Accountability and teacher effectiveness models will need to evolve to support this team-based approach.
- **Facilitate the implementation of high-quality curriculum.** One reason that students may not be challenged at the appropriate level is a disconnect between learning expectations and curriculum. State leaders can help establish criteria for selection of high-quality, coherent curriculum that is not only standards aligned but also engaging in a digital environment—an essential criterion in a blended learning world.
- **Align assessments to high-quality curriculum.** Alignment between assessments and curriculum can help teachers place students in a curricular progression (versus using only grade level or age as a proxy for identifying learning needs) and more effectively close learning gaps. In states where districts are using multiple high-quality curricula, selecting assessments that can be configured to align with those curricula is ideal.
- **Support the adoption of adaptive assessments administered throughout the school year.** Administering adaptive, standards-aligned, interim assessments that identify student learning needs (regardless of age/grade level) and performance relative to learning expectations will be critical. Now more than ever, educators will need rhythmic,

---

**COVID-19 could present teachers with even more learners than usual for whom they have little context.**

■  
The goal of through-year assessment is to increase coherence by using one assessment system to yield different types of data.

shorter-cycle evidence of student learning to make good decisions for each learner.

■ **Expand the value of statewide assessment.**

States can support systems that provide timely learner-specific feedback as well as coherence relative to summative expectations by providing standards-aligned formative and interim assessments for districts. They can also rethink statewide summative tests so that students can demonstrate proficiency when they are ready (competency-based and through-year assessment models, for example) as opposed to current systems, which expect all students to proceed on the same timeline.

## Innovative State Assessment Systems

The Every Student Succeeds Act (ESSA) permits innovation in summative assessment either statewide or within a subset of a state's local educational agencies (LEAs) or schools through the Innovative Assessment Demonstration Authority (IADA) program. Nebraska is moving toward statewide innovation, and Georgia, Louisiana, Massachusetts, New Hampshire, and North Carolina are exploring innovative approaches through the IADA pilot program (see also article, page 40).

This movement to change assessment began before the pandemic, with many states looking for ways to reduce overall testing and better support districts in their efforts to foster student learning. However, the cancellation of spring 2020 statewide summative testing across the nation due to school closures opens the door even wider for states to reconsider the best way to gauge how well schools are serving students.<sup>12</sup> This rethinking could usefully go beyond measuring whether learning happened at year's end to providing educators with timely, actionable data for improving learning throughout the year.

NWEA has partnered with Nebraska, Georgia, and Louisiana on their innovative assessment efforts, which focus on these through-year models. The models are designed to streamline and connect assessments that districts use to inform instruction with those the state designed to meet accountability requirements. Thus a single assessment system produces information for districts during the school year and summative scores, which along

with other measures are critical to provide a holistic view of school performance for school improvement efforts.

Most districts administer interim assessments during the school year to inform instruction and measure academic growth plus a summative test to meet accountability requirements in the spring. Through-year assessment eliminates the need to double-test in the spring and produces timely results during the school year showing where students are in their learning as well as how they are performing relative to summative grade-level expectations. This information supports teachers in addressing individual student learning needs while keeping proficiency and college and career readiness goals in view so all students are challenged to reach their full potential.

In summary, there are several benefits to through-year models:

- **Coherence.** A unified, cohesive approach avoids the use of one compass in the classroom and another in accountability frameworks.
- **Efficiency.** A single assessment system eliminates the need to administer both interim and summative assessments in the spring.
- **Utility.** Information on student performance throughout the year helps teachers target instruction and take action to ensure student mastery of foundational skills and concepts.
- **Fairness.** A fall-to-spring growth measure provides a more accurate view of schools' impact on learning than models rooted in summative performance alone.<sup>13</sup>

Even so, through-year assessment models are not without potential challenges. Some critics express concerns that using a single assessment solution to serve more than one purpose could compromise the value of the information produced. However, the goal of through-year assessment is not to produce one type of data to serve multiple purposes. Rather, it is to increase coherence by using one assessment system to yield different types of data. Advances in technology and test design support the development of models that address clarity in purpose and validity of score calculations.

In contrast, separating formative, interim, and summative assessments at the expense



of coherence in design can send a confusing message to educators about which actions to take. A disconnect between the assessments used to inform instruction and the ones used to ensure accountability for instructional outcomes can result in competing priorities. The coherence created with through-year assessment avoids this by supporting a match across all tests in the system with theories of learning and teacher action.

That said, innovation takes time, patience, and resources. The six states and participating districts currently pursuing innovative approaches are making sacrifices to effect change; it can take several years to make innovative assessments operational. If state leaders do not feel ready to take the leap, they may wish instead to innovate by creating more efficient and coherent summative tests and/or investing in balanced systems of assessment—for example, providing high-quality, standards-aligned formative and/or interim assessments to districts to ensure they have instructionally useful information during the school year.

## Aligning Accountability with Teaching and Learning

Even as states and districts improve education systems to accelerate learning, they must consider how much growth is realistic in a year. Particularly in the wake of the pandemic, where students may lose as much as a year of learning, state leaders should evaluate the fairness of using proficiency standards set under different learning conditions for accountability purposes. That does not mean proficiency expectations should be lowered, but it is worth reckoning how long it will take to get many learners back on pace to be college, career, and life ready.

This crisis gives state boards a chance to reassess the best way to hold schools accountable and identify those most in need of support—not just in 2021 but in the years that follow (box 1). As more states move toward interim assessments as part of balanced assessment systems or toward through-year models of assessment, they should leverage those assessments to measure fall-to-spring and cross-year growth, regardless of summative performance, for two reasons:

- **Avoids attribution of seasonal learning loss (or gain) to schools.** School effectiveness

### Box 1. How State Boards Can Catalyze Innovation

When reimagining systems of learning, teaching, and assessment, purpose is a driving factor. With this in mind, state boards should ask the following:

- What are the biggest challenges, learning gaps, and pain points our state and district leaders, teachers, students, and families are facing?
- What do we most want our schools to provide for our students?
- What is working well in supporting student learning in our state, and what evidence do we have of this?
- Are the board, the state education agency, and districts invested in and in sync on effecting change, and are roles clearly defined?
- How can technology help teachers develop caring, personal relationships with students and meet individual needs in a scalable way?
- What questions do we wish to answer by using assessments, and what decisions will we make with the resulting data?
- What are the impacts of new models—and new data—on teacher evaluation and accountability?
- What trade-offs are we willing to make? Are we willing to reimagine our system and how?

Discussion generated by these questions can reveal who state boards need to convene, as well as the policy shifts required to design systems that better foster student learning.

measures are sensitive to summer loss, and this will continue to be a concern as long as agrarian school calendars are the norm. An NWEA research study found that school accountability data based on students' fall-to-spring growth were often different from data based on spring-to-spring growth, a common practice under ESSA.<sup>14</sup>

- **Avoids using short-term student growth to measure school efficacy in achieving**

**Reimagining accountability does not mean that states should stop looking at student performance on summative tests or growth in proficiency.**

**long-term goals (college and career readiness).** Schools held accountable for short-term student growth are often very different from those that would be identified for improvement under ESSA if more years of data were used.<sup>15</sup>

Reimagining accountability does not mean that states should stop looking at student performance on summative tests or growth in proficiency over time. These data are crucial to identifying and addressing educational inequities. Rather, states should assess this information alongside additional measures such as growth over time, regardless of proficiency level, and reconsider whether an annual review period is appropriate.

## The Road Ahead

It has been heartening to see state leaders, district administrators, teachers, students, and families rising to the challenges created by the pandemic and working hard together to keep students learning. Yet unless state leaders seize this chance to effect change, the crisis risks making longstanding opportunity-to-learn problems even worse.

Everyone working in education has had to foster learning in new ways and make related decisions about systems of learning, assessment, and accountability. Approaching these changes as a stopgap to get through a period of uncertainty would be a missed opportunity to make education more equitable, resilient, and effective.

State boards of education and other education leaders can drive discussion and action around models of blended learning that employ student-centered practices and embrace the global digital world in which we live. Specifically, they can support systems that

- leverage digital tools, increase learning opportunities, extend beyond age- and grade-based structures, and maintain common, high expectations for all students;
- adopt high-quality curriculum, professional learning, and coherent assessment models that not only measure learning but also support teachers in fostering learning throughout the year; and
- create fair, meaningful accountability models that more effectively reveal which schools—across all communities—are thriving and which need additional support.

By reimagining education in the wake of an extraordinary experience shared across the world, state boards can better ensure that every student can develop the skills and knowledge they need to succeed in college, careers, and life. ■

<sup>1</sup>National Assessment of Educational Progress, The Nation's Report Card, "How Did U.S. Students Perform on the Most Recent Assessments?" web page, <https://www.nationsreportcard.gov/>.

<sup>2</sup>Megan Kuhfeld and Beth Tarasawa, "The COVID-19 Slide: What Summer Learning Loss Can Tell Us about the Potential Impact of School Closures on Student Academic Achievement" (Portland, OR: NWEA, April 2020).

<sup>3</sup>Megan Kuhfeld et al., "Projecting the Potential Impacts of COVID-19 School Closures on Academic Achievement," EdWorkingPaper (Providence, RI: Annenberg Institute at Brown University, May 2020).

<sup>4</sup>NWEA, "Researchers Estimate Students Coming Back after COVID-19 Closures May Have Greater Variances in Academic Skills," press release, May 27, 2020.

<sup>5</sup>John Cronin, "How Many Students and Schools Actually Make a Year and a Half of Growth during a Year?" *Teach. Learn. Grow. The Education Blog* (June 16, 2016).

<sup>6</sup>Chiefs for Change and Johns Hopkins University Institute for Education Policy, "The Return: How Should Education Leaders Prepare for Reentry and Beyond?" (Washington, DC: Chiefs for Change, May 2020).

<sup>7</sup>Gwen Baker, Lauren Schwartz, and Amy Chen Kulesa, "Unfinished: Insights from Ongoing Work to Accelerate Outcomes for Students with Learning Gaps" (Washington, DC: Bellwether Education Partners, September 2019).

<sup>8</sup>Rebecca Cohen and Keith Wardrip, "Should I Stay or Should I Go: Exploring the Effects of Housing Instability and Mobility on Children" (Washington, DC: Center for Housing Policy and MacArthur Foundation, February 2011).

<sup>9</sup>New Classrooms, "The Iceberg Problem: How Assessment and Accountability Policies Cause Learning Gaps in Math to Persist below the Surface . . . and What to Do about It" (Washington, DC: New Classrooms, 2019).

<sup>10</sup>The New Teacher Project, "The Opportunity Myth: What Students Can Show Us about How School Is Letting Them Down—and How to Fix It" (Washington, DC: TNTP, September 2018).

<sup>11</sup>Michael J. Petrilli, "When It Comes to Catching Kids Up, Let's Stop with the Generalities," *Flypaper Blog* (Washington, DC: Fordham Institute, May 6, 2020).

<sup>12</sup>Lynn Olson and Craig Jerald, "The Big Test: The Future of Statewide Standardized Assessments" (Washington, DC: FutureEd, April 2020).

<sup>13</sup>Andy Hegedus, "Evaluating the Relationships between Poverty and School Performance" (Portland, OR: NWEA, October 2018).

<sup>14</sup>Jim Soland and Yeow Meng Thum, "School Effectiveness, Summer Loss, and Federal Accountability" (Portland, OR: NWEA, March 2019).

<sup>15</sup>Jim Soland, Yeow Meng Thum, and Gregory King, "Reconciling Long-Term Education Policy Goals with Short-Term School Accountability Models" (Portland, OR: NWEA, March 2019).

Abby Javurek is vice president of government affairs and partnerships at NWEA, where she leads State Solutions' vision and impact strategy. Jason Mendenhall is president of the State Solutions Division at NWEA. He leads the organization in developing student-centered, instructionally relevant statewide assessment solutions.

# We Should Listen to the Canaries

Students with disabilities and English learners rank high among those facing the greatest academic challenges in public education. Like canaries in the coal mine, these students are often the first to reveal how conditions in schools are not adequately supporting students' academic growth.<sup>1</sup> Their mandated participation in alternate statewide assessments offers insights into why that might be.

First, the testing requirement for these groups incentivizes state and local educators to teach these students appropriate academic content and keep them in classrooms with their age-appropriate peers. Keeping students with disabilities in these classrooms, rather than segregating them in isolated classrooms, helps them and their nondisabled peers in a variety of ways, including improved academic achievement and self-regulation.<sup>2</sup> Similarly, English learners and their English-proficient peers both benefit from academic and social interactions in ways that enhance their achievement and linguistic skills.<sup>3</sup>

Second, high-quality alternate academic assessment and English language proficiency assessments involve some degree of individualized administration: Students taking the test interact directly with their teachers as it is administered. There is no secret about what the tests are asking students to do, and teachers see and evaluate students' performance in the moment. As a result, teachers gain insights into what their students know and can do directly from the process of testing rather than having to wait until scores become available.<sup>4</sup> They can use those insights in their classrooms that very day, and students can reap immediate benefits from their assessment-drive interactions with their teacher.

## A Plethora of Tests

The break from testing that the pandemic instigated in spring 2020 only

renewed old questions about the purpose of the entire assessment system and qualms about the number of summative state tests. While the full repertoire varies to some degree across states, some components are common to all. All states administer assessments in English language arts and mathematics in grades 3 through 8 and at least once in high school as well as science assessments in at least one grade in each of the 3 to 5, 6 to 8, and high school grade ranges. All states also administer English language proficiency assessments to K-12 English learners each year.

Every one of these general assessments also has an alternate version for students with significant cognitive disabilities who cannot access the general assessments.<sup>5</sup> Some states also offer alternate academic assessments in languages other than English to allow English learners who are not yet proficient in English to demonstrate their knowledge and skills. In sum, a state must administer sixty assessments annually, with individual students taking as many as four in a given year.

## History of Alternate Assessments

The federal government has required academic assessment only since the fall of 1994, when the Elementary and Secondary Education Act (ESEA) was reauthorized as the Improving America's Schools Act (IASA). It required states to administer annual, statewide English language arts and mathematics assessments to all students in specific grades. And it took nearly a decade for all states to comply with that mandate.<sup>6</sup>

Through two subsequent reauthorizations—No Child Left Behind and the Every Student Succeeds Act—states have been required to set their own standards, create their own assessments, and design their own accountability systems within a set of parameters. No federal or national sets of standards, accountability

*How alternate assessments for students with disabilities and English learners can point us toward better systems for all.*

---

Ellen Forte

---

**A state must administer sixty assessments annually, with individual students taking as many as four in a given year.**



assessments, or accountability systems have ever been imposed.

Before IASA and the Individuals with Disabilities Education Act (IDEA) of 1997, there was no explicit requirement for educators to apply the same academic expectations for students with disabilities as for all other students. From the early 1970s until 1994, ESEA required assessments only in schools receiving Title I funding, required only testing of students directly receiving services paid for under Title I of ESEA, and allowed districts to choose which assessments to administer.

In those days, the options were mostly limited to tests yielding norm-referenced scores as opposed to scores associated with standards-based achievement expectations. If a district were using a large-scale assessment for accountability or program evaluation purposes, it could choose to use that assessment or an entirely different one for the Title I students in its Title I schools.

Alignment to standards was not required, and if some students—say, those with cognitive disabilities or without English language proficiency—happened to miss out on testing, *c'est la vie*. There were no negative consequences for not testing students, including those whose absence was essentially a blanket exemption based on their disability, language learning status, or even the likelihood that they would not perform well on the tests. Thus the very students who ESEA was meant to support, with the billions of dollars it directed to education agencies each year, could easily go undetected in evaluations of school performance.

IASA addressed this deficit by requiring states to test all students, regardless of whether they were receiving services paid for by Title I or attending schools receiving Title I funds. The message was that states, districts, and schools must hold the same high expectations for all students, including those who face added challenges to achieving academic success. IASA explicitly included English learners under the all-means-all umbrella.

To allow for all students to participate, a state must offer a general assessment, a general assessment with accommodations, and alternate assessments. Under IASA and its successors, alternate assessments must be based on the same academic content standards as the general assessments but

have different performance or academic achievement standards, which define what constitutes proficient performance, performance that exceeds grade-level expectations, and performance that does not yet meet those expectations.

While IASA introduced the requirement for testing all students, it did not impose consequences for schools' failure to do so. States simply had to report aggregated assessment scores for all students as well as scores that were disaggregated for several student groups, including students with disabilities and English learners, on state and local report cards. NCLB attached accountability to these reporting requirements by mandating that at least 95 percent of the students in each student group, such as students with disabilities and English learners, had to participate in the assessments. In addition, each student group—rather than just the total number of students tested—had to meet performance expectations for adequate yearly progress. Under NCLB, all meant all in terms of both participation and performance.

**Students with Disabilities.** Meanwhile, the 1970 reauthorization of ESEA included as its Title VI grants program the Education of the Handicapped Act, subsequently established as the stand-alone Education for All Handicapped Children Act of 1975. The act required local education agencies to provide a “free and appropriate education” for students with disabilities, which included individualized education programs (IEPs) for each student as well as access to the least restrictive environment (LRE) for their instruction. These requirements were extended and expanded when the act was reauthorized as IDEA in 1990.

In 1997 and 2004, IDEA was reauthorized, in part to align with IASA and NCLB requirements, respectively. Both reauthorizations required states and local school districts to provide alternate assessments suitable for students with significant cognitive disabilities for every test they require their students to take.

**English Learners.** Assessment requirements for English learners stem from a combination of legislation, memoranda from the Office of Civil Rights in the U.S. Department of Education, and federal case law.<sup>7</sup> Title VI of the Civil Rights Act of 1964 specifically prohibited “discrimination on the basis of race, color, or national origin” in federal or federally funded programs.

While that legislation did not explicitly mention English language proficiency, Congress subsequently passed the Bilingual Education Act of 1968 as Title VII of the 1968 ESEA reauthorization, which provided federal funding “for programs to help language-minority students overcome linguistic barriers that prevent them from meaningfully participating in their educational program.”

The NCLB reauthorization of ESEA introduced in its Title III a state education agency role for supporting English learners (then referred to as “limited English proficient” or LEP students) that extended the standards, assessments, and accountability components of the systemic reform model to English language proficiency. All states had to establish English language proficiency standards, assessments aligned with those standards, and an accountability model that used scores from those assessments to identify districts where English learners were not acquiring, or making progress toward acquiring, English language proficiency. ESSA retained these requirements but pulled them into Title I, which applies to all schools and school districts rather than only those receiving Title III funding for English language support services.

## Why Offer Alternate Tests?

The logic behind all these testing mandates stems from the model of systemic school reform that emerged in the 1980s and early 1990s.<sup>8</sup> Systemic reform seeks to shift the focus away from mere compliance with rules for the use of school funding and toward evidence that this funding was being used effectively to improve student achievement.<sup>9</sup> Implementing systemic reform means

- articulating expectations for student achievement in each grade in the form of academic content and academic performance or achievement standards,
- assessing student achievement using tests aligned with these standards, and
- evaluating schools based on their students’ performance on these tests.

From the beginning and across the reauthorizations of the federal legislation, three key words have been at the core of the testing

mandate: accountability, alignment, and all.

Accountability means that these tests must yield scores that states use to hold schools accountable for student learning. While the scores may also inform some curriculum or programming decisions, accountability uses are required. That is, states must use the scores to identify schools where students appear to be underperforming and provide resources to support students’ achievement in those schools.

Alignment means that the assessments must measure the range of knowledge and skill expectations defined in the state’s academic content and performance standards—not just some of them and not some random list of knowledge and skills. Tests that do not align with those clearly articulated academic expectations, which also should be used to guide classroom instruction, cannot yield information of use to teachers, administrators, or parents.

And all means all. All students in the tested grades must participate in the statewide assessment system. Some may despise large-scale testing or at least large-scale accountability testing of the type ESEA requires, often for justifiable reasons. But it is often the “all” requirement that gives many educators, parents, and policymakers pause. All includes every student in every classroom in every school in every neighborhood. All includes every student with disabilities and every English learner, even those who are not yet proficient in English. It is an uncompromising requirement bent on equity and remains a fundamental part of the federal policy logic.

As these long legislative histories reflect, one answer to why states keep testing students is that federal law and dozens of other legal memoranda and legal precedents—some of which are grounded in the Equal Protection Clause of the U.S. Constitution—demand it. ESEA provides significant funding to states and school districts, and student assessment has been a key element in holding state and local agencies accountable for how they use the millions of dollars they get each year. In return for the funding, these agencies agree to be held accountable for the achievement of all of their students.

But the more compelling answer is that alternate assessments benefit students and give incentives to schools to make sure that students with disabilities and English learners

---

**Three key words have been at the core of the testing mandate: accountability, alignment, and all.**

**Instead of debating how many tests are appropriate, state leaders might instead ask why there are not more tests that involve teacher-student interactions.**

have opportunities to learn and meet rigorous expectations.

There are legitimate concerns about the quality of some of the assessments for these groups. It's true that traditional methods for assessing these students' knowledge and skills often fall far short of what is necessary to elicit valid, reliable information about what they can do. Scores for these individual students may reflect more errors in targeting knowledge and skills than scores for other students. However, the benefit that students, their families, and their communities accrue through the use of inclusive strategies and support for their academic achievement outweighs concerns about testing precision.

High-quality alternate academic assessments have been recently developed, such as those administered by the Multi-State Alternate Assessment and the Dynamic Learning Maps consortia.<sup>10</sup> Recently developed academic English language proficiency assessments, such as Access for ELLs and ELPA21, are far more technically sound and rigorous than their predecessors.<sup>11</sup> Such progress in testing rigor would not have occurred without the ESEA policy mandates.

## What State Boards Should Take Away

Students with disabilities and English learners offer a window into what large-scale testing could look like for all students if state and federal policies loosen their footing in traditional testing models and allow for more innovative, student-centered approaches. The tenets of accountability, alignment, and "all" would still apply—and likely with greater success—if state leaders would embrace and expand the lessons of alternate assessment.

In fact, high-quality alternate assessment and English language proficiency assessment formats should prompt state boards of education to rethink the nature of general assessments. Instead of debating how many tests are appropriate, state leaders might instead ask why there are not more tests that involve teacher-student interactions, administered at times and in ways that directly support students' learning. Certainly, there are time and cost implications for shifting to more student-centered, performance-based assessment formats. But if

students, their teachers, and their parents can actually learn from the testing process itself, that seems an investment worth exploring. Especially when the techniques necessary for building such tests would also demand stronger alignment with learning expectations and deep attention to the equity and accessibility of learning and assessment contexts.

The models on which current statewide accountability tests are based reflect tradition more than best practice. Innovative, student-centered assessments that are grounded in instruction and elicit rich evidence of student thinking appear more likely to effect real, sustainable changes in student learning than assessments that are distal to and separate from classroom interactions.<sup>12</sup> ■

<sup>1</sup>Charlene Tucker and Ellen Forte Fast, "Autopsy of a Canary: Searching for a Theoretical Framework to Understand the Underperformance of Students from Poor Families and Students of Color," paper presented at the CCSSO Annual Large-Scale Assessment Conference, Snowbird, Utah, 1999.

<sup>2</sup>Thomas Hehir and Lauren Katzman, *Effective Inclusive Schools: Designing Successful Schoolwide Programs* (Indianapolis: Jossey-Bass, 2012).

<sup>3</sup>Diane August et al., "Recent Research on English Language and Literacy Instruction: What We Have Learned to Guide Practice for English Language Learners in the 21st Century," in Marilyn Shatz and Louise C. Wilkinson, eds., *The Education of English Language Learners: Research to Practice* (New York: Guilford Press, 2010).

<sup>4</sup>Ellen Forte, Rachel Quenemoen, and Martha Thurlow, "NCSC's Theory of Action and Validity Evaluation Approach," NCSC Brief #9 (Minneapolis: University of Minnesota, National Center and State Collaborative, 2016); Rachel Quenemoen, Claudia Flowers, and Ellen Forte, "The Curriculum, Instruction, and Assessment Pieces of the Student Achievement Puzzle," in Fred Spooner, ed., *More Language Arts, Math, and Science for Students with Severe Disabilities* (Baltimore: Brookes Publishing, 2014).

<sup>5</sup>General assessments are those academic or English language proficiency tests taken by most students and are distinguished from "alternate" assessments.

<sup>6</sup>William Erpenbach, Ellen Forte Fast, and Abigail Potts, "Statewide Educational Accountability under NCLB: Central Issues Arising from an Examination of State Accountability Workbooks and ED Reviews under the NCLB Act of 2001" (Washington, DC: CCSSO, 2003).

<sup>7</sup>Ellen Forte, *The Administrator's Guide to Federal Programs for English Learners*, 2nd ed. (Washington, DC: LRP, 2017).

<sup>8</sup>Michael Smith and Jennifer O'Day, "Systemic School Reform," *Journal of Education Policy* 5, no. 5 (1990).

<sup>9</sup>David Cohen, "What Is the System in Systemic Reform?" *Educational Researcher* 24, no. 9 (1995): 11–17, 31.

<sup>10</sup>Arizona Department of Education, "Assessment" website, <https://www.azed.gov/assessment/msaa/>; Dynamic Learning Maps, website, <https://dynamiclearningmaps.org/>.

<sup>11</sup>WIDA, "Access for ELLs," website, <https://wida.wisc.edu/assessment/>; ELPA21, "The ELP Assessment System," website, <https://www.elpa21.org/>.

<sup>12</sup>Linda Darling-Hammond, "Developing and Measuring Higher Order Skills: Models for State Performance Assessment Systems" (Washington, DC: CCSSO, 2017).





## Performance Assessments: Promises and Pitfalls

Statewide performance assessments have gone in and out of vogue for over 30 years. While embraced as authentic representations of student work, they often prove burdensome for standardized assessment purposes. They can be a strong addition to a balanced assessment system—when implemented well for the right purposes. State policymakers looking to bring performance assessments to their schools need to think deeply about their purpose and intended use.

A performance assessment may comprise one or more tasks. A performance task is any activity that asks students to do something to demonstrate their knowledge, understanding, and proficiency. Performance tasks yield a tangible product and/or performance that serve as evidence of learning.

Performance assessments let students show what they know in an applied fashion. They are commonplace in the arts, where students are judged on

*By learning from the past, state boards can add depth and relevance to their assessment systems.*

---

**Marianne Perie**

■  
**The amount of subjectivity in assigning and scoring the tasks led to low levels of reliability and incentives to game the system.**

painting, sculpture, musical performance, or dance. Even in academic subjects, performance tasks may be more representative of work done in the classroom, be it running science experiments, writing essays, or recreating an event in history.

But because they are difficult to standardize and expensive to score, states have often abandoned efforts to administer performance assessments. This article will describe their benefits and challenges.

## Learning from Past Attempts

Before Congress passed No Child Left Behind (NCLB) in 2001, three states were using performance assessments as part of their accountability systems: Kentucky, Vermont, and Maryland. Kentucky began requiring writing portfolios at grades 4, 8, and 12 in response to the 1990 Kentucky Education Reform Act. A portfolio consists of multiple samples of student work taken throughout the school year to demonstrate progress.

The policymakers' goal was to improve their students' writing and use the work to hold schools accountable for providing better instruction that increased the progress of all students. The initiative required professional development for all teachers, not just English teachers, to create and score writing tasks. However, the program ultimately failed due to the high-stakes nature of the consequences for teachers. Teachers were not incentivized to challenge students, leading to a wide degree of variability in the tasks assigned. Additionally, there were concerns with the reliability of the results, as teachers scored their own students.

Vermont implemented a similar program in 1991 for writing and mathematics. Schools were encouraged to use results for individual instruction and intervention, but aggregated scores were used at the state level for school accountability. They, too, struggled with standardizing tasks and implementing comparable processes for providing feedback and allowing revisions. As in Kentucky, the amount of subjectivity in assigning and scoring the tasks led to low levels of reliability and incentives to game the system.

The Maryland School Performance Assessment Program (MSPAP) was a performance-based assessment implemented in 1991 that covered reading, writing, language usage,

mathematics, science, and social studies. It was implemented in grades 3, 5, and 8, and the assessment tasks were unique in that they involved group work. This format led to results that could be generalized at the classroom, school, and district levels but did not provide individual scores. This program appeared to have some early success, with large majorities of teachers and principals reporting that they saw improved instruction. Test prep activities became much more broadly focused on application of knowledge. However, the large focus on writing across all subjects interfered with the interpretation of scores in subjects like math and science.<sup>1</sup> Ultimately, the MSPAP had to be terminated with the passage of the NCLB, which required student-level scores in grades 3–8.

## Current State Efforts

Currently, most performance assessments consist of writing tasks—either for ELA alone or combined with history/government/social studies—and computer simulation models for science, technology, engineering, and mathematics (STEM) subjects. New Hampshire has gained notoriety for being the first state to implement a performance assessment under the innovative assessment flexibility offered under the Every Student Succeeds Act (ESSA). The Performance Assessment of Competency Education (PACE) project, which began in 2014, includes locally developed, locally administered performance assessments tied to grade and course competencies determined by local school districts that are aligned with the state's challenging academic content standards. In each grade and subject, one common complex performance task, administered by all participating schools and districts, helps with calibration and provides evidence about the comparability of judgments related to student achievement across schools and districts. The state has continued their traditional statewide assessments in select grades and subjects to provide an ongoing audit of the innovative assessment system.<sup>2</sup>

Rhode Island has a performance assessment graduation requirement. However, it is intentionally subjective and not intended for statewide accountability. Since 2005, Rhode Island has required an individual learning plan to be designed at grade 6 and revisited throughout

high school. Students work with their teachers to determine whether they will develop a portfolio, exhibition, or capstone product in one or more subject areas selected by the student and approved by their teachers. The product is intended to demonstrate students' culminated applied learning skills and knowledge. Portfolios include performance-based entries required by the school district and some selected by the students. Exhibitions or capstone products tend to demonstrate deeper understanding of a selected topic. The state provided guidance on developing strong performance tasks, but districts ultimately set the graduation requirements.

## Guide to Adaptation

All these state stories provide lessons for state boards of education interested in adopting performance assessments in their states or recommending them to districts. Assuming past is prologue, performance assessments require a level of local flexibility to survive in the long term.

State boards should ask questions about the purpose, content area, and criteria for proposed performance assessments. Bringing in resources to enrich student learning is very different from introducing a new assessment in order to evaluate teachers or schools. State leaders should also consider whether a performance assessment should be purchased, built from the ground up, or should leverage existing performance banks and standardized rubrics. Likewise, training educators to develop good performance tasks and strong rubrics can have effects beyond a single assessment. Finally, when introducing performance assessment to a balanced system of assessments, educators will need training on interpreting and using the results to inform teaching.

**Purpose.** The first question policymakers need to ask is what the purpose of the performance task is. If the intent is to use results for school and district accountability under ESSA, there are much stricter guidelines for reliability and comparability. If the goal is to incorporate performance tasks into instruction to help teachers better assess the depth of student understanding or to allow for more authenticity, then the requirements are much more flexible. A third vector is the use of performance tasks for student accountability by incorporating them into grades or graduation requirements,

which requires high reliability, but the comparability requirement can be set locally.

Once the purpose is established, it will drive the remaining decisions. For instance, if the goal of adopting formative assessment is primarily for school use, then the focus should be on looking for banks of performance tasks and training on administering and scoring them. If the plan is to use them for accountability and the tasks need to be kept secure, then a vendor with experience in developing and scoring performance tasks would be a better choice. Again, New Hampshire has had success with a hybrid approach of common tasks plus locally developed tasks.

**Subject.** The second consideration is the subject or content area of interest. The area with the longest history of performance tasks is, of course, writing. Asking students to write an essay, either with or without a reading passage for context, is a task that states have been requiring for years. Within writing, there are several approaches. First, the rubric can be standardized and circulated for teachers to use throughout the school year. Then, only the writing prompt is a secure test question. This approach became more commonplace under NCLB, where the approach to writing was to have students respond to a passage they read on the test. The writing prompt could still be narrative, informative, or persuasive, with a rubric for each genre, but the score was based on organization, clarity, and ability to reference the passage sensibly.

Some groups are more interested in a student's ability to respond to feedback, so the score is based both on the final writing sample and on the growth from the first draft. Others are interested in a student's ability to research a topic and then write about it, so the task may be given over several weeks rather than as an on-demand prompt. This approach is often used in college courses. Again, the rubric is typically provided ahead of time.

One area that has garnered much research is whether students should be allowed to choose the prompt on which they wish to write. Because prior knowledge is an important component in the ability to write effectively, the theory is that if students choose a topic they are interested in, they will produce a better product. This approach should also lead to more

---

**Performance assessments require a level of flexibility to survive in the long term.**



equitable products, as students have varying exposure to different topics.

Research on Advanced Placement tests indicated that students who were given a choice of prompts tended to choose the one that looked easiest rather than the topic they were most drawn to. Instead, letting students choose the topic without first seeing the prompt should better match the theory and minimize the effects of unequal prior knowledge. More research is needed in this area, but allowing students some freedom in the topic would ultimately increase the fairness of the task.

Performance tasks in mathematics are most commonly developed as on-demand tasks in which students must perform several steps and show their work. Their score is based on process as well as product. A task with greater scaffolding to help struggling students often has several questions that build on one another to help guide the student to the final question. Other performance tasks involve a single scenario with multiple questions regarding a similar concept at varying degrees of difficulty. This latter approach is useful for classroom instruction to help teachers determine the depth of understanding within a single standard or concept.

With the release of the next-generation science standards, intense work began on developing computer-based simulations to test students process skills in addition to their content knowledge. Students may be given a scenario and allowed to try different approaches before answering a series of questions. For example, a physical science question may be about the relationship between a projectile and its velocity. Students may be given a simulation of a model rocket where they can vary the length of the tube, the shape of the nose cone, or the number of fins on the tail. After running the simulations, they then answer questions about the relationship between those characteristics and the distance it can travel before gravity pushes it into descent. The National Science Teachers Association gathered information on science simulations to provide teachers resources for using them in their classroom.<sup>3</sup>

When teacher evaluation became a part of the NCLB waiver requirements, much work was done in the area of performance tasks for teachers of subjects other than English and mathematics.<sup>4</sup> Teachers were either trained to create performance tasks or were given a bank

to select from. Typically, two to three items per year were administered, and the teachers were evaluated on how much each student grew in their abilities throughout the year. Denver Public Schools created such banks for teachers of all subjects and provided professional development for teachers to create and field test their own tasks that could then be added to the bank.<sup>5</sup> Although the efforts in many states were largely abandoned, the exercise taught us that teachers could be trained to both develop and implement such tasks effectively. More work was needed on demonstrating comparability across classrooms and schools, as these efforts focused primarily on within-classroom comparability, which was not fully aligned with the purpose of teacher accountability.

**Criteria.** Multiple repositories exist for performance tasks and rubrics. Additionally, companies offer professional development to train teachers to develop and field test their own tasks.<sup>6</sup> State policymakers may consider first providing access to a bank of performance tasks and training teachers on implementing and scoring them. Grant Wiggins and Jay McGighe have written extensively on how to evaluate the quality of performance tasks. They focus on the importance of content and process knowledge as well as the student's ability to generalize that knowledge. Any context should be age appropriate, and the rubrics should be standardized. Their guidance for rubrics suggest evaluating students' products based on whether they meet the requirements of the task, the accuracy and thoroughness of the task, the evidence of following effective procedures, and the quality of the organization of the product (figure 1).

Other possible approaches involve hiring a vendor to create customized tasks and rubrics or train teachers to develop the tasks and rubrics, pretest them, revise them, and implement them effectively. As with most processes, evidence that the vendor has successfully completed such work previously is important.

Performance tasks can bring great depth to assessments and often result in tests that actually measure what teachers want to teach. However, they can be subjective and costly, making them difficult to implement at a large scale. Adding performance tasks to more traditional standardized assessments or building a performance assessment for specific grades or subjects may be the best approach to

**Figure 1. Four Types of Criteria with Sample Questions**



Source: Grant Wiggins and Jay McTighe, *Understanding by Design: Guide to Advanced Concepts in Creating and Reviewing Units* (Alexandria, VA: ASCD, 2012).

creating a balanced assessment system that also includes professional development for educators to design tasks for use within instruction. Several states doing this work are combining some degree of standardization with room for customization. By doing so, they can meet the demands of comparability while allowing teachers and students to focus on areas of interest and produce their best work. ■

<sup>1</sup>Daniel Koretz et al., “Final Report: Perceived Effects of the Maryland School Performance Assessment Program,” CSE Report 409 (Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing, 1996).

<sup>2</sup>New Hampshire Department of Education, “Performance Assessment of Competency Education,” web page, <https://www.education.nh.gov/who-we-are/division-of-learner-support/bureau-of-instructional-support/performance-assessment-for-competency-education>.

<sup>3</sup>See, e.g., Argenta Price, Carl Wieman, and Katherine Perkins, “Teachers Use Simulations for Student Motivation, Content Learning, and Engagement in Science Practices,” *Science Teacher* 86, no. 7 (March 2019): 46–52.

<sup>4</sup>See, e.g., Lisa Lachlan-Haché, Ellen Cushing, and Lauren Bivona, “Student Learning Objectives as Measures of Educator Effectiveness” (Washington, DC: American Institutes for Research, 2012).

<sup>5</sup>The Commons, “Student Learning Objectives (SLOs) for SSPs,” web page <http://thecommons.dpsk12.org/Page/417>.

<sup>6</sup>Grant Wiggins and Jay McTighe, *Understanding by Design: Guide to Advanced Concepts in Creating and Reviewing Units* (Alexandria, VA: ASCD, 2012).

# States Experiment with Assessment through Innovative Pilots

The Innovative Assessment Demonstration Authority (IADA) program, written into the Every Student Succeeds Act, gives up to seven state educational agencies the opportunity to launch innovative assessment pilots in their public schools, with the goal of eventual statewide implementation.<sup>1</sup> As of August 2020, the U.S. Department of Education (ED) had approved five states' applications.

Beginning with Louisiana and New Hampshire in 2018, Georgia and North Carolina in 2019, and most recently Massachusetts in 2020, each state received a federal waiver for regular assessments so they could pilot assessment systems in select districts or schools over five years. Largely, the pilots focus on building a competency-based system that is more locally informed and better aligned to coursework.

## Louisiana

As the first state to receive an IADA waiver, Louisiana is furthest along, implementing its first pilot in the 2018–19 school year. With a through-course assessment model based on the Guidance curriculum, which about 70 percent of the state's school systems use, participating districts merge English language arts and social studies content in the same through-year assessments to cover students' knowledge of passages aligned to their social studies instruction or books they have covered in class.<sup>2</sup> Louisiana officials said they responded to research showing that “students need deep knowledge of a subject in order to think creatively or critically about it.”<sup>3</sup>

Aligning content in this way helps avoid testing students on their comprehension of passages dealing with subjects they have not previously

learned. Students thus can show they have mastered content gained in class as part of their assessment experience, said Chanda Johnson, deputy assistant superintendent at the Office of Academic Policy and Analytics, Louisiana Department of Education. “By connecting these assessments to that unit of instruction, we’re ensuring a more equitable experience for all of our students.”

As the pilot progresses, Louisiana will gather feedback through surveys and focus groups with teachers and school leaders to determine what they want to see in assessment data, what instruction they would change because of it, or what extra professional development they would like to receive. “[T]he best way to understand what’s best for kids is to talk to people who are the closest to those kids—make sure everyone is engaged in the process,” Johnson said.

Though unable to complete the 2019–20 pilot year due to the pandemic, participating schools were able to complete two of the three assessment windows before schools closed, thereby providing at least some data for the school year, Johnson said.

## New Hampshire

The second state to receive approval, New Hampshire applied in 2018 for its IADA pilot with its Performance Assessment of Competency Education (PACE) program. The PACE initiative, for which ED previously granted permission in March 2015, lets schools and districts provide multiple means for assessment other than standardized tests, with an emphasis on performance assessment. “The PACE system is based on the belief that a rich system of local and common ... performance-based assessments is necessary for supporting deeper learning as well



as allowing students to demonstrate their competency through multiple performance assessment measures in a variety of contexts,” according to New Hampshire’s IADA application.<sup>4</sup>

Established as a competency-based educational approach, PACE encompasses a combination of local, common, and state-level assessments administered to students in grade 3 English language arts (ELA), grade 4 mathematics, grade 8 ELA and mathematics, and grade 11 ELA, mathematics, and high school science. Each grade and subject receives one common performance task, called the PACE Common Task, which is developed and administered by all participating schools.<sup>5</sup>

“When PACE started off, they noticed the districts were all over the place and there wasn’t a lot of coherence on what proficiency is,” said Melissa White, administrator for academics and assessment at the New Hampshire Department of Education. PACE gathers input from local schools and teachers to help inform the state academic assessments with the content they cover in class.

While some states may have staff capacity sufficient to implement a new assessment system, others will rely on outside entities for operation and data management, as New Hampshire has. Thus, as response to the COVID-19 pandemic eats into budgets, states looking at implementing new assessment systems should be clear about the cost and the capacity requirements, White said.

## Georgia

ED approved the Georgia Department of Education’s application in 2019 to implement two innovative assessment consortia models: the Georgia MAP Assessment Partnership and the Putnam County Consortium.<sup>6</sup> The first partners with NWEA to develop MAP Growth for Georgia, a through-year assessment that leverages adaptive interim assessments to provide timely insights on students’ command of grade-level standards, measure academic growth, provide norm-referenced test results, and produce summative proficiency scores. The second group is working with Navvy Education LLC to develop an on-demand assessment system that provides real-time diagnostic data.

“Both assessment systems seek to assess students throughout the school year in order

to provide feedback to teachers while instruction is still occurring,” said Allison Timberlake, Georgia’s deputy superintendent for assessment and accountability. “The main difference is that the MAP Growth for Georgia model is a computer-adaptive assessment and the Navvy model is a standards-level assessment.”

Georgia’s objective in the IADA program is to leverage district work to explore alternate, innovative assessment solutions, Timberlake said. Ultimately, they will explore multiple models and work with stakeholders to select an assessment to implement statewide.

Georgia originally applied for three district consortia assessment systems, but the third—a model developed at Cobb County School District—did not meet ED’s requirements and selection criteria because they did not align with state academic achievement standards.<sup>7</sup>

## North Carolina

In October 2014, the North Carolina State Board of Education convened the Task Force on Summative Assessment to address concerns over the amount of time students spend on standardized assessments and the usability of the data at the classroom level. The task force recommended a through-grade assessment model called NC Check-Ins to replace the end-of-grade summative assessments.<sup>8</sup> After receiving feedback on the original IADA proposal, the department of public instruction revised its NC Check-Ins design to develop the North Carolina Personalized Assessment Tool (NCPAT), which measures students’ knowledge of content standards in three intervals across the school year. The first two NCPAT assessments, taken at the end of trimesters two and three, provide detailed feedback so classroom instruction may be tailored immediately to individual needs and serve as indicators to determine an appropriate starting point for developing the third NCPAT assessment. At the end of the school year, the final NCPAT assessment will feature multiple “clusters” of items based on results of the first two assessments.

“Matrix sampling will be used to design these mini clusters to ensure they are aligned to grade level test specifications, and across all NCPAT 3 clusters, the full breadth and depth of grade level content standards will be assessed,” states the North Carolina IADA application addendum.<sup>9</sup>

**States looking at implementing new assessment systems should be clear about the cost and the capacity requirements.**

With IADA approval in June 2019, North Carolina will expand this program from a first-year sample of participating students in through-grade assessments for mathematics and ELA to two districts participating in the second year, and a sample of at least 15 percent of students statewide in years three and four.

Early on, the task force engaged stakeholders as members and discussants, which provided vital input on the use, development, and technical requirements for a more balanced assessment system.<sup>10</sup>

## Massachusetts

Massachusetts received its waiver from federal assessment requirements in April 2020, and state leaders plan to focus on increasing access to deeper learning, piloting an assessment system in science and technology/engineering for grades 5 and 8.<sup>11</sup> The new design will combine the current Massachusetts Comprehensive Assessment System (MCAS) with a new hands-on session, where students will be assessed in dynamic, interactive simulations, much as they might experience a task in a science class. “We want to make sure the assessment reflects the kinds of tasks that we’re asking teachers to give their students,” said Sam Ribnick, special advisor for innovative assessments and data at the Massachusetts Department of Elementary and Secondary Education.

The IADA pilot will expand another Massachusetts pilot, the Kaleidoscope Collective for Learning, in which select schools and districts organized classroom instruction around deeper learning. About half of the IADA pilot group are Kaleidoscope schools.

Its pilot is part of a broader state effort to reduce disparities and make deeper learning more ingrained into the instruction all students receive. “Any state that’s getting into this process has to be really clear on what are the goals and the intended purpose of doing this and let that drive the initial conversation about what the new design will look like,” Ribnick said. ■

<sup>10</sup>Office of Elementary and Secondary Education, “Innovative Assessment Demonstration Authority,” webpage (Washington, DC: U.S. Department of Education, 2020), <https://oese.ed.gov/offices/office-of-formula-grants/school-support-and-accountability/iada/>.

<sup>11</sup>Louisiana Believes, “Louisiana Innovative Assessment Pilot,” fact sheet (Baton Rouge, LA: Louisiana Department of Education, N.d.), [https://www.louisianabelieves.com/docs/default-source/key-initiatives/louisianas-key-initiatives--innovative-assessment-pilot.pdf?sfvrsn=a6219f1f\\_18](https://www.louisianabelieves.com/docs/default-source/key-initiatives/louisianas-key-initiatives--innovative-assessment-pilot.pdf?sfvrsn=a6219f1f_18).

<sup>3</sup>Daniel T. Willingham, “Knowledge and Practice: The Real Keys to Critical Thinking,” issue brief (Knowledge Matters, March 2016), <http://knowledgematterscampaign.org/wp-content/uploads/2016/05/Willingham-brief.pdf>.

<sup>4</sup>New Hampshire Department of Education, “New Hampshire: Application for the New Authorities under the Innovative Assessment Demonstration Authority” (Concord, NH: author, 2018), <https://www.education.nh.gov/sites/g/files/ehbemt326/files/files/inline-documents/nhpceaapplication.pdf>.

<sup>5</sup>New Hampshire Department of Education, “Performance Assessment of Competency Education,” webpage (Concord, NH: author, N.d.), <https://www.education.nh.gov/who-we-are/division-of-learner-support/bureau-of-instructional-support/performance-assessment-for-competency-education>.

<sup>6</sup>Office of Elementary and Secondary Education, Georgia’s IADA approval letter (Washington, DC: U.S. Department of Education, 2019), <https://www2.ed.gov/admins/lead/account/iada/gaiaadaapproval2019.pdf>.

<sup>7</sup>Ibid.

<sup>8</sup>Office of Elementary and Secondary Education, “North Carolina: Application for New Authorities under the Innovative Assessment Demonstration Authority” (Washington, DC: U.S. Department of Education, 2018), <https://www2.ed.gov/admins/lead/account/iada/nciadaappdec2018.pdf>.

<sup>9</sup>North Carolina Department of Public Instruction, “Items that Require Additional Information or Revision in North Carolina’s Innovative Assessment Demonstration Authority Plan” (Raleigh, NC: author, April 2019), <https://www2.ed.gov/admins/lead/account/iada/ncapplicationaddendum.pdf>.

<sup>10</sup>Office of Elementary and Secondary Education, “NC: Application for New Authorities.”

<sup>11</sup>U.S. Department of Education, “Massachusetts Becomes First State to Qualify to Test New and Innovative Ways to Assess Student Achievement Next School Year,” press release (Washington, DC: author, 2020), <https://www.ed.gov/news/press-releases/massachusetts-becomes-first-state-qualify-test-new-and-innovative-ways-assess-student-achievement-next-school-year>.

## cont’d from page 11...A Shifting Landscape

recently published “Blueprint for Testing: How Schools Should Assess Students During the Covid Crisis,” drawing on recent work, to provide a ready reference for busy education leaders and policymakers as they plan for standardized testing now and into the future.<sup>7</sup> ■

<sup>1</sup>Lynn Olson and Craig Jerald, “The Big Test: The Future of State Standardized Assessment” (Washington, DC: FutureEd, April 2020).

<sup>2</sup>Deepa Srikantaiah, “How State and Federal Accountability Policies Have Influenced Curriculum and Instruction in Three States: Common Findings from Rhode Island, Illinois, and Washington” (Washington, DC: Center on Education Policy, October 2009).

<sup>3</sup>Lynn Olson, “The New Testing Landscape: State Assessment under the Every Student Succeeds Act” (Washington, DC: FutureEd, September 2019).

<sup>4</sup>Olson and Jerald, “Big Test.”

<sup>5</sup>Megan Kuhfeld and Beth Tarasaw, “The Covid Slide: What Summer Learning Loss Can Tell Us about the Potential Impact of School Closures on Academic Achievement” (Portland, OR: NWEA, April 2020).

<sup>6</sup>Thomas Toch, “Don’t Abandon Standardized Testing in Schools Next Year: Rethink It,” op-ed, *The Hill* (May 26, 2020).

<sup>7</sup>Lynn Olson, “Blueprint for Testing: How Schools Should Assess Students during the Covid Crisis,” (Washington, DC: FutureEd, July 2020).

# We the Media

## *Communicating in Anxiety-Laden Times*

I am an exhausted parent of two daughters, ages 2 and 5. Months of juggling work and home teaching have depleted my husband and me, and we are ready to see schools reopen. We wanted our eldest to start first grade with friends. In a school.

Fairfax County Public Schools began reopening planning in mid-June with a survey to parents and staff outlining three options: a virtual start to the school year (dependent on state reopening phases); a hybrid in-school, at-home learning model; and a 100 percent virtual model for students and teachers at high risk. They asked for our preferences and experiences with spring online learning. Days later, our superintendent presented a detailed plan to the local board. I was impressed by the thoroughness of the presentation, but I saw holes in the plan and wanted to retake the survey. I wasn't alone.

Two town halls and another school board work session later, the revised plan was presented: Families could opt for two days a week in school, with asynchronous learning at home for three, or for 100 percent virtual synchronous learning. The district asked families to commit for the full year—a hard ask. Not surprisingly, these options met with resistance and questions.

What safety precautions would schools be taking? How would a school manage a COVID outbreak or regular illness? How would the models be implemented? Would children be able to continue immersion, IB, AP? Safety aside, which learning model is best? And what about accommodations for special education? The answers to these and other questions were, “We don’t know. We need families’ preferences first.”

How frustrating. We wanted to make informed decisions, yet district and school leaders could not help. We turned to friends and family, whose advice may not have been accurate (tip: never underestimate the power of social media in swaying opinion). Many parents nationwide faced similar chicken-or-egg scenarios: Parents need good information to make confident choices; districts need to know parents’ intent in order to plan. This mutual dependency left parents to take leaps of faith.

Our ultimate decision came down to safety. And apparently, so did our school district’s. Only a week after we submitted our preference to keep our daughter home, our superintendent and school board reversed course and are starting the school year virtually for everyone. A wise choice, I think, but I wish I could get back those sleepless nights spent agonizing over our decision!

It will take a village to pull this off. Parents will need supports to supplement learning at home and help their children safely socialize. We will need continued flexibility from our jobs and child care. I especially hope schools leverage community resources to fill gaps for so many struggling families.

In the face of shifting, sometimes conflicting guidance, reopening schools is easier said than done. For communicating in uncertain times, my advice is this: Manage change with authority. State boards should make every effort to involve stakeholders early in planning. Be transparent when the information that undergirds the plan changes, and point stakeholders to trustworthy sources so they can make the best decisions possible. ■



**Renée Rybak Lang**  
Communications Director





**Robert Hull**  
President/CEO

## from the President's Pen

### *Looking in the Mirror*

State boards of education spend a great deal of time discussing evaluation processes and accountability systems. Based on their respective areas of authority, state board agendas are peppered with lengthy, substantive discussions on teacher evaluation, student performance, program analysis, as well as state chief and state education agency evaluation, which all lead to discussions about accountability based on the results of such data. All important topics, for sure.

However, most board agendas are eerily silent when it comes to the single most important area of evaluation and accountability within the purview of all state boards of education: self-evaluation. As the single-most powerful education policy group within each state and territory, state boards of education owe it to themselves, their fellow state policy leaders, and, most important, their families and students, to engage in a regular, thorough evaluation of their effectiveness as education policy leaders.

I recently had the privilege to lead a state board through an examination of their effectiveness, and it was powerful for both the board and for me as the leader of NASBE. The NASBE model of board effectiveness self-evaluation looks at four broad areas: 1) mission and goals, 2) state board agenda, 3) board operations, and 4) stakeholder engagement, each of which has a strong emphasis on equity and excellence. These areas are assessed through a state-specific survey of individual members and, when coupled with hard data on student achievement and program effectiveness, should lead to in-depth discussions that renew the board's focus on priorities, align direction

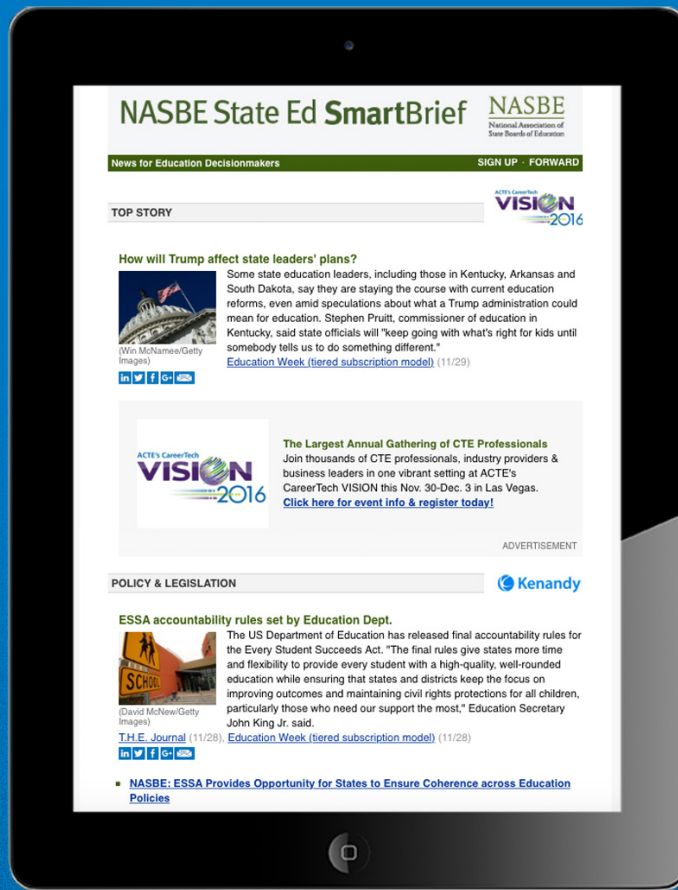
with those priorities, and increase efficiency and effectiveness. It also permits boards to highlight their accomplishments and recollect milestones of achievement. It is not only about growing as an effective board but also about taking time to stop and enjoy progress.

While the self-evaluation process may seem daunting given the limited time boards have to meet, it is well worth the investment, as it streamlines discussions moving forward and aids in prioritizing agenda items. Hence, greater effectiveness leads to enhanced efficiencies.

Once a board completes a self-evaluation, it is important to share those results with constituents and colleagues in the spirit of public transparency and to gain the support and trust of those the board serves. State boards are the citizen leaders in the education policy landscape and must be willing and eager to assess their own performance and effectiveness before prescribing such measures for teachers, schools, districts, and state agencies.

The relentless pursuit of equity and excellence begins at the board table—not at the schoolhouse door. State boards must heed the words of the great musical philosopher, Michael Jackson, and start with the man in the mirror. “I’m starting with the man in the mirror.... And no message could have been any clearer. If you want to make the world a better place, Take a look at yourself, and then make a change.”

Just as self-reflection is key to personal growth, board self-evaluation is the first and most crucial step in the collective journey of equity and excellence for *all* students. ■



The smarter way to  
stay on top of news for  
education decisionmakers.

## NASBE State Ed **SmartBrief**

A weekly e-newsletter dedicated to keeping  
education decisionmakers on top of the news.

**NASBE**  
National Association of  
State Boards of Education

**SIGN UP TODAY:** [SmartBrief.com/NASBE](http://SmartBrief.com/NASBE)





# College Board Opportunity Scholarships

---

A new scholarship program for all students.  
Earn up to \$40,000 for college.

**[cb.org/opportunity](https://collegeboard.org/opportunity)**

© 2020 College Board.