EDITOR’S NOTE
Effective assessments can illuminate strengths, pinpoint learning gaps, and better guide education. This Spotlight will help you evaluate effective ways to offer students feedback; examine how some states are transitioning to through-year testing models; hear from educators regarding pressure for student success on standardized tests; gain insights into data on student test results; review expert advice for improving assessments for English learners; and more.
Few Educators Say A-F and Numeric Grades Offer ‘Very Effective’ Feedback for Students

By Alyson Klein

Traditional grading systems aren’t getting an A-plus from most educators.

In fact, fewer than 1 in 6 educators—13 percent—surveyed by the EdWeek Research Center earlier this year say that A through F or numeric grades are a “very effective way” to give feedback to students.

Still, educators clearly see the system as having some merit. Nearly half—42 percent—of the teachers, principals, and district leaders find A through F or numeric grades at least “somewhat effective.” About a quarter find them “somewhat ineffective,” while more than 1 in 5—21 percent—find them “very ineffective.”

Despite those mixed reviews, traditional systems appear to remain the norm across the country, the survey found. More than three-quarters—77 percent—of educators surveyed said that their districts use either the A through F grading system, a numerical grading system, or a combination of the two. Just 11 percent of educators said their districts use another type of system. The EdWeek Research Center nationally representative survey of 863 educators was conducted from March 29 to April 11.

Still, some teachers worry that these systems aren’t nearly nuanced enough to capture student progress, and that students with a succession of low grades may get discouraged and give up trying to master the material.

“The traditional 0-100, A-F grading system does not communicate learning,” wrote Jonathan Medeiros, a language arts teacher at Kaua‘i High School in Hawaii, in an email. “It communicates behavior, privilege, and positional. Worse than that, the A-F scale promotes giving up and cheating to get the grade, moving students away from the desire to grow and learn.”

Micah Miner, the district administrator for instructional technology and social studies for the Maywood, Melrose Park, Broadview school district outside Chicago, agreed that the A through F system doesn’t give students and their parents enough of a picture of which material they’ve mastered and what they are still struggling with.

Over the past four years, Miner’s district has moved to a standards-based system, where students are given a 1 through 5 rating to show how close they’ve come to grasping a particular skill or concept.

For instance, a social studies teacher might tell a student that they have a good understanding of how the American Revolution influenced civics by giving them a 5 on that standard. But the student may still struggle with some of the historical specifics, meaning they would only get a 3 on the history standard for that unit, Miner explained.

That kind of system can give students a more detailed understanding of their progress, strengths, and areas for improvement, Miner said. However, he acknowledged that teachers may struggle to make the transition to this type of grading.

“It’s a harder thing to do,” said Miner, who helped his district move to the standards-based system. “It’s more objective, and you’re giving more specific feedback and that requires more time.”

Parents may have an even more challenging time making the switch, others argued.

Zack Kleypas, the superintendent of Texas’ Thorndale school district and a former principal, said that while A through F and traditional numeric systems aren’t perfect, they offer an easy-to-grasp metric for families.

“It easily communicates to most parents whether a kid’s doing really good, kind of good, not so good, or bad,” Kleypas said. “For that reason, I like it, because if you change that paradigm too much, then you have to spend a lot of your energy training parents to comprehend a completely different system.”

The system though, could be tweaked to give students and families more information about how students performed on each of the standards that make up that overall letter grade, Kleypas said.

In your view, how effective are traditional A-F or number grades at providing useful feedback to students?

Very ineffective: 13%
Somewhat ineffective: 21%
Somewhat effective: 42%
Very effective: 23%
Frstrated with an accountability system that revolves around once-a-year standardized exams, more states are looking to redesign the tests, hoping they provide teachers and districts with more timely and useful feedback on their students’ progress.

Last month, the Montana education department secured a rare waiver from the U.S. Department of Education, allowing the state to forgo its usual standardized tests used for federal accountability purposes and instead field test a new system of exams given at various points throughout the 2023-24 school year. In Missouri, the state education department has awarded “innovation waivers” to a group of 20 school districts so they can pilot through-year programs in addition to the state standardized tests, and the state is hoping to secure a federal waiver like the one Montana received to start offering the through-year tests in lieu of the current system. And in Florida, schools have switched from an end-of-year test to a “progress monitoring” model that involves testing three times a year.

The U.S. Education Department’s decision to approve Montana’s request for the Field Test Flexibility waiver on Aug. 10 is an extremely rare instance of the department allowing a state to forego federal testing requirements without a natural or public health disaster, like COVID-19, driving that decision.

“The Department is committed to partnering with States to develop more innovative approaches to assessments and supporting States in their provision of timely academic achievement and student progress data to educators, parents, and families,” a department spokesperson said in an email to Education Week.

At least 13 states have started exploring through-year test models, according to Education First, an education policy organization. The slow movement reflects educators’ opinions that state-mandated tests aren’t useful for teaching. But a wholesale transition away from the testing regime would require a federal law change dispensing with the two-decade-old accountability system introduced under the No Child Left Behind Act. No such change, however, appears to be in the offing.

The old accountability model was built on outdated assumptions that the primary audience for test scores should be policymakers, not teachers, students, or parents, said Mike Fulton, lead facilitator of the Success-Ready Student Network, the group in Missouri leading the work to change testing.

“That is a model centered on the notion if we just weigh the schools and report out on how they’re doing, improvement will occur,” Fulton said. “But the assessment actually has nothing to do with informing continuous improvement in real time, making adjustments in real time, informing instruction, informing school design on how time and structures ... are used to support learning.”

In a recent EdWeek Research Center survey, nearly 80 percent of educators said they feel moderate or large amounts of pressure to have students perform well on state exams, with 49 percent of educators saying they feel more pressure now than they did prior to the COVID-19 pandemic.

But most educators don’t find end-of-year assessments useful.

An outdated system

Tony Lake, superintendent of the Lindbergh school district in St. Louis, compares the current end-of-year state testing model to being a quarterback playing in the Super Bowl in the 1970s.

“If you watched the Super Bowl last year, you saw [Kansas City Chiefs quarterback] Patrick Mahomes sitting on the sidelines with his tablet getting real-time information to help him win that Super Bowl or win that game, and they won that Super Bowl,” Lake said. “Kansas City won Super Bowl IV [in 1970] when Len Dawson was the quarterback. Len Dawson got his information from the game about four days later.

“When you think of the current assessment, Len Dawson is our current state assessment system.”

District leaders like Lake would prefer to live in a world where testing gives educators real-time feedback on student progress so they can ensure they’re meeting student needs throughout the school year. Lindbergh is one of the 20 districts in Missouri piloting new testing models.

The district is using NWEA tests administered multiple times throughout the year for students in 3rd through 9th grade in addition to Missouri’s year-end standardized tests. (NWEA creates the series of MAP tests designed to be given throughout the year to measure student progress.) The idea is to give
students, teachers, and families the full information on how the student is progressing throughout the year.

In the Show Me State, it’s individual school districts driving the change. But it’s the same thought process as in Montana, where State Superintendent Elsie Arntzen and her team developed the Montana Alternative Student Testing Pilot Program, which the state has used in a small set of schools since September 2022. It is a through-year test administered five times a year to all students in grades 3-8.

Before becoming superintendent, Arntzen spent some of her career as an elementary school teacher and became aware of the current testing model’s pitfalls.

“That test score just didn’t recognize my work as a teacher or the work that [my students] did,” Arntzen said.

The federal education department’s Field Test Flexibility waiver, issued Aug. 10, allows Montana to expand the pilot, and removes the usual year-end testing so teachers don’t have to double-up exams for their students.

The new testing model is “going to bring parents and students and teachers all together on the same page,” Arntzen said. “It’s going to recognize that the accountability of the classroom is spread between family, the student, and the teacher.”

**The issue of accountability**

One of the major criticisms of state-mandated standardized tests is that they aren’t an accurate measurement of school performance. Standardized tests can favor districts with more resources and leave districts with high rates of poverty struggling to catch up.

Many educators would like to see states and the federal government put more emphasis on other measures to assess school performance, like teacher turnover and attrition rates and school climate surveys. As for the tests educators find useful, a majority of respondents to the recent EdWeek Research Center survey listed teacher-created formative assessments and teacher-created end-of-unit exams.

But for accountability purposes, through-year testing may give policymakers a more accurate view of how students are doing and provide more transparency to families, Fulton said.

Fulton, Arntzen, and Lake would all like to see state and federal policymakers provide more opportunities for flexibility from current accountability requirements, like the waiver from the U.S. Education Department.

There are options built into federal law to allow states and individual districts the flexibility to overhaul their testing systems or opt for different tests. But there haven’t been many takers. Only a handful of states ever showed an interest in an Innovative Assessment pilot, for example, that federal lawmakers hoped might pave the way for a new generation of tests.

In an email, an Education Department spokesperson said the agency is working to strengthen the Innovative Assessment pilot, while also providing states with grant funding for more assessment innovation.

Teachers, meanwhile, have often argued they’re overburdened with testing, and that test prep takes away from time that could be spent on student learning: Thirty-six percent of educators in a recent EdWeek Research Center survey said they or teachers in their district spend pretty much the entire school year on test prep. Some teachers’ unions have argued that through-year testing can add to that burden, especially as schools are still required to do year-end testing.

Those pushing a real-time testing model argue that it’s one ingredient in a system that better promotes learning.

“Replace heavy-handed accountability with an absolute laser-like focus on learning,” Fulton said. “Then support student learning with good instructional practices and policy that encourages people to take risks based on research.”
The time is now to assess essential academic skills

Realize the full potential of every student, every teacher, and every school...with the powerful and effective Acadience® suite of research-based, preK–8 assessment tools.

Including:

- Acadience® Reading Pre-K: PELI®
- Acadience® Reading K–6
- Acadience® Reading 7–8
- Acadience® Math
- Acadience® Learning Online

With the Acadience suite of assessment solutions, educators can evaluate the effectiveness of instructional strategies, gain a deeper understanding of their students’ skills, and implement targeted intervention to guide them on the path to success.

Learn more at voyagersopris.com/acadience
“t was the best of times, it was the worst of times,”
—Charles Dickens

(A Tale of Two Cities, 1859)

These century-and-half-old words describe what it is like being an educator today.

On one side, educators say, “It is the best of times” because we know more now than ever about how students learn to read...brain imaging, consensus research, sharing on social media, increased funding for education, technology... on and on.

However, on the other hand, many educators today say, “It is the worst of times!”

Educators are facing unprecedented challenges:

• Growing workloads and less time to provide students with the individual attention educators know they need
• Politicians who’ve never studied the science of learning trying to tell educators what they can and should teach
• The push to deal with learning loss
• Required summer school and tutoring
• Divergent needs including ALL students
• Conflicting messages and approaches from ever-present social media

And despite all the demands, U.S. Secretary of Education Dr. Miguel Cardona recognizes the challenge and recently commented, “And through it all, you’re focused on what matters most—your students.”

No truer words!

How can educators sort through the noise, ease their angst, and begin to feel more confident about the decisions they make each day in planning instruction for their students?

How do educators make the best use of assessment resources and results to plan effective evidence-based instruction for all students?

There are two proven practices educators can feel confident about utilizing in their classrooms to ensure positive outcomes and literacy success for all students:

1. Structured Literacy—The Science of Reading
2. MTSS (Multi-tiered Systems of Support)

Structured Literacy—The Science of Reading

Decades of reading intervention research provides compelling evidence that teaching phoneme awareness, phonics, orthography, morphology, syntax, and semantics in an integrated, woven, simultaneous manner through explicit systematic, cumulative interactive practice with opportunity to read connected text results in improving reading outcomes for all students struggling to learn to read, including those with dyslexia (Al Otaiba, Rouse & Baker, 2018; Torgesen et al., 2001; Vellutino et al., 1996; Vellutino, Scanlon, & Lyon, 2000). This body of evidence makes up Structured Literacy and provides a solid foundation for reading instruction known as the science of reading.
MTSS (Multi-tiered Systems of Support)

Multi-tiered Systems of Support (MTSS) is the framework for implementing the Structured Literacy practices in the science of reading. MTSS includes three tiers of instruction in which all students are given the type and amount of instructional support they need to be skilled readers.

Implementation of MTSS requires a comprehensive system of assessments to address four purposes: screening, diagnostic, progress monitoring, and outcome evaluation. The goal of assessment is to guide instruction and intervention...matching intervention to assessment data.

In the revision of the book I coauthored, Next STEPS in Literacy Instruction: Connecting Assessments to Effective Interventions, these practices are highlighted, and we include the Oral Reading Fluency Decision Tree, which you can [download when you visit](https://www.edview360.com/podcast) the EDVIEW360 podcast, on which I was a guest. This figure gives educators an easy-to-use framework with the steps/flow chart for analyzing student data to plan effective instruction.

For example, when educators use the Oral Reading Fluency Decision Tree (at right), they begin to ask questions as they review student data:

**Step 1**
- Is the student’s score at benchmark level with accuracy 95 percent or above?
- If the student’s accuracy score is below 95 percent, or if the score is below benchmark, it’s time to dig deeper.

Answers to these two initial questions lead to planning purposeful effective intervention.

**Step 2**
- Dig deeper: Are there specific diagnostic assessments to help answer questions about how and why the problem is happening?

**a. Assess phonics:**
- Example: Acadience Reading K–6 Nonsense Word Frequency

**b. Assess phoneme awareness:**
- Example: Acadience Reading K–6 Phoneme Segmentation Fluency provides insight into phoneme awareness development. Look for patterns (e.g., no segmentation, initial sound only, inconsistent, final sound)
- Example: The Quick Reveal Phoneme Awareness Tool (Next STEPS-Revised). Look for patterns (isolate first sound, final sound, all sounds, blending all sounds?)

**c. Assess High-Frequency Word knowledge:**
- Screening High-Frequency Word List Top 248 (Read and Spell) (Next STEPS in Literacy Instruction Connecting Assessments to Effective Interventions, Smartt & Glaser, Brookes Publishing, 2024)
- CORE Graded High Frequency Word Survey (Assessing Reading Multiple Measures-Diamond & Thorsnes, 2018)

I hope you’ll tune into Voyager Sopris Learning’s EDVIEW360 podcast, Determining the Right Literacy Intervention: Using Assessment to Guide Your Course, where I discussed this in detail.

Explore Acadience assessment solutions at voyagersopris.com/acadience
Educators Feel Growing Pressure for Students to Perform Well on Standardized Tests

By Libby Stanford

A majority of educators find that state-mandated standardized tests aren’t useful in the classroom despite feeling a large amount of pressure to have their students perform well on those exams, according to new data from the EdWeek Research Center.

Just 25 percent of educators said state-mandated tests provide useful information for the teachers in their school in an online survey of 870 teachers, principals, and district leaders administered from July 26 through Aug. 20.

But nearly half of educators, 49 percent, said they feel more pressure now than before the COVID-19 pandemic to ensure students perform well on state tests. Forty-two percent of educators said the amount of pressure has remained about the same since 2019, while 9 percent said it has decreased.

The data come at a time when state and federal leaders are rethinking the value of end-of-year standardized tests, which critics argue only capture a snapshot of student achievement at a specific moment in time. In a speech to educators in January, U.S. Secretary of Education Miguel Cardona said standardized tests should be “a flashlight” on what works in education rather than “a hammer” to drive outcomes.

Cardona’s sentiment is shared by many educators, but it goes against more than two decades of federal education policy that has required annual testing and set out sanctions for lower-performing schools.

Nearly 80 percent of educators said they feel moderate or large amounts of pressure to have their students perform well on standardized tests. In the survey, educators said they’d like to see graduation rates, school climate surveys, and rates of teacher turnover and attrition used to measure school success over standardized test performance.

Standardized tests are “overwhelming,” said Olga Neurauter, principal of Longfellow Elementary School in Raton, New Mexico. “There’s too much pressure put on these kids for testing, and there’s too much testing. We need to be able to teach the curriculum, and sometimes when we do those state-mandated tests, it takes away” from learning.

In the survey, 41 percent of respondents said the amount of time they and teachers in their districts spend preparing students for standardized tests has grown since 2018-19, the last full school year before the pandemic. Forty-nine percent said they spend about the same amount of time, and 10 percent said they spend less.

Educators find year-round testing more useful

Only a quarter of educators who responded to the survey said they found state-mandated standardized tests useful. At the same time, 74 percent said they find in-class, teacher-created, formative assessments to be useful, 59 percent said unit tests and final exams are useful, and 50 percent said diagnostic exams created by teachers at the start of the year are useful.

In conversations with Education Week, teachers said they find diagnostic exams like I-Ready, which is administered multiple times throughout the year, to be the most useful for guiding classroom instruction, compared with state-mandated standardized tests.

“There’s too much pressure put on these kids for testing, and there’s too much testing. We need to be able to teach the curriculum, and sometimes when we do those state-mandated tests, it takes away.”

OLGA NEURAUTER
Elementary School Principal
Raton, NM
Assessment

How useful to YOU are the results of the state-mandated standardized tests taken by your students?

![Chart showing results from teachers, principals, and district leaders.]

Results show responses from teachers, principals, and district leaders
SOURCE: EdWeek Research Center survey, August 2023

[exam] when I group my students to differentiate instruction,” said Nino Cuzco, a middle school Spanish teacher in New York. “For example, if I have a student who scores low on the ELA I-Ready, I know they will need more support in my classroom because when I teach Spanish there may be vocabulary or grammatical concepts they may be lacking in background knowledge.”

The I-Ready exam has also been helpful in determining when students need interventions to help them get on track, Cuzco said. Cuzco said she’s also found New York’s state-mandated tests, which are administered to students in 3rd through 8th grade every spring, to be useful in determining students’ skill levels. But the I-Ready tests give a better sense of student growth throughout the year, she said.

Standardized tests often just feel like a snapshot with little useful information for educators, said Nanette Murray, an English language development teacher in Tucson, Ariz.

Teachers already have a good understanding of where their students are academically through their work in the classroom, Murray said.

“We’re spending a gazillion dollars for a tiny piece of information,” Murray said.

A push for less emphasis on standardized tests in school accountability

Graduation rates, school climate surveys, and teacher turnover rates were all more popular ways to measure school performance among educators than standardized tests, according to the survey.

Cuzco said that she finds standardized tests useful in providing a sense of students’ academic performance, but she doesn’t think they should drive school accountability, especially as students are struggling to recover from the COVID-19 pandemic.

“My opinion may have been different 10 years ago or so,” Cuzco said. “Now I know there is generational trauma and every child is going to jump back from it in a different way or need different support, but I am seeing learning gaps. The tests do show that.”

Murray would like to see lawmakers find a way to measure schools based on parent engagement, including levels of parent outreach and parent support. Parents play a large role in helping students succeed, and schools should work to facilitate that, she said.

“Schools need to be held accountable for meeting the kids and the parents where they are and giving them the supports they need to succeed,” Murray said.

Additional Resource
View this article’s charts
Teachers Say Those Ubiquitous, Every-Few-Months Tests Don’t Always Capture What’s Taught

By Sarah Schwartz

At some point during the year, most students in U.S. schools will take a benchmark test. These tests, also called interim assessments, are designed to measure students’ progress toward instructional goals and are given every few months. Districts can purchase off-the-shelf versions—MAP, i-Ready, and Star are common tools, for example—or create their own tests.

But even though the vast majority of schools use these assessments, teachers and principals have different perceptions of their effectiveness.

More than 80 percent of principals say that, for the most part, the tests they use align to state standards and the end-of-year state-administered tests used for school accountability.

But about a third of teachers say that these interim tests don’t align with the curriculum and don’t accurately measure what students have learned. And teachers who report this mismatch are more likely to say that it’s difficult for them to figure out how the tests should inform their instruction.

These are the findings in a report this week from the RAND Corporation, which analyzed nationally representative survey data from about 1,500 principals and over 6,000 math and English/language arts teachers.

“We know that many education leaders and school systems are working really hard to help kids recover from the effects of the pandemic,” said Ashley Woo, an assistant policy researcher at RAND and the lead author on the paper. “Part of that work entails knowing where students are and how to adjust learning to meet students’ needs.”

The findings underscore an important feature of interim tests: Different assessments are designed to serve different purposes. While some are meant to measure how well students will do on a state summative test, others claim to show whether students have mastered skills and knowledge taught in the curriculum.

Understanding each test’s intended purpose is crucial to using the data it produces well, assessment experts have stressed. That’s challenging because the commercial market-place for these tests is a bit of a black box: Not much is known about the technical properties or assumptions that undergird most of the popular off-the-shelf exams.

Using multiple types of tests was common, according to the principal survey. On average, schools administered three different benchmark assessments in both ELA and math—though the report notes that not all tests would necessarily be given to all students.

This could be because different types of tests are better suited to different purposes, Woo said.

“For instance, you might have something like an i-Ready or an NWEA MAP test, which is more standardized and allows you to compare student progress across different schools and districts because lots of different schools and districts are using these assessments,” she said.

Others might measure whether students have learned the skills and knowledge teachers cover in class—think unit tests, term papers, or other major assignments. “Those kinds of assessments that are specifically designed to align with curriculum are the kinds of assessments that are giving teachers more real-time data,” Woo said.

The report also tracked test use over time, including during the pandemic. The researchers found small but steady growth in schools’ use of commercial assessments between the 2018-19 and 2021-22 school years.

Use of the locally created assessments fluctuated, declining during the 2020-21 school year, and then jumped back up during...
2021-22. Some of this change could stem from pandemic-related shifts in instruction and assessment, Woo said.

The overall trend in locally created test use during the period of the study is positive, though.

**Where principals’ and teachers’ perspectives on interim tests differ**

Principals generally thought that interim assessments could gauge whether students were meeting broad instructional goals. More than 80 percent of principals said that the interim tests they used aligned to state standards and end-of-year summative assessments.

The surveys asked teachers about a different metric: curriculum alignment. A majority, 64 percent, said that interim assessments did align with their curriculum. Still, though, about a third of teachers said they only partially aligned with the curriculum, or didn’t align at all.

These numbers varied slightly across different groups. For example, ELA teachers were less likely than math teachers to say that their assessments were aligned. Teachers who created their own interim assessments, and teachers who spent more time in professional development analyzing test data, were both more likely to report curriculum alignment.

In general, it seems somewhat difficult for teachers to figure out how interim test data should inform their instruction. Only 37 percent said that was easy to do; another 37 percent said it was difficult, and about a quarter said it was neither.

It was especially difficult for teachers who felt that the interim assessments weren’t well-aligned with their curriculum: Among those teachers, just 14 percent said it was easy to address student needs identified by their benchmark tests.

In interviews that the researchers conducted with 45 teachers, several mentioned unclear messaging from leadership about how to use the data that these tests provided.

“This is probably the biggest, most frustrating thing of our district,” said one elementary ELA teacher. “They have adopted so many different ways to benchmark our kids, different formats and different resources. ... And all of these different tests are all different standards, and they don’t align to our curriculum necessarily. ... Looking at the data, it’s sometimes hard for us as teachers to know which data [are] important, which do I need to use to show that my children are progressing.”

Guidance from district and school leaders is necessary, Woo said. Leaders should know whether benchmark tests align to standards, year-end state tests, curriculum, or some combination—and then use that information to convey messages to teachers about how to use them, she said.

“We think it’s important for state and local leaders to consider the slate of assessments ... and the role that they expect each assessment to fulfill,” said Woo.
Students have made some progress toward academic recovery, but overall achievement hasn’t yet reached pre-pandemic levels, new test results show.

The data are from the COVID-19 School Data Hub, a 2021 project launched by Emily Oster, a professor of economics at Brown University. The hub has recently aggregated 2023 student test-score results across grades 3-8 in math and reading for about half of all states.

The results show that most states have made up some ground in math, compared to the 2020-21 school year. But in reading, some states are making progress while others have regressed. Only a few states have recovered to pre-pandemic levels: Iowa and Mississippi in both math and ELA, and South Carolina and Tennessee in ELA.

“For us, the main takeaway has been the dramatically different recovery patterns between English/language arts and math,” said Clare Halloran, the associate director of the COVID-19 School Data Hub. “That was surprising.”

The group of states includes those that did not make changes to their assessments in 2021 or 2022, and have already released scores this year. The hub team plans to update the collection results as more states release their 2023 scores.

Testing experts say it’s hard to know exactly how to interpret these results.

“The overall story that there was a negative impact [of the pandemic] is not too controversial. Once you get past that, it’s much more controversial,” said Derek Briggs, a professor in the Research and Evaluation Methodology program at the University of Colorado Boulder.

Because different states design their tests differently, it can be difficult to compare student progress across locations. It’s also hard to know what exactly caused score jumps in certain states.

Education Week spoke with several testing experts about how to interpret the new state assessment data. Here are five considerations they raised.

Why ELA and math scores might be different

The data show that students are progressing more steadily in math than they are in English/language arts. In ELA, states have vastly different trends from one another. Testing experts say there are a few reasons this might be the case.

Math tests are “more sensitive to changes in instruction,” said Marianne Perie, the director of assessment research and innovation at WestEd. Previous research has shown that it’s easier to move math scores up with consistent teaching than it is to increase ELA scores.

Some curriculum experts have criticized reading tests for ultimately being a measure of how much students know rather than how well they read.

ELA tests also measure a combination of discrete skills—reading, writing, and speaking/listening, said Andrew Ho, a professor at the Harvard Graduate School of Education. States make different choices about how to balance and evaluate the combination of these skills in their tests.

So if students got less practice with writing during the pandemic, for example, that could result in big negative effects on some states’ tests—but smaller effects on others.

“Because different states measure in different ways, measurement experts have been a little wary of accurate measurement of progress in ELA through the pandemic,” Ho said.

Students in different grade levels could be recovering at different rates

The state test results that are reported in the COVID-19 School Data Hub are aggregated across all grades, 3-8.

“I think that can mask certain things,” said Benjamin Shear, an assistant professor in the Research and Evaluation Methodology program at the University of Colorado Boulder, referencing differences by grade level.

For example, in a few of the states Perie works with, 3rd grade ELA scores specifically are down. Most of these students were learning online in 1st grade, an important year for early reading foundations.

“Does that have more of an impact than being online in kindergarten?” Perie asked.

It’s possible, she said, that certain cohorts of students—who were online during key moments in their educational experience—might see more persistent score declines than others.

We don’t know exactly why some states are outliers

A few states stood out in the data, where students had reached or exceeded pre-pandemic achievement levels. This was the case in Iowa and Mississippi in both subjects, and in Tennessee and South Carolina in reading.

It’s hard to know exactly why some states are seeing these strong results, said Briggs.
The real story of the pandemic has been less about decline and recovery than it has been about inequality, inequality, inequality.”

ANDREW HO
Professor
Harvard Graduate School of Education

“Just because you see what looks like a trend, then attributing that trend to a change in policy [like a specific pandemic-recovery effort] is really, really hard to do,” he said.

Still, states have claimed that certain pandemic-era choices relate to their outcomes. In a press release about 2023 test scores, the Iowa Department of Education Director McKenzie Snow cited the state keeping school buildings open as a reason why students experienced a smaller decline in test scores during the pandemic than other states.

“There is some pretty strong national evidence that states who sent their kids back to school sooner had fewer issues,” said Perie. Even so, other states where many districts reopened for the 2020-21 school year—such as Arkansas—aren’t making similar academic progress.

Analyzing trends between student groups vary

Many other analyses of student test data—from the interim assessments that students take a few times a year to the federally administered National Assessment of Educational Progress—have shown the uneven toll the pandemic took on student achievement. Students from low-income backgrounds fared worse than their peers, and students who were already behind saw widening gaps.

This analysis also shows that in many states, the gap in student proficiency between high-income and low-income districts is increasing.

“The real story of the pandemic has been less about decline and recovery than it has been about inequality, inequality, inequality,” said Ho.

Still, he said, measuring gaps solely by using changes in the percent of students proficient could distort those comparisons. The reason why is technical, but important. In some situations, small changes in students’ scores could tip a lot of students over the edge from not proficient to proficient—or vice versa. It could look like a state was making big gains or losses, when in reality student scores only changed a small amount.

Analysts should consider other metrics as well, like average scale scores, that can provide a more precise understanding of gaps between student groups, Ho said.

Comparing anything to 2021 data comes with caveats

The state testing landscape in 2021 looked a lot different than in previous years.

In some states, such as Colorado, not every grade was tested in every subject. Across the country, student participation was much lower than usual.

“2021 has a lot of error around it because so many states didn’t get strong participation,” said Perie, noting that school enrollment rates changed during this time period, too.

All of these factors make it possible that the population of students who participated in testing between 2021 and 2023 changed in such a way that skews the comparison, said Shear.

Oster, Halloran, and their team took this variability in 2021 into account. For states in which 70 percent of students or fewer participated in testing in 2021—including Colorado, Delaware, Michigan, Oregon, and Nevada—they used 2022 results as their first post-pandemic measure.

Briggs said it could be useful to juxtapose state test results with other data, such as the state-level NAEP results that are released every two years.
Assessments in K-12 schools have traditionally been designed with a monolingual English-speaking student in mind. But the English learner population, speaking a variety of home languages, continues to grow across the country.

How effective, then, are traditional assessments in measuring what multilingual learners know in a given subject if they are limited to testing in English while they are still learning that language?

That topic was discussed in a recent webinar on multilingual learner engagement from the nonprofit Center for Applied Linguistics. Expert speakers offered advice on what it would take to rethink how English learners’ academic progress is measured at both the systemic level and classroom level, and what else can be done if assessments themselves can’t change overnight.

Rethinking what counts as assessments

When thinking about how to change assessments to best measure English learners’ progress, Micheline Chalhoub-Deville, a professor of education research methodology at the University of North Carolina at Greensboro, said much of what is considered best practice for these students benefits all students.

That includes not solely relying on one point in a school year when a test is given and ensuring that assessments are valid and reliable. Assessments should change to allow students to access and engage with academic content in whatever language can best demonstrate their content knowledge. In practical terms, that could mean adapting tests with simplified instructions or multilingual test practices, Chalhoub-Deville said.

“Rather than saying the ideal model here is a monolingual English language student in grade 5, the ideal should be somebody who can utilize whatever language is at their disposal in order to achieve in science, in math, in the arts,” Chalhoub-Deville said. “This is what we care about when it comes to educating our students.”

Such reimagining is currently underway in North Carolina, where the state education department received a grant from the U.S. Department of Education to look at how state assessments help level the playing field for students, said Ivanna Mann Thrower Anderson, a multilingual/Title III consultant for the state agency.

Specifically, the state wants to rethink science assessments after data analysis found gaps in achievement for English learners in this subject. Currently, science assessments are only given in English, Anderson said.

The state agency has now done focus groups with students, teachers, and families to get a better sense of what students are learning in science and what the assessments are not showing. In the next stage, the state plans to put together a group of researchers to develop tools, performance tasks, and rubrics teachers can use.

The work needed beyond changing assessments

Chalhoub-Deville shared how important it is for researchers like herself to work directly with policymakers after years of policy engagement being optional for researchers and scholars. Their work would more typically just filter through to policy, she said, but as accountability testing becomes even more critical for major decisions including education reform, researchers need to more proactively engage with policymakers.

Families of English learners should also play a bigger role in informing what should go into assessments for their multilingual students, Chalhoub-Deville said.

But changing statewide assessments doesn’t happen quickly even with strong collaborations between policymakers and researchers.

“If we can’t get these tests in their languages, we don’t have the capacity to do so, or we can’t get legislation to let us do that, then what work needs to be done with our classroom teachers to make sure that the instruction provided to them is making sure the content is accessible,” said Kia Johnson, director of Pre-K-12 language and literacy at the Center for Applied Linguistics.

Teachers need to think about other ways to measure progress in the classroom for English learners. They can use project portfolios, demonstrations, graphic organizers, promoting group work to learn from peers, and more, Johnson said, all of which can help students access academic content they will be tested on.
Acadience® Learning Online (ALO) is an assessment and data management platform that allows educators to enter scores or digitally administer measures, and receive immediate results and actionable data through an interactive experience. ALO enables educators to quickly identify student needs and make informed decisions regarding instruction and intervention.

A Resource That Supports the Outcomes-Driven Model for Student Success

ALO supports decision-making within an Outcomes-Driven Model, allowing educators to more efficiently analyze data.

Learn more:
voyagersopris.com/acadience
800.547.6747

To see a list of available assessments, please contact your Voyager Sopris Learning® Sales Representative.
Grades and Tests Can Undermine Learning. Do They Have To?

Assessments in this country can send the wrong message to students

By Rick Hess

Amid the heated debates over testing, “grading for equity,” and test-optional college admissions, it’s easy to throw up one’s hands in exasperation. That’s what makes a new book by Ethan Hutt and Jack Schneider so timely. In Off the Mark: How Grades, Ratings, and Rankings Undermine Learning (But Don’t Have To), they mostly make the case against traditional testing and grading—but they’re sharp enough to have penned a provocative volume that complicates simple-minded dogmas. Hutt is a professor at the University of North Carolina and Schneider at the University of Massachusetts-Amherst, where he leads the Beyond Test Scores Project. While I frequently disagree with their prescriptions, I also value their acuity and tend to learn a lot from them. So, I thought it worth chatting with them about the volume. Here’s what they had to say.

—Rick

Jack: We tried to do three things in this book. The first was the easiest: to make clear the unintended consequences associated with what we call “assessment technologies.” Most people don’t think of grades, test scores, and transcripts as technologies, but that’s exactly what they are. And they are highly problematic technologies that end up distorting the learning process. The second goal was to explain how we got here. In most cases, the people who created these technologies didn’t have bad intentions—they were trying to solve real problems. Understanding those problems is essential if we want to generate successful reforms. The third goal of the book was to do the impossible: to suggest a fix for this mess.

Rick: What’s wrong with traditional approaches to tests and grading?

Jack: You’ll need to turn over several more columns to us if you want a full answer to that question. But let’s touch on a couple. One major problem is that our assessment technologies are maximally reductive; a single letter or number is supposed to tell us what a student knows. Because of this, the symbol—the grade or the test score—ends up being what students focus on. The message they receive is clear: It doesn’t matter what you learn, provided you get good marks. A second problem is that grades, test scores, and transcripts are permanent—it’s as if they’ve been etched in stone. But students grow and change. As a result, their assessment records tend to reflect the past more so than the present. Several decades ago, I didn’t know how to ride a bike; today, however, that previous lack of knowledge is completely irrelevant. Imagine if I had to carry around a permanent record of all the things I used to be unable to do.

Rick: Given these concerns, I was struck that you guys seem to have a fairly nuanced stance on standardized testing. Can you speak to that?

Ethan: We would definitely win more friends in reform circles if we adopted a simple slogan like “abolish testing.” But that isn’t particularly realistic or productive. Standardized tests have historically served two distinct functions in our system. The first is synchronization. Our system is very decentralized—we don’t have common standards, curricula, or textbooks—which means that despite being in the same grade or even the same course, students in different areas can have very different experiences. Yet, at key moments, we need the disparate pieces of our system to fit together; that’s
what standardized testing facilitates. The other critical function standardized tests serve is communication. Students’ scores get used and abused in all kinds of ways, but they also can provide important information about those students. You don’t see special education professionals clamoring to get rid of diagnostic tests, for instance.

Rick: I was going to say, given your testing skepticism, that I was wondering whether some ardent testing critics may think your stance here isn’t as strong as they’d like. Has that come up?

Jack: I’m not sure anyone has been more outspoken than I have about the excesses of standardized testing. Test scores are a rotten measure of school quality; the high-stakes use of test scores in state accountability systems is totally misguided and has resulted in a host of unintended consequences; and our use of test scores to sort students has exacerbated inequity. But let’s also recognize that a well-designed test can offer useful information. Educators use test results in productive ways every day, and, as Ethan mentioned above, they tie our system together in ways that we can’t just dismiss.

Rick: What do you guys make of the push by many in higher education to move away from the SAT and ACT?

Ethan: One main argument in the book is that reforms have to be considered in the context of the dynamics of the whole system. A change in one part of the system is likely to reverberate. In this case, eliminating the SAT or ACT is certain to ramp up the competition for distinction in other parts of students’ applications: their grades, Advanced Placement scores, letters of recommendation, extracurriculars, and essays. From an equity standpoint, that’s a problem. Many of these elements are even more strongly associated with socioeconomic status than are SAT and ACT scores.

Rick: I’m also curious of what you think of the push for “equity grading,” no-zero policies, and similar proposals?

Jack: I don’t think they’re going to do any harm, and in some cases, they’ll make a positive difference. This is an example of the kind of tinkering that educators have the freedom to do in their classrooms, and I support it. But it isn’t going to address the fundamental problems with our assessment technologies. A no-zero policy, for instance, isn’t going to change the fact that students are still focused on the acquisition of a token—the grade.

Rick: Are there any points made by defenders of traditional grading that you find particularly persuasive or that deserve more consideration than they usually receive?

Ethan: Grades and test scores both serve critical functions in our system. When defenders of traditional grading say that they believe grades should honestly communicate how a student is doing in class, we completely agree. Our argument is that this critical communication function has been undermined because of all the other tasks we’ve burdened grades with. Part of our goal is to figure out how we can create more space for teachers to communicate candidly about students’ work in the moment without students fearing for their futures.

Rick: So, Ethan, if just “abandoning” traditional assessment and grading isn’t the way to go, what’s a better path forward?

Ethan: I think the first step at all levels is to look at current practices and ask what can be removed or adjusted while still preserving the three core functions of assessment: motivation, communication, and synchronization. We know, for instance, that much of the perceived overkill in standardized testing comes not from federal or state mandates but from district decisions. The amount and frequency of testing could be pared back to de-emphasize testing and make more room for learning.

Rick: Given that, what kind of practical advice can you offer to teachers and school leaders?

Jack: At the classroom level, there are things that educators can do, like anchoring assessment in projects that are meaningful to students. But we also need to be acting at the school level, the district level, the state level, and even the national level. College-admissions practices, for instance, need to change.

Rick: OK, so what are a couple of actions that you’d encourage district leaders to adopt?

Jack: States don’t mandate specific approaches to grading, so there are definitely actions districts can take on that front. For instance, we discuss making grades “overwritable.” In essence, students would have the ability at particular points in the school year to demonstrate competencies; and if they develop a competency, new information should overwrite old information.

Rick: Ethan, you have a final thought on that count?

Ethan: Beyond just overwriting old information with new, district leaders could also add new information to a transcript by making it what we call “double-clickable.” Imagine if a college or employer could click on a grade and see a representative piece of work from the class. We could shift the focus from earning a grade—by whatever means—to a focus on what students are able to do.

Rick Hess is a resident senior fellow and director of Education Policy Studies at the American Enterprise Institute.
Let’s Dump the Obsession With Standardized Testing

By Larry Ferlazzo

What are other ways than standardized-test scores to evaluate the effectiveness of schools?

Today, Denita Harris, Jennifer Mitchell, Rebecca Alber, and Amanda Kipnis offer their answers.

‘Formative Assessments’

Denita Harris serves as a district administrator in the MSD of Wayne Township in Indianapolis. Harris is currently serving in her 24th year in education as the new chief diversity, equity, and inclusion officer for her school corporation:

Standardized-test scores should not be the only indicator that defines the effectiveness of a school. Schools should be evaluated by using a holistic approach that encompasses all educators are required to do when it comes to servicing students.

The days of solely teaching academics are long gone, if they ever existed. Instead, schools are working to educate the “whole” child. A whole-child approach requires that students’ emotional, relational, intellectual, academic, and mental health are cultivated. To my knowledge, there is no standardized test that is able to measure the effectiveness of the number of schools that go beyond schooling students to educating them to thrive in a global society.

When people want to know how effective a school is in educating the whole child, student and family/caregiver voice, academic growth, and student interests/activities, which create a sense of belonging, are key. Student and family/caregiver voices are essential when evaluating the effectiveness of a school. Educators are in the service industry, and like any other industry that provides a service, we should want to know what our students and families/caregivers have to say about the educational services we provide.

We should be asking how our students and families feel when they enter our school buildings. We should want to know if students find the curriculum engaging and challenging or are our students often bored or completely lost when we do not explain things well. Our families could provide meaningful feedback regarding communication from the school and if they truly feel like they are partners in their child’s education.

Analyzing academic growth on a frequent and consistent basis by utilizing formative assessments is a much better way to determine how effective a school is in educating students. As adults, we understand that not all students will excel in all things; however, as educators, we have a responsibility to set the bar high, maintain high expectations, and focus on students’ assets and not their deficits. When we measure individual growth and make the necessary instructional shifts, we remove the danger of competition and comparison, which can result in low student self-efficacy.

There is an innate desire in all of us to have a sense of belonging and, as a result, be valued for our contributions. Opportunities that allow students to exercise their gifts—sports, drama, speech, sewing, debate, art, etc.—allow students to shine and stand in their truth. Evaluating schools by what opportunities they allow students to get involved in and how many students participate in these opportunities is important in evaluating how diligent and attentive schools are in ensuring every student has an outlet and an opportunity to be a part of the larger school community.

One number on one test should never determine the effectiveness of a school, just like one number on one test should never determine a student’s educational trajectory.

‘Goal-Setting & Improvement’

Jennifer Mitchell teaches ELs in Dublin, Ohio. Connect with her on Twitter: @readwritetech or on her blog:

Twenty years after the passage of No Child Left Behind, many educators and students agree that with increased standardized testing and the pressures attached to it, in fact, too many students are being left farther behind. What have we gained from these years of attempting to quantify the unquantifiable (what does it really mean to score “435” on an English test that included 30+ multiple-choice questions and several essays?) and measure unique individuals against poorly defined metrics and each other, with such high stakes that students and teachers feel their futures are on the line over a few spring mornings? Stress, confusion, frustration at the lack of agency, and a sense of hopelessness.

Instead, let’s measure students, teachers, and schools by what we hope they do: foster...
academic and social-emotional growth over time, while affirming the unique identities of each student and staff member. At the end of each school year, are all students and staff more skilled in the areas they’ve practiced than they were at the beginning? Are they happier and healthier, with a better sense of who they are and how they can leverage their uniqueness to impact the world? Do they feel a greater sense of belonging and connection to their school community?

And when we measure these areas, we need to include the participants in the process. Students and staff members should actively contribute to their own assessments by reflecting on their current skills and emotions, demonstrating and analyzing their growth over time, and setting meaningful goals for continued improvement.

In my English-learner classroom, students take pride and ownership in showing their growth throughout the year with frequent goal-setting and reflection in a variety of key areas:

- Reading, writing, and academic / SEL skills forms completed three times a year (in the first month, at the semester, and year-end)
- Current skill-set reflection completed in the first month and at semester, using their form responses
- Daily overview of current learning targets and mini-reflections at the end of each class
- Weekly reading goal and thinking tracker, completed every Monday (green boxes) and Thursday (orange boxes)
- Writing reflection sheet, completed after every major writing assignment and/or unit
- Monthly personal-goal work around a chosen academic or SEL skill, with check-ins every Monday
- Weekly reflection form including sections on academic and SEL progress, as well as a chance to tell me anything else they want (which often yields incredible insights!)
- Frequent conferring with me, while I document key insights to refer to in future conferences (students do a preconference form for writing)
- Some unit-specific goal setting and reflections, such as personalized goals and reflections for our “Slice of Life” blogging unit
- Quarterly class feedback forms where students can voice their perspectives about our class activities and environment
- Year-end class anthology submissions where each student fills two pages with selections of their best and/or favorite writing to submit to a spiral-bound class anthology; copies are given to all class members, the principal, and kept in the classroom for future students to read!
- Year-end annotations of two pieces of their best writing from their anthology selections; students copy the two selections into one Google Doc and use the commenting feature to highlight and explain the writing strategies demonstrated in each piece
- A year-end creative display of their reading growth, visually grouping their book choices in some way (by genre, difficulty, how much they enjoyed them, etc.), accompanied by short reflections on the impact of each book on their growth as a reader. Students design their own visual representation; some students have created reading ladders, maps, bookshelves, or interactive digital presentations.
- Year-end reflections on their reading and writing growth, analyzing evidence from the above reflections and other classwork

When my students are directly involved in the process of analyzing their own work, reflecting on their growth, and setting informed goals for future improvement, assessment becomes an empowering, participatory, ongoing process. They know where they are with specific skills and how they have improved and they know how to use feedback to inform their next steps. Moreover, they feel supported and capable. That’s much more meaningful than any number on a standardized test.

‘Digital Portfolio’

Rebecca Alber is an instructor at UCLA’s Graduate School of Education. A teacher educator and literacy specialist, she believes education’s true purpose is for liberation and transformation:

Standardized tests are typically one sitting and on one day. Along with the pressure of limited time and conditions (fluorescent-lit, cramped classroom, multiple choice only), can this really produce reliable and valid results? It cannot. In fact, no single measure properly captures what a learner knows and can do and who they are. Simply put, a standardized, multiple-choice test is just a single “snapshot” from one day.

What we need is a “photo album” for each learner. Curriculum experts Carol Ann Tomlinson and Jay McTighe describe this photo album as containing various “photos” (assessments) given in different contexts to reveal a more accurate “portrait” of a child’s knowledge and skills. This means presenting students with assessments on different days using varied modalities. But as we know, this takes time, resources, and commitment from a school/district/state. And the single measure, standardized, multiple-choice test is the least expensive (computers do the scoring) and also supports the meritocratic educational system currently and firmly in place in this country—one built around ranking schools and children.

Students with neurodiversity, English-learners, and poor test-takers particularly get a raw deal when a school overrelies on the single measure and modality standardized test to determine its success. (It’s important to note that when a school does do this, what is revealed is not actually the limitations of the test-taker but rather the limitations and deficits of the school.)

So, let’s circle back to the photo album approach to assessment: What might this look like?

Think digital portfolio. In that portfolio, a student may have artwork, written pieces that are brief and poetic, and longer pieces that are more detailed and explanatory. There may be audio recordings of the student narrating thoughts on new content learned or explaining a new concept or idea and reflecting on the benefits of this new learning. The portfolio could include slide presentations; quiz and test results; problem-solving charts, graphs and reports; multimedia projects; and also self-assessments, along with peer reviews and teacher reviews of the student’s work.
Teachers should then be provided with pupil-free days so that they can periodically look through these portfolios together, discuss, evaluate, and determine individual and collective student growth. This data can then be shared and discussed with the school leadership team.

Additionally, when a school is determining how effective it is, the school should survey students, families, and faculty so they can provide feedback to the leadership team on such things as the learning environments and school culture and community.

If our education system moved away from this obsession with the single measure of standardized testing and moved toward a more holistic and humanizing approach to evaluating student learning, think of all the ways this could transform our schools and positively impact children.

School Culture

Amanda Kipnis is a passionate educator who teaches a 3rd-5th grade special day class for students with moderate to severe disabilities in Lemon Grove, Calif. Amanda was named to the 2020 class of Curriculum Associates Extraordinary Educators:

I, along with thousands of other educators, have appreciated the limited pandemic shift away from standardized testing. I know it’s only a matter of time before we go back to the “old ways;” however, this pandemic has forced us to think outside the box and find new ways of assessing student growth and analyzing the effectiveness of a school. One topic I find promising is the focus on school culture and its correlation to student success.

Kent D. Peterson from the University of Wisconsin-Madison defines school culture as “the set of norms, values and beliefs, rituals and ceremonies, symbols and stories that make up the ‘persona’ of the school.” School culture is more than just the feeling of a school. It is the soul of a school and includes everything school-related.

School culture and its connection to academic achievement was not truly researched until the 1970s. Since then, we’ve learned that a positive school climate is correlated with increased student achievement while a negative school culture is associated with decreased student success. Therefore, assessing school culture can give you a glimpse into how students are performing academically.

Fortunately, assessing school culture can be very simple with an anonymous survey. A very simple survey exists that evaluates the three main aspects of school culture: collaboration, collegiality, and teacher efficacy. The School Culture Triage Survey consists of 17 questions and a rating scale from 1-5 (ranging from Never to Always/Almost Always). There are five questions about professional collaboration, six questions about collegial relationships, and six questions about teacher efficacy/self-determination. Questions about collaboration look at how and why staff collaborate and includes determining if teachers and staff discuss instructional strategies and curriculum issues. Questions about collegial relationships look at staff behavior in and out of the school environment and includes asking if teachers and staff tell stories of celebrations that support the school’s values. The section on teacher efficacy looks at how staff members influence each other and how they respond to adversity. It includes questions asking if members of the school community seek alternatives to problems/issues rather than repeating what they have always done.

Knowing this correlation exists, many educators find it’s frustrating how few staff meetings and professional developments have been devoted to school culture. The good news is that districts focusing on staff and student well-being may influence more than just what SEL curricula they decide to use. The questions in the School Culture Triage Survey also provide concrete ideas leaders can immediately implement to improve their school’s culture. With a healthy school culture and high expectations for students, one can only assume a school is maximizing student learning and fostering student success!

Thanks to Denita, Jennifer, Rebecca, and Amanda for contributing their thoughts!

Larry Ferlazzo is an English and social studies teacher at Luther Burbank High School in Sacramento, Calif.