



—Patat/Stock/Getty

# Assessment

## EDITOR'S NOTE

The field of Assessment continues to evolve and adapt. This Spotlight will help you learn about formative assessments; examine the debate over standardized testing; understand research on the link between food stamp disbursements and students' exam scores; gain insight on interpreting state test results; and explore the value of federally mandated state tests.

How Can Teachers Better Understand Students? A New Breed Of Assessment Will Try to Help..... **2**

How Two Years of Pandemic Disruption Could Shake Up The Debate Over Standardized Testing..... **3**

Timing of Food Stamps Can Affect Students' Test Scores, Study Finds..... **5**

State Test Results Are In. Are They Useless?..... **6**

## OPINION

What Federally Mandated State Tests Are Good For (And What They Aren't)..... **8**



Researchers hope to create new assessments to help teachers gain deeper insights into the identities and strengths of their students, like these 6th graders at Cape Cod Lighthouse Charter School in East Harwich, Mass.

Published July 21, 2021

## How Can Teachers Better Understand Students? A New Breed of Assessment Will Try to Help

By Catherine Gewertz

**R**esearchers and education entrepreneurs launched a project that aims to create a new breed of assessments that will offer teachers a window into young students' emerging identities and build on their strengths to enhance their learning.

The idea behind the program, called Assessment for Good, is to help teachers understand how students 8 to 13 years old are starting to see themselves in the world, culturally, socially, and academically, and how that constantly shifting sense of self takes shape—and can support—learning.

Leaders of the initiative want to create formative assessments that teachers could weave frequently into classroom instruction. Those tools and strategies would explore many facets of students' identities, from how they think of themselves as learners to their race, gender, or disability, so they could better provide instruction that values the diversity of all their students, Temple Lovelace, the project's director, said in an interview with Education Week.

"We want to take an identity-affirming,

**“**We want to take an identity-affirming, strength-based approach. We want to look at what threads are emerging, what we can build on.”

**TEMPLE LOVELACE**

Project Director, Assessment for Good

strength-based approach,” said Lovelace. “We want to look at what threads are emerging, what we can build on.”

Before leading Assessment for Good, Lovelace was an associate professor of special education and assessment at Duquesne University. One of her areas of focus was “eco behavioral assessment,” which tries to measure how ever-shifting factors in a student's envi-

ronment can affect their learning and behavior. She has a particular interest in reducing the disproportionate referral of Black students to special education.

It's too early to know details of what the assessments will look like. Its organizers issued requests for information and proposals, hoping that teachers, researchers, and others will jump in with ideas to get the ball rolling.

### Conducting research in new ways to address big problems

Assessment for Good is part of a larger initiative that's focused on tackling education problems that disproportionately affect Black and Latino students, and those from low-income families. That initiative, the Advanced Education Research & Development Fund, or AERDF, is supported by \$200 million from three foundations that work in the K-12 sector: The Bill & Melinda Gates Foundation, the Chan Zuckerberg Initiative, and the Walton Family Foundation.

The first \$50 million, awarded in 2019 by the Gates and Chan Zuckerberg philanthropies, supports a project called EF+Math, a cluster of research efforts that aim to enhance the executive functioning skills of Black and Latino students in grades 3-8 to boost their math proficiency.

With a new, additional \$150 million from Gates, Chan Zuckerberg and Walton, AERDF was created to house EF+Math and Assessment for Good. By the end of 2023, three more projects will be added, all focused on addressing stubborn educational problems that disproportionately affect Black, Latino and low-income students, AERDF CEO Stacey Childress, who is also the CEO of the NewSchools Venture Fund, said during a call with reporters. Evaluations by independent researchers will track the effectiveness of each project, she said.

The AERDF projects will use an “inclusive R&D” approach, bringing together educators, researchers, and assessment designers from the start, “with all perspectives respected equally,” Childress said. They will also be guided by advisory councils that include educators, students, and caregivers.

The Assessment for Good project enters a landscape in which schools increasingly want to assess many more student characteristics than just math or reading ability. Many schools have been trying to teach and test social and emotional skills such as self-awareness, self-management, and responsible decisionmaking.

Schools are already using an array of products to measure those kinds of skills, such as

—Elise Amendola/AP



# Interrupted Learning and the Impact of Assessment

The need for assessment is especially critical to support elementary and secondary students facing interrupted learning challenges in literacy and math. Without evidence-based assessment solutions, it can be challenging to identify which skills students need help developing or if instruction and intervention efforts are even working. Backed by research, the Acadience® Learning suite of solutions empowers educators with the data they need to identify areas of struggle, monitor student progress, and modify instruction or intervention to get pre-K–8 students on grade level in reading and math.



Authored by research experts Dr. Roland Good and Dr. Ruth Kaminski, and built upon more than 30 years of research, Acadience reading and math assessment solutions include:

**Acadience® Reading Pre-K (PELI):** Storybook-embedded assessment used for ages 3–5 to measure pre-literacy and oral language skills needed for kindergarten

**Acadience® Reading K–6:** Designed to predict early reading success and identify students experiencing difficulty in the acquisition of foundational literacy skills

**Acadience® Reading 7–8:** Universal screening and progress-monitoring assessment that measures the acquisition of content-area literacy skills for grades 7–8

**Acadience® Math K–6:** Math assessment used to identify students experiencing difficulty acquiring foundational math skills

Through quick and reliable measures, Acadience Learning assessments and tools provide individualized data, allowing educators to act early to help students realize their full potential.

See which Acadience assessment solutions are right for your school or district.

► [Learn More](#)

student surveys, teacher rating scales, and performance-based assessments. Clark McKown, an associate professor of behavioral science at Rush University in Chicago, who has studied social-emotional learning assessments, said it could be difficult to measure students' emerging identities.

"Identity is such a highly complex and abstract concept," said McKown, whose compa-

ny, xSEL Labs, makes a performance-based social-emotional learning assessment for schools. "It could be helpful to educators. But how to define it, much less assess it, is a little unclear to me."

Scott Marion, the director of the Center for Assessment, which advises states and districts on testing systems, said that understanding—and nurturing—students' identi-

ties could be a powerful way to enhance their learning. But it remains to be seen, he said, whether the new tools can provide information teachers can use.

"Ultimately, the question about assessment always is, what can I do now?" Marion said. "There has to be a productive way to act on the information you get. Otherwise, what's the point?" ■

Published March 5, 2021

## How Two Years of Pandemic Disruption Could Shake Up The Debate Over Standardized Testing

By Evie Blad

**T**he week the U.S. Department of Education told states it wouldn't issue blanket waivers from mandated annual assessments, the creators of a national guide instructing parents on how to opt their children out of the standardized tests reported a spike in web traffic to the site.

"Parents are hopping mad," said Bob Schaeffer, the interim executive director of FairTest, an organization that promotes testing opt outs and created the guide. "If schools don't cancel the tests, parents will."

Advocates for testing — including civil rights organizations, a vocal group of lawmakers, and some educational leaders concerned about equity — say such suggestions may be overblown. But, after states cancelled tests entirely in 2020, some of those same advocates fear that two consecutive years of disruption in state testing—be it through opt outs, modifications, or complete cancellations — could amount to the country taking its foot off the gas in its commitment to broad assessments.

"What I fear is we just don't know enough right now," said Cheryl Oldham, vice president of education policy at the U.S. Chamber of Commerce, which has advocated for federal testing mandates. "We just won't know the implications of this year for a while."

Although President Joe Biden criticized high-stakes testing as a candidate, one of his Education Department's first acts was to leave the tests in place, even as many testing opponents argued that unprecedented disruptions caused by the COVID-19 pandemic gave a good reason to cancel them.



“  
I just think we have a moral responsibility to understand how all of our students are doing, where we are falling short...”

**U.S. SEN. PATTY MURRAY**  
D-Washington

In addition to concerns about the reliability of scores for tests given during the pandemic, schools will face a host of logistical challenges, like deciding how to test remote learners, finding off-site space that allows for greater social distancing, and finding adults to supervise testing at a time when staffing is already a challenge.

Testing supporters were heartened when the Education Department said in its Feb. 22, 2021 guidance that states must conduct tests. But the guidance provides a lot of wiggle room, allowing states to bypass requirements that they use the

scores to rate schools, to delay when they are administered, and to forgo the requirement that 95 percent of students participate, a change that could open the door for more opt outs.

States are still sketching out their response to the department's directive. While some have committed to conducting full assessments in person, others have said they will need to make changes to address the realities of the pandemic.

Those changes could leave a big hole in the data states use to track schools' performance over time and become talking points in debates about the future of testing, those testing supporters have said.

And some critics of testing hope they are right.

### An alliance in support of assessment

Congress reaffirmed its commitment to test-based accountability when it passed the Every Student Succeeds Act in 2015 after years of debate. The federal education law maintained many of the testing mandates that were the hallmarks of its predecessor, the No Child Left Behind Act.

Under ESSA, states have to test students in reading and math in grades 3-8 and once in high school. They also have to break down the resulting test scores to track results for targeted populations, like English-language learners, various racial groups, and students from low-income families.

Critics of NCLB had argued that it led schools to place too much emphasis on tests and to use their results in punitive ways. Under ESSA, lawmakers aimed to address those concerns by giving states more flexibility in

how scores are used and by allowing them to limit the amount of time schools spend testing. They also created a program to pilot new innovative assessments.

But an alliance that included civil rights groups, business organizations, and prominent congressional Democrats successfully pushed for Congress to reject more-dramatic changes to testing mandates when it passed ESSA.

Broad annual data are necessary to ensure schools are serving all students adequately, they insisted. And standardized tests offer consistency that other forms of feedback, like teacher observations and classroom assignments, may lack, they said.

And, as the Biden administration considered how to handle testing, many of those groups and lawmakers revived the same arguments. They insisted that, after fits and starts of in-person instruction and a full-year of remote learning in some areas, many parents will find value in gauging what subjects their children mastered and where they may need more help. And the data may be necessary to help direct federal aid, including an additional \$128 billion in K-12 relief, much of it targeted to responding to the effects of interrupted learning time, they've said.

"I just think we have a moral responsibility to understand how all of our students are doing, where we are falling short, and we have to use data to make sure that we are doing the right thing and sending the dollars to where they are needed the most," U.S. Sen. Patty Murray, a Washington Democrat and chairwoman of the Senate education committee, told Education Week in December 2020. "That's called education equity."

But in a sign of splits within the Democratic Party about education policy, newer members of Congress, like Rep. Jahana Hayes of Connecticut and Rep. Jamaal Bowman of New York, both former educators, have said it doesn't make sense to conduct tests in the 2021-22 school year.

Hayes, asked in late February 2021 if she thought the Biden administration's decision not to grant blanket waivers from standardized tests undercut its other efforts to assist schools, spoke extensively about how tests wouldn't help students or educators, and wouldn't capture the real impact of COVID-19.

Hayes shares that viewpoint with national teachers' unions and some Republicans. They include Sen. Richard Burr of North Carolina, the ranking member of the Senate education committee, who called upon U.S. Secretary of Education Miguel Cardona to cancel the tests during Cardona's confirmation hearing.

### How much flexibility should there be from testing requirements?

As states sketch out their responses to the federal directive, it remains unclear exactly how much flexibility the Education Department will allow.

While some states have committed to carrying on with testing plans, leaders in some places, including the District of Columbia, Michigan, and Pennsylvania, have said they will push forward with requests to cancel the tests entirely.

Groups that support test-based accountability have acknowledged the unusual circumstances created by the health crisis, but they've also urged states to do as much as is practical to conduct assessments.

"In our minds, the right thing is you've got to make every effort to test as many kids as possible," said Oldham, of the U.S. Chamber of Commerce. "In this year, of all years. Are you kidding me? This is a time when we absolutely have to know."

Testing experts like Andrew Ho, a professor at the Harvard Graduate School of Education, have said there are ways to present this year's test scores that would add helpful context for the public.

But testing critics have organized campaigns to urge states to push for as much flexibility as possible, including full waivers. They've also called on states to pause state-level accountability efforts, like use of reading test scores to determine if students should be promoted beyond the 3rd grade.

Opt-out advocates have urged mass disruption of state tests over the last decade, citing concerns about how high-stakes tests are conducted and how their scores are used to rate schools and educators. And they've reported a fresh surge of concern about testing during the pandemic.

It's unclear how many students will sit out testing. In a February 2021 parent poll conducted by the National PTA, 52 percent of respondents said they favored end-of-year testing, and 60 percent said they are worried their child is behind and want more information on where their child is academically.

But signs point to pockets of opt-out interest in some areas. And, if enough students sit out, it could amount to a de facto test cancellation, especially if the population of students who do participate doesn't mirror the demographics of the student body as a whole, some testing experts have warned.

New York City's departing Chancellor Richard Carranza encouraged parents to consider opt outs at a Feb. 25, 2021 news confer-

“

In our minds, the right thing is you've got to make every effort to test as many kids as possible. In this year, of all years. Are you kidding me? This is a time when we absolutely have to know.

#### CHERYL OLDHAM

Vice president of education policy at the U.S. Chamber of Commerce

ence, and said the school system would make them aware of their rights.

"We do not want to impose additional trauma on students that have already been traumatized," said Carranza, who has since announced plans to resign at the end of March 2021.

And, if new converts to the opt-out movement engage with activist organizations and join their email lists, they may play a more active role in debates over testing in the future, said Oren Pizmony-Levy, a professor of international and comparative education at Columbia University who has studied the opt-out movement.

"Once you change a habit, once you challenge something that is taken for granted, you all of a sudden start to question this behavior," he said.

Congress has not expressed any appetite to reauthorize ESSA or to rethink its testing requirements in the near term, but Pizmony-Levy cites past research that found that people who engage with opt-out movements are more likely to contact their congressional representatives when conversations about education policy occur.

### Is there a middle path in the testing debate?

Some veterans of the testing debate suggest that two years of disruption will lead parents, policymakers, and educators to tackle exactly what standardized state assessments should look like and how they are used.

# The Power of Progress Monitoring

By **Acadience Learning**

The research and assessment measures and tools authored by Dr. Roland Good and Dr. Ruth Kaminski have a proven track record of successfully helping educators determine if and where students are struggling.

Leading their webinar, "The Power of Progress Monitoring: Considerations for Reading Instruction and Intervention," the duo emphasizes the impact progress monitoring has on a student's pathway to literacy.

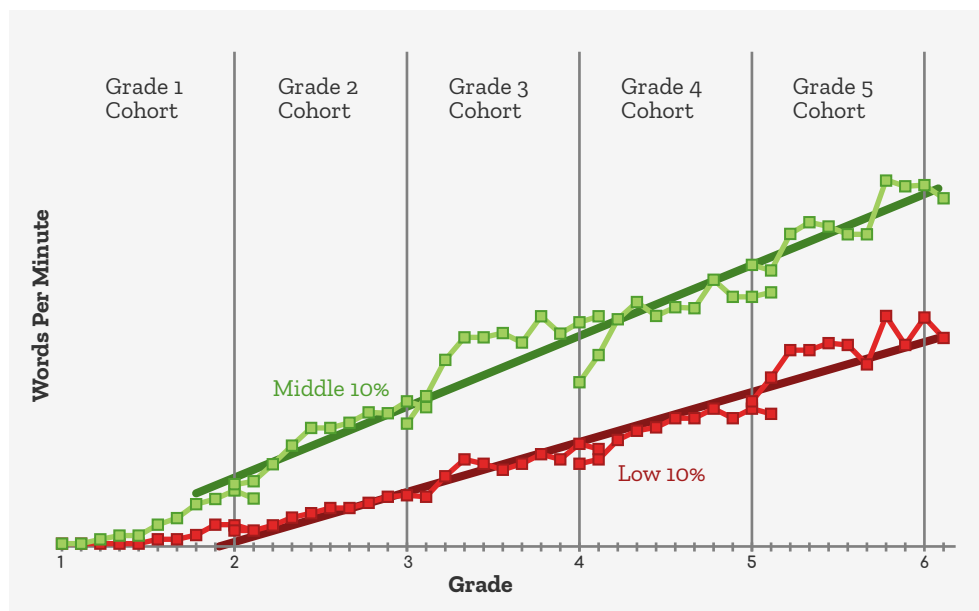
## What's Known About Learning Trajectories

As presented in the webinar, this chart maps the average trajectory for the lowest 10% of students and the average trajectory for the middle 10% of students from grades 1–6. Each group starts in the same place and improves over time, but the low group doesn't catch up to the middle group. In fact, the gap between them noticeably widens.

Learning trajectories are established early. Students on low trajectories tend to stay that way and fall further and further behind their peers. To catch up, students in the low group need to learn at a faster rate of progress than the average group.

## How to Change Trajectories

Still, with the right methods of intervention and instruction, teachers can change learning trajectories for the better. Drs. Good and Kaminski offer these guidelines for



doing so:

- Focus assessment on indicators of important outcomes
- Focus instruction on essential early literacy and reading skills
- Use ongoing assessment to make educational decisions to improve outcomes for all students

## The Role of Progress Monitoring

Research shows that providing goals for where students should be is an important part of changing outcomes. These goals should be set early on after identifying which students need support.

Then, educators should focus on essential reading skills, monitor progress, and use assessment information to alter instruction as needed.

Regarding the frequency of this process, Dr. Good says that the more concerned an educator is for a student, the more frequently that student should be monitored. However, he adds, it's important not to do so more than once a week.

"This is not about testing," says Dr. Good. "It's about teaching. Progress monitoring is a part of instruction and intervention." As part of a teacher's overall system, it can help prevent reading difficulties and disabilities, including dyslexia.



### Progress Monitoring and Dyslexia

The biggest impact a teacher can have on students with or at risk for dyslexia is early in their learning. If a student is provided with effective classroom instruction but is challenged by essential early literacy and reading skills, it's critical to begin progress monitoring. From there, if they experience a sustained lack of adequate progress in learning these skills, they can be screened for dyslexia.

Teachers should use an outcomes-driven model for decision-making that emphasizes prevention, early intervention, or, if needed, remediation. Educators should prioritize the things that can be changed, and administrators should empower teachers by providing them with the resources to follow through.

### Evaluating the Support Provided

It's not enough to evaluate the student. Educators must also evaluate each other and the instruction the student is receiving because it's wrong to assume all teachers are being effective.

"We're progress monitoring to evaluate the effectiveness of the support because the support is the thing we can do something about," says Dr. Kaminski. "The idea behind the power of progress monitoring is that if the support isn't working, you do something to modify it."

Just because a teacher makes a decision for the student does not mean it will be the right one. Assessing educators is how to find the instruction method that works best for that individual.

“  
**This is a circular process where I keep planning, implementing, evaluating, and revising until it works for the student.**

—Dr. Roland Good

”

### Assessment Amid Uncertainty

Reflecting on the pandemic and how it's affected assessment, Dr. Kaminski offers that it's never been more important to progress monitor students. "We have procedures for administering Acadience® Reading remotely, and teachers have done it," she says. "We're not going to lower benchmarks."

One of the advantages of Acadience is it provides a direct measure of a student's skills, which is critical in a remote or partially remote setting. Teachers need clear, vivid information about where students are and what should be taught. Acadience provides that.

### 'Our Mission and Our Job'

No matter the setting, as Dr. Good puts it, it's an educator's mission and job to change a student's learning trajectory. Acadience Reading is one tool that can be used to improve outcomes because it provides actionable assessment data through progress monitoring.

"As a parent, quite frankly, I felt I never got information about how well my child was doing, where they needed to be, if they were making adequate progress toward it," says Dr. Good. In part, that's why he and Dr. Kaminski have authored Acadience Reading. They aim to make goals and feedback more tangible for both student and teacher so that progress can be monitored, instruction can be tailored, and student achievement can be accelerated. ■

Dig deeper into this important conversation with Dr. Good and Dr. Kaminski by watching the webinar.



Watch Now

Good, R. H., Simmons, D. C., & Smith, S. B. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School Psychology Review*, 27, 740-753. [Joint publication with Educational and Child Psychology.]

“If we don’t [take] this opportunity to think about how we want to make our system better, we really are missing the boat on this one,” said Chris Minnich, the CEO of NWEA, a nonprofit organization that helps develop assessments. Minnich was also the executive director of the Council of Chief State School Officers when ESSA was enacted.

In response to the federal guidance, some states have suggested they will delay testing until fall 2021. Officials in New Mexico have requested permission to test a representative sample of students, an idea that has been float-

ed in broader policy discussions in the past. In South Carolina, officials have asked to use formative assessments conducted throughout the school year in place of a year-end, summative test.

That’s an idea NWEA has promoted, working with states like Georgia through ESSA’s innovative assessment pilot. But reverse engineering such an approach—by deciding on the back end to use those incremental tests for year-end results—could be imperfect, Minnich said.

Some critics of what Minnich calls

“through-year assessments” say formative and summative assessments serve distinct purposes that should not be mingled. But they acknowledge that states’ creative efforts may provide talking points for future discussions.

“It introduces the next chapter in an ongoing discussion of the proper type of assessment, the proper function of assessment, and ultimately how that benefits students,” said David DeSchryver, the senior vice president of Whiteboard Advisors, an education consulting firm. “How does that information support equitable educational opportunities?” ■



Food stamps can be used in some farmers' markets, as at this one in Topsham, Maine. New research shows a link between timing of the aid and student performance on key tests.

Published February 19, 2021

## Timing of Food Stamps Can Affect Students' Test Scores, Study Finds

By Stephen Sawchuk

**S**tudents who took a college-entrance exam more than two weeks after their families received food stamps scored more poorly than those who took the test soon after receiving the aid, a research paper finds.

The results have implications for public policy, the researchers argue. Other social science research has shown that households tend to increase their food spending and consumption immediately after receiving food stamps, but then taper off before their next disbursement. In effect, the research indicates, stu-

dents with the misfortune of a bad testing date are probably hungrier and not as able to put their best foot forward on the high-stakes SAT.

While the declines in test scores are modest, they can have very real consequences for low-income students, who are less likely to attend a four-year college right after high school, the research concludes.

The study is among the first to connect the inner workings of vital nutrition policies to specific education outcomes.

“It is a huge life impact for a small, random policy that students had absolutely no control over,” said Timothy Bond, an associate professor of economics at Purdue University and one

of four researchers who conducted the study. “We have a quirk with a huge and random impact on people’s long term lives. It’s just sort of stunning to think about.”

It also potentially raises new questions about the relative importance of college entrance exams, which have come under increasing criticism in some quarters.

States, not school districts, control disbursement rules for food stamps, formally known as the Supplemental Nutrition Assistance Program or SNAP. But schools could potentially minimize the effects of SNAP benefit timing by offering the SAT during the week—when low-income students typically receive school meals—rather than on Saturdays.

In all, the findings underscore the link between child nutrition and test scores, even as the COVID-19 pandemic lays bare the extent of food insecurity among Americans.

Fewer students are currently signing up for school meals, probably for logistical reasons amid the pandemic, and more families say they sometimes go without food, troubling signs that have prompted a federal response.

### ‘Winners and losers’ when test dates and food stamp disbursements don’t align

The research was released as a working paper from the National Bureau of Economic Research. Bond’s Purdue colleague Jillian Carr, as well as Analisa Packham of Vanderbilt University and Jonathan Smith of Georgia, also contributed to it. It has not yet been published in a peer-reviewed journal.

States use different ways to determine which day of the month a family receives its food stamp benefits; some use case ID numbers, while others use the first letter of the family surname. The researchers limited



their study to the seven states that use the latter policy.

They collected data on SAT scores, college attendance, and college selectivity data for high school cohorts between 2009 and 2014 from a variety of sources, including the College Board, which administers the SAT.

No database directly connects SNAP and education data, so the researchers used household income information reported by the students, tallies of students who received fee waivers on the exam, and U.S. Census data to determine which students were most likely to be low income and to qualify for SNAP benefits.

Then the researchers used disbursement schedules to see how students who received their SNAP benefits just before an exam date fared on the test, compared to those who received them just after, controlling for race, gender, and other factors that could skew the results.

They found that students in households receiving less than \$60,000 in annual income who sat the exam at the end of their benefit period scored about six points lower, an effect size of about .06 of a standard deviation.

That is generally considered a modest effect size, and social scientists have debated how relevant such effects are for public policy. But the researchers make a case for why: The declines were concentrated among lower-achieving students, not higher achieving ones.

“We’re talking about kids who are already at the margin,” Bond said. “These are kids where 10 points could put them above an automatic admission threshold at their state flagship colleges.”

It’s “a small thing,” he said, but the randomness of benefit cycles yields “a very arbitrary selection of winners and losers.”

### Small effect has large consequences

Indeed, the researchers found evidence that low-income students who took the exam near the end of the benefit cycle also were more likely to attend a two-year college, less likely to enroll in a four-year one, and enrolled in less selective institutions. They estimated that about 1,150 fewer students initially attended a four-year college in that time period, out of about 170,000 students in the overall sample, because of the test timing.

There are various ways to minimize the

harm to students who don’t end up taking the SAT on a date that aligns well to the SNAP benefit cycle. A more complicated one would require states to smooth out benefits, to avoid peaks and valleys throughout the month.

But school districts could also play a role by agreeing to hold the SAT during the school week, rather than on weekends, for the simple reason that low-income students would likely be able to access school meals on those days.

About half of the class of 2020 took the exam on a school day, according to the College Board, and the nonprofit has encouraged schools to offer the SAT during the week; about 10 states also provide the exam at no cost during the school day. 46 percent of school day test-takers also attend high-poverty schools, the organization said.

Testing centers could also consider offering breakfast when the exam is given on a Saturday, Bond suggested, though that would incur some new costs.

The study also has implications for testing research in general.

Other studies have shown that a constellation of factors, like heating and air conditioning, exposure to violent crime, and sleeping patterns, can all affect how students do on measures of achievement. College-entrance exams in particular have been criticized because of the link between family income and scores; the University of California last year made the decision to phase out the exams by 2025.

The pandemic has further upended the industry, with scores of colleges saying they will forgo the scores for college admission in the 2021-22 school year.

But many of the other factors that play into college admissions—selective academic tracks, extracurriculars, essay preparation, and so forth—are also highly dependent on family resources. ■

Start your day with us.

## EdWeek Update

Stay on top of everything that matters in K-12

[SIGN UP](#)

The graphic shows a smartphone displaying the EdWeek Update newsletter. The newsletter header reads "EdWeek Update" and "Education Week's EdWeek Update". Below that is a photo of a man in a suit, and the main headline is "Building a Community for Black Male Teachers". Other headlines include "Colorado's new teacher standards" and "How to lead in a hybrid classroom".

Published October 21, 2021

## State Test Results Are In. Are They Useless?

By Catherine Gewertz

**E**ducators have been bracing for them, and now they’re here: the first state test results since COVID-19 interrupted K-12 schooling. Districts, states, and schools are poring over the data from spring 2021 tests, hoping to understand exactly how—and how badly—the pandemic affected children’s learning.

But even though educators are hungry for insight, assessment experts are urging caution. The 2021-22 school year, more than any in recent memory, calls for extreme care and restraint when analyzing statewide test scores, drawing conclusions, and taking action, they say.

Like schooling itself, standardized testing was deeply disrupted in many ways, which may have distorted the meaning and utility of the results. In some cases, state test data will be virtually useless, the experts say. In others, with thoughtful analysis, the data can yield insights that could help leaders and educators allocate resources and help children rebuild academic muscle.

Here are some key considerations—and important cautions—for state, district, and school leaders, and teachers, to bear in mind as they review state test scores.

### A lot happened with state tests in 2021 that could affect the results

In 2020, the U.S. Department of Education allowed states to skip federally required assessments. In 2021, however, states had to administer those tests. But that doesn’t mean it was business as usual.

In a handful of states, some students took tests remotely, while others took them in person. Massachusetts, for instance, allowed students in grades 3-8 to take remote tests if their schools were in remote learning mode, and more than 15 percent of those students did so.

Some states made other changes to their testing regimens. A few gave shortened versions of their tests. Colorado gave its English/language arts test only in grades 3, 5, and 7 and its math test only in grades 4, 6, and 8. In

California, some districts gave the Smarter Balanced test, and others used assessments of their choosing.

Many states saw fewer students take the test than usual, though, and that is the factor poised to exert the most widespread influence on the validity and comparability of state test data. According to the Center for Reinventing Public Education, which has been monitoring states' responses to COVID-19, of the 30 states that have released test results so far, only 14 reported test-participation rates of 90 percent or more.

Some states reported participation rates as low as 10 percent (New Mexico) and 30 percent (Oregon). Participation also varied markedly within states: Colorado reported regional participation rates ranging from 51 percent to 88 percent.

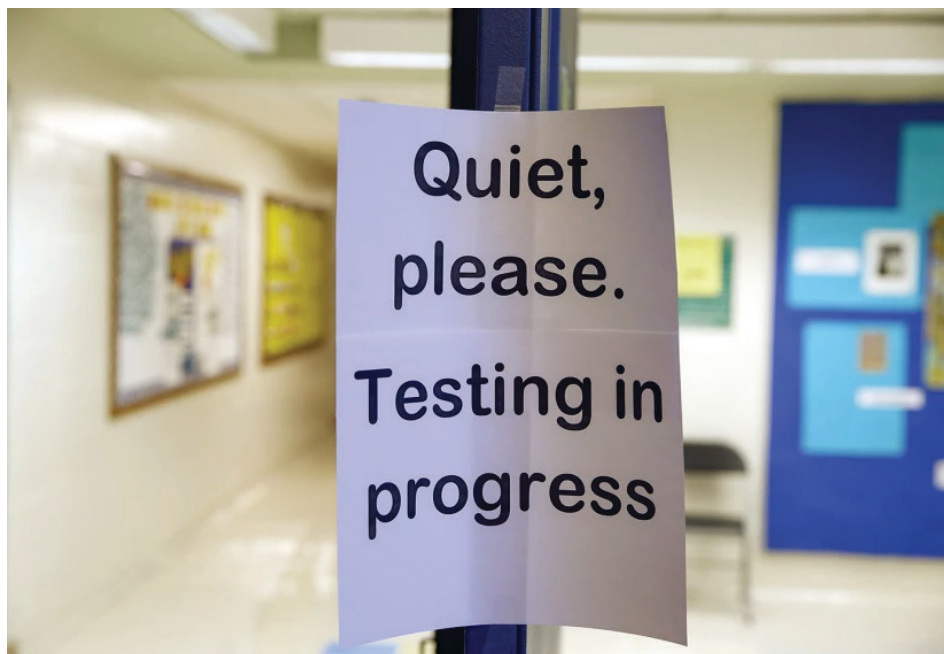
A number of factors fueled low participation rates, including many parents who chose not to send their children into school buildings simply to take a test. And schools likely felt less pressure to insist that students show up for testing, since the Education Department waived its accountability rules that normally penalize schools for testing fewer than 95 percent of their students.

"There was a wide variety in the ways testing played out," said Terra Wallin, who advised the Education Department on assessment and accountability from 2014 to 2017 and now oversees those issues for Education Trust, a civil rights advocacy group. "There are still ways states could look at general patterns [in test-score data], do a higher-level examination, to help them think about how best to use federal funding for recovery, but they need to proceed with caution."

### Ask key questions before deciding how to use the data

Experts say it's important to ask three crucial questions about your state test data.

- **Did any of our students take the test remotely?** If so, those scores shouldn't be viewed as comparable to the scores of students who took it in person. That "mode effect" is a key tenet of assessment: Whether a student takes a test online or with paper and pencil can influence the results.
- **Did we use the same test as in 2019?** If you switched tests, or changed the length or frequency of your test, a detailed expert analysis could be needed to confirm the validity of the 2021



A sign is posted at the entrance to a hall for a test-preparation class. Assessment experts say educators should use data from spring 2021 tests with caution.

—Alex Brandon/AP

results—were there enough questions in each strand of the academic standards, for instance, to generate a valid score?—and to establish that those results can be compared with 2019 results.

- **How many of our students—and which ones—took the test?** This "participation rate," experts say, is very important in understanding what state tests say—or can't say—about student learning. They urge educators to dig deeper than the overall state or district participation rate and find out who took the test and who didn't.

Imagine that an analysis shows that the students who skipped the test were disproportionately those who scored low in previous years. That would skew test results artificially high, and stalled progress might appear less severe than it actually is.

That isn't just speculation, either. It's likely that remote learners account for many missed tests and it became increasingly apparent during the pandemic that low-income, Black, and Latino students were far likelier to be learning remotely than other students. And emerging multistate research on state test results is finding that COVID's impact on learning isn't concentrated just in elementary schools, or among traditionally low-performing students, as early analyses of interim tests suggested; it's broader, affecting students at

all grades and achievement levels.

Enrollment declines, widely documented in many grades, can also play havoc with sound interpretations of test scores. Again, it's important to understand the academic and demographic profiles of who stopped coming to school, experts say.

"If you aren't paying attention to how the population is changing, you're misinterpreting your scores," said Andrew Ho, a Harvard University professor of education who focuses on assessment. He urges state leaders to perform a three-dimensional analysis of their test scores to ensure valid comparisons. This is done by separately comparing each group—the students who took tests and those who didn't—only to groups who performed similarly in the past.

"We've just got to avoid a naïve analysis" of 2021 test-score data, said Derek Briggs, a University of Colorado professor who leads the National Council on Measurement in Education, whose members design and study K-12 assessments.

"The danger here is that we report 2021 scores as observed in 2021, without doing any other analysis. People want to compare them to 2019, and they're going to interpret the difference as the effect of COVID." But the pool of students who took the tests in 2021 changed, and that requires deeper analysis than in other years, he said.

Briggs is worried that districts and states won't take the shifting test pool into account,

## EXCLUSIVE INTERVIEW

# Dr. Roland Good on Assessment and the Equity Gap

Since the pandemic began, a widening equity gap has been identified in education and it has had devastating effects on our schools, particularly on minority students in low-income communities. The lockdown and ensuing switch to remote learning limited access to teachers and made education contingent upon technology and the internet.

The **equity gap** refers to disparities in educational outcomes and student success metrics across race, ethnicity, socioeconomic status, gender, physical or mental abilities, and other demographic groups.

Other factors contributing to the gap—as identified by Dr. Roland Good in a recent interview on the EDVIEW360 podcast—included the time, coordination, and self-motivation required to connect with teachers virtually. These are all things vulnerable students were likely to struggle with and, ultimately, caused them to fall behind.

“**Many students learn to read or learn mathematics almost as if we don’t have to teach them. Those students are less likely to be affected by the pandemic. But many students learn *because we teach them*. Those are the students we need to impact the most.**”

—Dr. Roland Good

“For those students, we’re going to have to switch from preventing a problem to remediating a problem,” says Dr. Good. “The earlier we can intervene, the earlier we can improve teaching and instruction—and the earlier we can get students on track.”

Dr. Good urges educators to use tools like Acadience® Learning to assess each student’s skills to understand where they stand and how instruction must be modified to meet their needs.

“We have brief assessments that can tell us where a student is on a skill and where they need to get to in order to make adequate progress,” he says. “We can track their progress toward that goal on a week-by-week basis, and we can make adjustments to instruction depending upon whether or not the student is learning and having success.” ■

To better understand how assessment can help eliminate the equity gap in your district or school, catch the entire interview with Dr. Good for EDVIEW360.



Listen Now



and they'll take reassurance from a falsely rosy picture. That's a particular danger in any state or district where fewer than 90 percent of students took the test, he said. Smaller margins of missing students means less of a chance those missing scores affect overall results.

Participation rates below 50 percent would make it tough to draw any meaningful conclusions from test results, said Marianne Perie, the president of Measurement in Practice, which advises states on test design and use.

Sean Reardon, who leads a Stanford University project that analyzes the links between test scores and children's learning opportunities, said the insight into learning offered by spring 2021 test scores is very limited because of all the factors influencing the scores.

"If you had a random sample of kids [in the testing pool], then that would be fine," he said. "But testing in 2021 wasn't random. Kids and families chose whether they took the test. Unless you have a lot of information to support a claim of comparability, I think the default assumption for 2021 is that they're not comparable [to 2019 test scores]. I wouldn't draw too many conclusions based on them and I'd use a lot of caveats."

### Consider ways to get insight into motivation and learning conditions

Ellen Forte, the chief executive officer and chief scientist at edCount, which advises states and districts on testing, said educators should bear in mind that millions of students, anxiety-riddled during COVID-19, were likely less motivated to do well on tests. Given that distortion, and the fact that state tests are not designed to yield highly detailed pictures of students' achievement, she wouldn't want to see students' test scores used to make instructional decisions.

"Remember, these tests were designed for accountability," Forte said. "The unit of focus should be the school, district, or state. Not the student."

It also would behoove educators to understand more about the conditions in which students were learning, said Scott Marion, the executive director of the Center for Assessment, a consultant to states on testing. The organization has helped several states create student surveys that asked about things like their access to live-streamed instruction and how much they'd

learned compared with the previous year. Teachers were asked, among other things, whether they'd been adequately supported with good professional development during the pandemic.

In a year like 2021, "I think it's important," Marion said. If a child tested in 2021 under conditions similar to 2019, educators can probably make sound—and very general—inferences about whether she gained or lost ground in those two years, Marion said. But what's missing is the "why." Gathering other data, from surveys, teacher observations, formative strategies, and interim assessments embedded in good curriculum, can shed light on "why my kids did poorly and what I might need to do differently," he said.

### Takeaway message: Multiple sources of data are more important than ever

Most experts consulted for this story agreed that with the right kinds of analyses, states can probably glean valuable information about patterns of low achievement so they can provide appropriate supports. They urged districts to press their states for detailed information and analysis to guide similar decisions at the district level.

In the classroom, though, experts differed on the role state test data should play in guiding instructional decisions for groups or individual students. Perie of Measurement in Practice said she wouldn't want to see scores used for high-stakes decisions like grade promotion but thinks they could help teachers create flexible groupings in math or reading or dive more deeply into strands where class scores seemed weak.

Even better, Perie and other experts said, would be to blend test-score information with a portfolio of other data from formative or diagnostic tests, reports from students' previous teachers, and other sources. "You've got to triangulate, leveraging other measures like you never have before," Harvard's Ho said.

Superintendents understand this, said Dan Domenech, the executive director of AASA, the School Superintendents Association. They know it's "critical to ascertain how much loss has taken place so they know where to begin," but they recognize that standardized tests, while valuable, provide only "a general overview." Accordingly, teachers will rely heavily on quizzes and other formative strategies to understand what their students need, he said. ■



—Visual Generation/iStock/Getty

## OPINION

Published May 5, 2021

# What Federally Mandated State Tests Are Good For (And What They Aren't)

By Stuart Kahl

Federally mandated state testing, which did not take place for the 2020-21 schools year because of the pandemic, is happening. However, with the waivers states have been allowed, involving the identification of low-performing schools, the shortening of tests, and changes to the percentage of students to be tested, this marks the second year that continuity under the Every Student Succeeds Act has been broken. For that reason, this is an opportune time to start shifting the focus of state testing back toward a purpose for which the tests are more appropriately designed.

In recent years, school administrators and teachers have put pressure on state policymakers and testing officials to address the desires of local educators for more immediate information from state tests to inform their day-to-day instruction while still satisfying the federal accountability requirements. Guiding day-to-day instruction is not something states' summative-test results can or should be expected to do, despite the ESSA's unmet requirements for quick, individual student results that are "interpretive, descriptive, and diagnostic."

Many people have questioned the need for spring 2021 testing, viewing the main purpose of state assessments to be school account-

ability for student academic performance for which states determine percentages of proficient students. This limited view overlooks the power of end-of-year state assessments for program evaluation and improvement.

Standardization in testing means comparability—a quality of state tests that enables local educators and residents in general to get an external perspective on how their local instructional programs are doing. How does our student performance compare with that of other schools serving similar populations?

State test results should raise questions that need further investigation to answer. Why are we underperforming in this subdomain of math? Why is this subgroup in our school underperforming their counterparts in other schools in our state? Are our new approaches in this area working? Finding the answers to such questions informs program-improvement efforts that may not immediately benefit the students tested but instead should benefit many more students in the future.

There is strong feeling among educators and noneducators alike that teachers and schools should not be penalized because of the impact of COVID-19 on student learning. I agree. But the need for program-evaluation information is greater than ever right now. How well have we served our students via online instruction, packets sent home, or any other method compared with other schools? What is the extent of learning loss during the pandemic?

Clearly, we should be prepared to see results reflect a negative impact from COVID-19. Although the stakes have been relaxed, state testing officials would do well to maintain a focus on program evaluation not only now but also in the future after the stakes are reimplemented.

State tests are designed to produce reliable total test scores for students in mathematics, English/language arts, science, and other subjects states may test. But the tests generally do not yield reliable subtest scores—e.g., geometry or measurement in mathematics or physical or life science within science. Typically, subtest scores are based on just a handful of machine-scorable items hardly representing a good sampling of the content and skills in a subtest area. These areas are still very broad, so the sparse coverage of relevant content and skills makes subtest scores neither reliable nor valid for making important instructional decisions for individual students. For a test to be truly diagnostic, it would require multiple items addressing narrowly defined learning targets.

The U.S. Department of Education could offer guidance on ESSA relaxing the unmet requirement for states to provide diagnostic

“**Standardization in testing means comparability—a quality of state tests that enables local educators and residents in general to get an external perspective on how their local instructional programs are doing.**

reports for individual students, while still requiring the reporting of total test scores for all students. Diagnoses of students’ specific learning gaps are best left to other components of a balanced assessment system. School average subtest scores are more reliable than individual student subtest scores even with the poor sampling of content. However, that poor sampling means important programmatic decisions for school programs are still questionable—and this is a validity issue. Local educators observing patterns of scores over a few years would be a better basis for major changes to programs.

There’s a place for end-of-year summative testing in both program evaluation and accountability assessment. For these purposes, I’d want to know how students in a school perform near the end of a school year. ESSA now allows accountability assessment to take place during the course of the school year. Interim benchmark tests covering recently taught material and interim general achievement measures are appropriate for purposes of early warning to identify students and curricular areas needing additional attention before end-of-year testing. However, many of these tests have the same limitations as the state summative tests because of the sparse coverage of relevant content and skills. And they don’t reflect student achievement at the end of the school year.

There are testing-program designs that can shorten the testing time for individual students yet broaden the coverage of a subject-area domain for school-level results. Reliable total test scores could still be provided for individual students, while reliable school results in subtest areas and even finer breakdowns of content and skills could be reported, offering more valuable information for program evaluation. Innovative curriculum-embedded performance as-

sessments a few times during the school year could tap higher-order cognitive skills that the efficient end-of-year tests shortchange and could yield immediate results from teacher scoring of student work on complex tasks or projects. These, too, could count toward federally required accountability results.

Challenges create opportunities. The pause in 2020 testing and waivers—because of the pandemic—make this a good time to renew our emphasis on state assessment as a powerful activity for informing program improvement. We should not think of it solely as a vehicle for producing accountability data or a source of information for real-time instructional use—something it cannot be. Since the collection of longitudinal student-achievement data for ESSA has to be restarted anyway, changes to assessment designs can be made in the 2021-22 school year or next that optimize their utility for program evaluation. Such changes would be a way to get state assessment back on track with respect to what it can legitimately do, not just for the short term but also for the foreseeable future. ■

*Stuart Kahl is an independent assessment consultant with 35 years of firsthand experience in designing, developing, and implementing state assessments and is the former CEO and founder of Measured Progress Inc., a company that operated state testing programs in more than half the states over several decades. He previously taught at the elementary, secondary, and graduate levels.*

Copyright ©2022 by Editorial Projects in Education, Inc. All rights reserved. No part of this publication shall be reproduced, stored in a retrieval system, or transmitted by any means, electronic or otherwise, without the written permission of the copyright holder.

Readers may make up to 5 print copies of this publication at no cost for personal, non-commercial use, provided that each includes a full citation of the source.

For additional print or electronic copies of a Spotlight or to buy in bulk, visit [www.edweek.org/info/about/reprints.html](http://www.edweek.org/info/about/reprints.html)

Published by Editorial Projects in Education, Inc.  
6935 Arlington Road, Suite 100  
Bethesda, MD, 20814  
Phone: (301) 280-3100  
[www.edweek.org](http://www.edweek.org)

EducationWeek®

# SPOTLIGHT

**Get the information and perspective you need on the education issues you care about most with Education Week Spotlights**

The Achievement Gap • Algebra • Assessment • Autism • Bullying • Charter School Leadership • Classroom Management • Common Standards • **Data-Driven Decisionmaking** • Differentiated Instruction • Dropout Prevention • E-Learning • ELL Assessment and Teaching • ELLs in the Classroom • Flu and Schools • Getting The Most From Your IT Budget • Gifted Education • Homework • **Implementing Common Standards** • Inclusion and Assistive Technology • Math Instruction • Middle and High School Literacy • Motivation • No Child Left Behind • Pay for Performance • **Principals** • Parental Involvement • Race to the Top • Reading Instruction • Reinventing Professional Development • Response to Intervention • School Uniforms and Dress Codes • Special Education • STEM in Schools • **Teacher Evaluation** • Teacher Tips for the New Year • Technology in the Classroom • Tips for New Teachers



**VIEW THE COMPLETE COLLECTION OF EDUCATION WEEK SPOTLIGHTS**

[www.edweek.org/go/spotlights](http://www.edweek.org/go/spotlights)